Inversion of Snow Parameters from Passive Microwave Remote Sensing Measurements by a Neural Network Trained with a Multiple Scattering Model

Leung Tsang, *Fellow, IEEE*, Zhengxiao Chen, Seho Oh, Robert J. Marks II, *Senior Member, IEEE*, and A. T. C. Chang, *Senior Member, IEEE*

Abstract— **The inversion of snow parameters from passive microwave remote sensing measurements is performed with a neural network trained with a dense media multiple scattering model. The basic idea is to use the input–output pairs generated by the scattering model to train the neural network. Once the neural network is trained, it can invert snow parameters speedily from the measurements. In this paper, we have performed simultaneous inversion of three parameters: mean-grain size of ice particles in snow, snow density, and snow temperature from five brightness temperatures. The five brightness temperatures are that of 19 GHz vertical polarization, 19 GHz horizontal polarization, 22 GHz vertical polarization, 37 GHz vertical polarization, and 37 GHz horizontal polarization. It is shown that the neural network gives good results for the inversion of parameters from the simulated data computed from the dense media radiative transfer equation which includes the effects of multiple scattering. For the simulated testing data, the absolute percentage errors for mean-grain size of ice particles and snow density are less than 10%, and the absolute error for snow temperature less than is 3 °K. We also use the neural network with the trained weighting coefficients of the three-parameter model to invert the SSMI data over the Antarctica region. The algorithm inverts 30 000 sets of 5-channel brightness temperatures of Antarctica in only 10 cpu min on a VAX 3500 workstation. Validity of the inversion results is discussed in view of the limited number of parameters that we used and the much more complicated real-life situation in the Antarctica.**

I. Introduction

Various techniques for solving inverse problems in remote sensing have been proposed in the last few decades [1]–[4]. The well-known technique in inversion is the Backus–Gillbert inversion technique [2]–[4]. In this technique, the approximation of single scattering is used so that the scattering measurements are linearly related to the medium parameters. However, in microwave remote sensing problems, especially for dry snow, multiple scattering can be a dominant effect. The relation between remote sensing measurements and the medium parameters are highly nonlinear. In this paper, we use an artificial neural network to invert snow parameters from passive microwave remote sensing measurements. The basic idea is to use the input–output pairs generated by the scattering model to train the neural network [5]. The scattering model we use includes multiple scattering effects. Once the neural network is trained, it can invert snow parameters speedily from the measurements.

An artificial neural network can be defined as a highly connected array of elementary processors called neurons. In this paper we consider the Layered Perceptron (LP) type artificial neural network [6]–[8]. The LP type neural network consists of one input layer, one or more hidden layers, and one output layer. Each layer employs several neurons and each neuron in the same layer is connected to the neurons in the adjacent layer with different weights. Signals pass from the input layer, through the hidden layers, to the output layer. Each neuron receives a signal which is a linearly weighted sum of all the outputs from the neurons of the former layer. The neurons then produce an output signal by passing the summed signal through the sigmoid function.

The backpropagation learning algorithm is used for training the neural network. This algorithm uses the gradient descent algorithm to get the best estimates of the interconnected weights, and the weights are adjusted after every iteration. The iteration process stops when a minimum of the difference between the desired and the actual output is reached by the gradient descent algorithm [7], [8].

The scattering model that is used to train the neural network is the dense media radiative transfer theory while the medium is modeled by a random collection of discrete scattering with size distribution [9]–[16]. The dense medium radiative transfer theory is different from the conventional or classical radiative transfer theory. In a dense medium with an appreciable fractional volume of scatterers (e.g., ice grains in snow), the assumption of independent scatterers, that is used in conventional radiative transfer theory, is not valid. This has been verified by controlled laboratory experiments [17], [18] and by Monte Carlo simulation by direct solution of Maxwell equation with configuration including up to 4000 particles [19]. Recently, we have developed the dense medium radiative transfer theory which accounts for correlated scattering and

Manuscript received April 22, 1991; revised January 15, 1992.

L. Tsang, Z. Chen, S. Oh, and R. J. Marks II are with the Department of Electrical Engineering, University of Washington, Seattle, WA 98195.

A. T. C. Chang is with the Hydrological Science Branch, Goddard Space Flight Center, National Aeronautics and Space Administration, Greenbelt, MD 20771.

IEEE Log Number 9201974.

which are derived from field theory using the quasicrystalline approximation of the Dyson's equation and the correlated ladder approximation of the Bethe–Salpeter equation [9]–[16]. The dense media theory also includes multiple scattering effects and agrees with controlled laboratory experiments [17], [18] and Monte Carlo simulations of direct solution of Maxwell equation [19]. The relations between brightness temperatures and the snow parameters are nonlinear under the dense media multiple scattering model.

We first use the dense media theory to compute the brightness temperatures for a half-space snow medium for the five channels using different combinations of input parameters of mean-grain size of ice particles in snow, snow density, and snow temperature. About 1000 sets of input–output pairs are generated in this manner that are well distributed in mean-grain size of ice in snow, snow density, and snow temperature. On a VAX 3500 workstation, it takes about 2 h cpu time to calculate all 5-channel brightness temperatures for these 1000 cases based on dense medium radiative transfer theory. These are used as training data for the neural network. Using the error backpropagation algorithm on these sets results in a set of weighting coefficients. The cpu time for training of 10 000 iterations with these 1000 sets of training data by this backpropagation algorithm is about 24 h on a DEC 3100 workstation. Thus the training time can be large. But once the training is complete, the actual inversion of parameters can be done speedily. We note that the multifrequency and two-polarization measurements are very important for the convergence of the weighting coefficients. Without either of them, there exist several inversions for the same observation. As a consequence, the weighting coefficients would not converge when these data were not used. The neural network then is tested by a set of simulated testing data which are also generated by the passive dense medium theory and are randomly distributed in mean-grain size of ice particles in snow, snow density, and snow temperature. We show that the neural network yields good results for the simulated testing data with absolute percentage errors for mean-grain size of ice particles in snow and snow density less than 10%, absolute error for snow temperature less than 3 °K. Our results demonstrate that neural network can perform speedy inversion of parameters from full multiple scattering model. Finally, we also use the neural network with the trained weighting coefficients to invert the SSMI data over the Antarctica region [12]. The algorithm inverts 30 000 sets of 5-channel brightness temperatures of Antarctica in only 10 cpu min on a VAX 3500 workstation. The results are quite encouraging in view of the fact that representation of the entire Antarctica region by three physical parameters is an oversimplified picture.

II. Dense Media Radiative Transfer Equation for Passive Remote Sensing

Consider thermal emission from a half space medium with dielectric particles of permittivity ϵ_s embedded in a background medium of permittivity ϵ. (Fig. 1, note for the case of snow, the background permittivity ϵ is ϵ_0 of free space and the particle permittivity ϵ_s will be that of ice.) The particle sizes obey a size distribution $n(a)$ which is the number of particles per unit volume with radii between a and $a + da$. The medium is of uniform temperature T. Then the dense media radiative equation for passive remote sensing assume the following matrix form [11] of dimension 2, for $0 \leq \theta \leq \pi$

$$\cos\theta \frac{\partial}{\partial z}\bar{I}(z,\theta) = -\kappa_e \bar{I}(z,\theta) + \kappa_e(1-\tilde{\omega})CT\begin{bmatrix}1\\1\end{bmatrix} + \frac{3}{8}\kappa_e\tilde{\omega}\int_0^{\pi} d\theta' \sin\theta' \overline{\overline{P}}(\theta,\theta')\cdot\bar{I}(z,\theta') \tag{1}$$

where

$$\bar{I}(z,\theta) = \begin{bmatrix} I_v(z,\theta) \\ I_h(z,\theta)\end{bmatrix} \tag{2}$$

and I_v, I_h are the vertical and horizontal specific intensities, respectively. Also in (1), $C = K_b K'^2/(\lambda^2 k^2)$, K_b is Boltzman's constant, λ is free space wavelength, K' is the real part of the effective wave number in region 1, k is the free space wave number, and $\tilde{\omega}$ is the albedo. In (1)

$$\overline{\overline{P}}(\theta,\theta') = \begin{bmatrix} p_{11}(\theta,\theta') & p_{12}(\theta,\theta') \\ p_{21}(\theta,\theta') & p_{22}(\theta,\theta')\end{bmatrix} \tag{3}$$

where

$$p_{11}(\theta,\theta') = 2\sin^2\theta\sin^2\theta' + cos^2\theta\cos^2\theta' \tag{4}$$

$$p_{12}(\theta,\theta') = \cos^2\theta \tag{5}$$

$$p_{21}(\theta,\theta') = \cos^2\theta' \tag{6}$$

$$p_{22}(\theta,\theta') = 1. \tag{7}$$

The boundary conditions for (1) are, for $0 \leq \theta \leq \pi/2$,

$$I_v(z=0,\pi-\theta) = r_v(\theta) I_v(z=0,\theta) \tag{8}$$

$$I_h(z=0,\pi-\theta) = r_h(\theta) I_h(z=0,\theta) \tag{9}$$

where

$$r_v(\theta) = \left|\frac{k^2\cos\theta - K'(k^2 - K'^2\sin^2\theta)^{1/2}}{k^2\cos\theta + K'(k^2 - K'^2\sin^2\theta)^{1/2}}\right| \tag{10}$$

$$r_h(\theta) = \left|\frac{K'\cos\theta - (k^2 - K'^2\sin^2\theta)^{1/2}}{K'\cos\theta + (k^2 - K'^2\sin^2\theta)^{1/2}}\right|. \tag{11}$$

In (9)–(11), K' is the real part of the effective propagation constant K of the medium which consists of the background and the dielectric particles with size distributions and $\kappa_e = 2$ Im(K) is the extinction rate of the specific intensity.

After (1) is solved subject to the boundary conditions of (8)–(9), the brightness temperatures in the direction θ_0, where

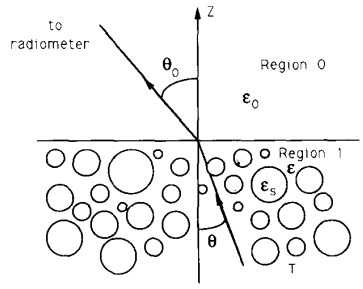


Fig. 1. . Passive remote sensing with observation of brightness temperatures of a half space medium with dielectric particles of permittivity ϵ_s embedded in a background medium of permittivity ϵ. The particles obey Rayleigh size distribution. The temperature of low half space is T. The observation is in direction θ_0.

$\theta_0 = \sin^{-1}(K' \sin\theta/k)$ is related to θ by Snell's law, in region 0 for vertical and horizontal polarizations are give by

$$\begin{bmatrix} T_{Bv}(\theta_0) \\ T_{Bh}(\theta_0) \end{bmatrix} = \frac{1}{C} \begin{bmatrix} (1 - r_v(\theta)) I_v(z=0,\theta) \\ (1 - r_h(\theta)) I_h(z=0,\theta) \end{bmatrix}. \tag{12}$$

The differences between the dense medium theory and the conventional radiative transfer theory are the calculations of K, the extinction rate κ_e and the albedo $\tilde{\omega}$ in terms of the physical parameters of the medium which are represented by ϵ_s and size distribution $n(a)$.

The cross pair distribution functions of multiple particle sizes are calculated through the Percus–Yevick approximation [13]–[16] that expresses the correlations in terms of the size distribution $n(a)$.

The calculation procedure for K, κ_e, and $\tilde{\omega}$ from ϵ_s and $n(a)$ is as follows. We first discretize $n(a)$ into L sizes: $a_1, a_2 \ldots a_L$, equally spaced at $\triangle a$. Then each size a_j is represented by a fractional volume f_j and the number density n_j which are

$$f_j = \frac{\triangle a}{2} \left[w(a_j - \frac{\triangle a}{2}) + w(a_j + \frac{\triangle a}{2}) \right] \tag{13}$$

$$w(a) = \frac{4\pi}{3} n(a) a^3 \tag{14}$$

$$n_j = \frac{f_j}{\frac{4\pi}{3} a_j^3}. \tag{15}$$

Using QCA-CP, the effective propagation constant K can be calculated from f_j, n_j and a_j as follows:

$$K^2 = k^2 + \frac{3K^2}{\hat{D}} \sum_{l=1}^{L} f_l \hat{y}_l \cdot \left(1 + i \frac{2K^3}{3\hat{D}} \left(a_l^3 \hat{y}_l + \sum_{j=1}^{L} a_j^3 8\pi^3 n_j \hat{y}_j H_{jl}(\overline{p}=0) \right) \right) \tag{16}$$

where

$$\hat{D} = 1 - \sum_{l=1}^{L} f_l \hat{y}_l \tag{17}$$

and

$$\hat{y}_j = \frac{k_s^2 - k^2}{3K^2 + (k_s^2 - k^2)}. \tag{18}$$

In (13)–(18), the subscript j refer to a particular species of particle size. H_{ij} is the three-dimensional Fourier transform of the total correlation function $h_{jl}(r)$ between a pair of particles of size j and l separated by a distance r. The total correlation function between two particles is : $h_{jl}(r) = g_{jl}(r) - 1$, where $g_{jl}(r)$ is the pair distribution function. The pair distribution function $g_{jl}(r)$ is proportional to the conditional probability that a particle of size j will be found at a distance r from a given particle of species l and vice-versa. Pair function of g_{jl} can be computed using the Percus–Yevick approximation [13]–[16].

The effective propagation constant is calculated by a simple iteration of (18). The initial guess for K is a real value K_0 determined by ignoring the imaginary term in (18) and solving the resulting nonlinear equation. The positive root is put into the right-hand side of (18) to iterate once to get K. The extinction rate is κ_e=2 Im(K). The albedo $\tilde{\omega}$ is calculated as follows:

$$\tilde{\omega} = \frac{2|K_0|^4}{\kappa_e |\hat{D}(K_0)|^2} \sum_{l=1}^{L} f_l \hat{y}_l(K_0) \cdot \left(a_l^3 \hat{y}_l(K_0) + \sum_{j=1}^{L} a_j^3 8\pi^3 n_j \hat{y}_j(K_0) H_{jl}(\overline{p}=0) \right). \tag{19}$$

Using the calculated results of K, κ_e and $\tilde{\omega}$ from ϵ_s and $n(a)$, the dense media radiative transfer equation can be solved readily to calculate the brightness temperature.

In the following numerical simulation and training of the neural network, we have used a Rayleigh size distribution [20] with

$$n(a) = \frac{\pi f a}{16\langle a \rangle^5} \exp\left(-\frac{\pi a^2}{4\langle a \rangle^2} \right) \tag{20}$$

where $\langle a \rangle$ is the mean radius and f is the fractional volume of all the particles. The fractional volume and the mean radius are defined by

$$f = \int_0^\infty da \frac{4\pi}{3} a^3 n(a) \tag{21}$$

$$\langle a \rangle = \frac{\int_0^\infty da a n(a)}{\int_0^\infty da n(a)}. \tag{22}$$

The advantage of the Rayleigh size distribution is that there are only two parameters f and $\langle a \rangle$.

Thus the medium input physical parameters of the dense medium radiative transfer equation using the Rayleigh size distribution are:

1. Complex permittivity of the particles ϵ_s
2. Fractional volume occupied by all the particles f

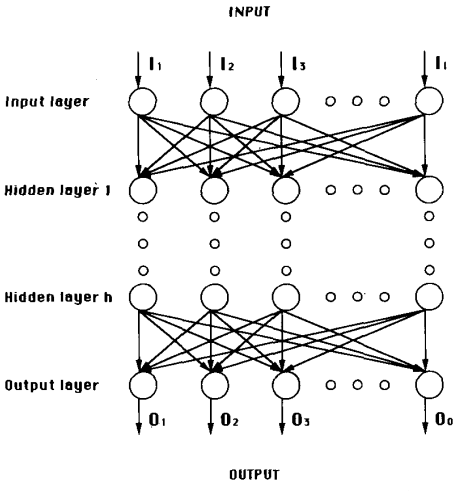


Fig. 2. Structure of a multilayered perceptron-type artificial neural network.

3. mean radius $\langle a \rangle$
4. physical temperature T

From these parameters, the brightness temperatures of vertical and horizontal polarization at specified frequencies and observation angles can be calculated by the above procedure.

III. The Layered Perceptron

A. Architecture

An artificial neural network (ANN) can be defined as a highly connected array of elementary processors called *neurons*. A widely used model called the layered perceptron (LP) ANN is shown in Fig. 2. The LP type ANN consists of one input layer, one or more hidden layers, and one output layer. Each layer employs several neurons and each neuron in a layer is connected to the neurons in the adjacent layer with different weights. Signals flow into the input layer, pass through the hidden layers, and arrive at the output layer. Each neuron receives signals from the neurons of the previous layer linearly weighted by the interconnect values between neurons. The neuron then produces its output signal by passing the summed signal through a sigmoid function [21]–[25].

A total of Q sets of training data are assumed to be available. Inputs of $\{\vec{i_1}, \vec{i_2}, \ldots, \vec{i_Q}\}$ are imposed on the corresponding target vectors, $\{\vec{t_1}, \vec{t_2}, \ldots, \vec{t_Q}\}$ on the bottom layer. In the training of the neural network with dense medium multiple scattering theory, we have used Q between 180 and 1000. An example of a calculation of a set of training data is as follows. We fix the complex permittivity if the ice particle in snow such that $\mathrm{Re}(\epsilon_s) = 3.2\epsilon_0$, $\mathrm{Im}(\epsilon_s) = 0.001\epsilon_0$. The three parameters are: $\langle a \rangle = 0.03$ cm, $f = 0.3$, $T = 270$ °K. Using dense medium transfer theory, we calculate five brightness temperatures at observation angle 53°, we get: $T_{Bv} = 233°K$, $T_{Bh} = 216°K$ at 19 GHz; $T_{Bv} = 219°K$ at 22 GHz; $T_{Bv} = 156°K$, $T_{Bh} = 142°K$ ar 37 GHz. These together give a set of training data:

$$\vec{i_1} = \begin{bmatrix} 233 \\ 216 \\ 219 \\ 156 \\ 142 \end{bmatrix} \quad \vec{t_1} = \begin{bmatrix} 0.03 \\ 0.3 \\ 270 \end{bmatrix}$$

All Q sets of training data are normalized between 0 and 1. The dimension of $\vec{i}$ is 5 and the dimension of $\vec{t}$ is 3. Note: In calculation of brightness temperatures using the parameters of ϵ_s, f, $\langle a \rangle$ and T from dense medium theory, the parameters are the input and the brightness temperatures are the output. However, in using the set of results as training data, the input $\vec{i}$ is the brightness temperatures, and the output $\vec{t}$ is the physical parameters. This is because we are using the neural network to do the inverse problem of finding the physical parameters from the brightness temperature.

The training is halted when the average error between the desired and actual outputs of the neural network over the Q training data sets is less than a predetermined threshold. The training time required is dictated by various elements including the complexity of the problem, the number of the training data pairs, the structure of network, and the training parameters used.

B. ANN Training

Although there are other methods, the *Generalized Delta Rule* (GDR) is the technique most commonly used to train the layered perceptron. The following procedure is used. The first vector, $\vec{i_1}$, is presented to the LP. The error is mean square error between the actual output and the desired output $\vec{t_1}$. The error is used to adjust the neural network weights. A similar adjustment is performed for the training pair $(\vec{i_2}, \vec{t_2})$, etc. After all Q data have been used, the process is repeated until the error is acceptably small.

The output O_i , from neuron i, is connected to the input of neuron j, through the interconnection weight W_{ij}. Except for the input layer, the state of the neuron j is :

$$O_j = f(\sum_i W_{ij} O_i) \tag{23}$$

where $f(x) = 1/(1 + e^{-x})$ and the sum is over all neurons in the adjacent layer. Let the target state of the output neuron be t_k. Thus, the error at the output neuron can be defined as

$$E_k = \frac{1}{2}(t_k - O_k)^2 \tag{24}$$

$$E = \sum_k E_k \tag{25}$$

where neuron k is the output neuron.

The gradient descent algorithm adapts the weights according to the gradient error, i.e.,

$$\triangle W_{ij} \propto -\frac{\partial E}{\partial W_{ij}} = -\frac{\partial E}{\partial O_j}\frac{\partial O_j}{\partial W ij} \tag{26}$$

Specifically, we define the error signal as

$$\delta_j = -\frac{\partial E}{\partial O_j} \tag{27}$$

With some manipulation, we obtain the following GDR:

$$\triangle W_{ij} = \eta \delta_j O_i \tag{28}$$

where η is an adaptation gain. δ_j in computed based on whether or not neuron j is in the output layer. If neuron j in one of the output neurons,

$$\delta_j = (t_j - O_j) O_j (1 - O_j) \tag{29}$$

If neuron j is not in the output layer,

$$\delta_j = O_j (1 - O_j) \sum_k W_{jk} \tag{30}$$

In order to improve the convergence characteristic, a momentum gain α is commonly addended to (28) [26].

$$\triangle W_{ij}(n+1) = \eta \delta_j O_j + \alpha \triangle W_{ij}(n) \tag{31}$$

where n represents the iteration index.

The training of the neural network is complete if the convergence of weighting coefficients has been achieved. The convergence criterion is that the mean square error at the output must be less than the desired error which is called tolerable error. Layered Perceptrons typically take a great deal of time to train. Once trained, however, for a given input, output can be generated quite rapidly. All that is required is a few multiplications, additions, and the sigmoid function calculation.

IV. Training and Testing of Neural Network Based on Simulated Data from Dense Media Theory

The procedure of training and testing of the neural network is summarized as follows.

The frequencies are set at 19 GHz, 22 GHz, and 37 GHz, which are the SSMI operating frequencies for microwave radiometer. The permittivities of the particles ϵ_s are fixed at values that correspond to dry snow at microwave frequencies. The observation angle θ_0 is set at 53° which is the observation angle used by SSMI.

The physical parameters of the snow medium are represented by:

1. Mean-grain size a_m of the ice particles in snow with assumed Rayleigh size distribution.
2. Snow density d. The fractional volume f occupied by the ice particles in snow is equal to the ratio of snow density to ice density. The ice density we use here is 0.91 g/cm^3
3. Temperature distribution of the lower half space. We assume the snow medium has a constant temperature $Temp$.

In the range of interest, We use dense medium radiative transfer theory and different combinations of $a_m(k)$, $d(k)$, and $Temp(k)$ to calculate the brightness temperatures $T_{19v}(k)$, $T_{19h}(k)$, $T_{22v}(k)$, $T_{37v}(k)$ and $T_{37h}(k)$ with 19, 22, 37 denoting frequencies in unit of GHz; v, h denoting vertical and horizontal, respectively, and k denoting the index of the data set. On a VAX 3500 workstation, it takes about 2 h cpu time to calculate all the 5-channel brightness temperatures for 1000 combinations of $a_m(k)$, $d(k)$ and $Temp(k)$. The input is $\vec{i}(k)$ = [computed brightness temperatures $T_{i\alpha}(k)$], $i\alpha$ = $19v$, $19h$, $22v$, $37v$, and $37h$, and the output is $\vec{t}(k)$ = [physical parameters of $a_m(k)$, $d(k)$, and $Temp(k)$]. $\vec{i}(k)$ and $\vec{t}(k)$ have to be normalized between 0 and 1 in order to be used by the neural network. Linear scaling is used here. To have better convergence in the training, the 1000 input–output pairs are randomly shuffled before supplying to the neural network to be trained. Sets of data are generated for the neural network in this manner. Part of them are used as training data, part of them are used as testing data.

Training is done by using the error back propagation algorithm described in Section III. The training is complete when error threshold is met. On a DEC 3100 workstation, it takes about 24 h cpu time to finish training of 10 000 iterations with 1000 sets of training data. The resultant weighting coefficients are used to invert a_m, d and $Temp$ from known brightness temperature $T_{i\alpha}$.

In this paper, the neural network we use has the following properties: 1) Number of hidden layers is three; 2) number of neurons in each hidden layer is five; 3) tolerable error at output is 0.00001.

A. Testing Results with Variation of Number of Iterations

In this case, we assume $\epsilon_s = (3 + i0.00025)\epsilon_0$ at 19 GHz, $\epsilon_s = (3 + i0.00028)\epsilon_0$ at 22 GHz, $\epsilon_s = (3 + i0.001)\epsilon_0$ at 37 GHz, $0.01\text{cm} \leq a_m(k) \leq 0.055\text{cm}$, $0.1\text{g/cm}^3 \leq d(k) \leq 0.7\text{g/cm}^3$, and $207°\text{K} \leq Temp(k) \leq 270°\text{K}$. We also have $T_{19v}(k)$, $T_{19h}(k)$, $T_{22v}(k)$, $T_{37v}(k)$, and $T_{37h}(k)$ as inputs, and $a_m(k)$, $d(k)$, and $Temp(k)$ as outputs to the neural network.

The testing results are shown in Figs. 3 (a), (b), and (c). Fig. 3(a) shows the error for mean-grain size of ice particles in snow. Fig. 3(b) shows the error for snow density. Fig. 3(c) shows the error for snow temperature. Figs. 3 (a), (b), and (c) demonstrate that increasing the number of iterations results in better convergence to the true value and hence lowers the errors. After 10 000 iterations, the absolute percentage error for mean-grain size is less than 10%; the absolute percentage error for snow density is less than 10%; the absolute error for physical temperature is less than 3 °K.

B. Investigation of Multifrequency and Polarization Effects

We assume $\epsilon_s = (3 + i0.00025)\epsilon_0$ at 19 GHz, $\epsilon_s = (3 + i0.00028)\epsilon_0$ at 22 GHz, $\epsilon_s = (3 + i0.001)\epsilon_0$ at 37 GHz, $0.01\text{cm} \leq a_m(k) \leq 0.055\text{cm}$, $0.1\text{g/cm}^3 \leq d(k) \leq 0.7\text{g/cm}^3$, and 207 °K ≤ Temp(k) ≤ 270 °K. We have considered the following three cases:

(1) Brightness temperatures $T_{19v}(k)$, $T_{22v}(k)$, and $T_{37v}(k)$ are inputs, $a_m(k)$, $d(k)$, and $Temp(k)$ are outputs to the neural network. In this case, the input has three vertical brightness temperatures, and has no horizontal brightness temperatures.

(2) Brightness temperatures $T_{19v}(k)$ and $T_{19h}(k)$ are inputs, $a_m(k)$, $d(k)$, and $Temp(k)$ are the outputs to the neural network. In this case, the input has both vertical and horizontal brightness temperatures. However, there is only one operating frequency at 19 GHz.

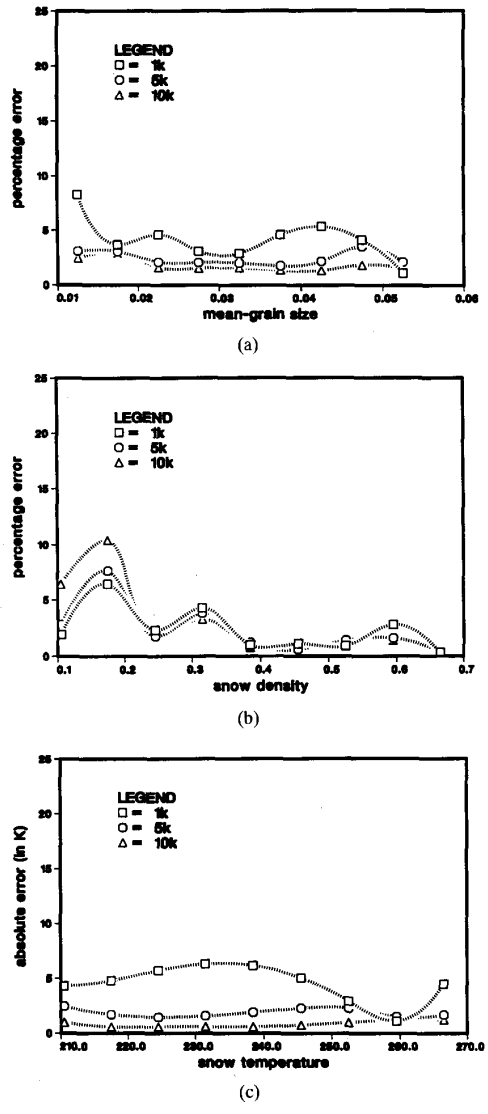


Fig. 3. Performance of the neural network with variation of number of iterations. In case (1), the number of iterations is 1000. In case (2), the number of iterations is 5000. In case (3), the number of iterations is 10 000. (a) The percentage error for mean-grain size of ice in snow. (b) The percentage error for snow density. (c) The absolute error for snow temperature.

(3) Brightness temperatures $T_{19v}(k)$, $T_{19h}(k)$, and $T_{22v}(k)$ are inputs, and $a_m(k)$, $d(k)$, and $Temp(k)$ are outputs to the neural network. In this case, input has both vertical and horizontal brightness temperatures at 19 GHz, and another vertical brightness temperature at 22 GHz.

The testing results are shown in Figs. 4 (a), (b), and (c). Fig. 4(a) shows the error for mean-grain size of ice particles in snow. Fig. 4(b) shows the error for snow density. Fig. 4(c) shows the error for snow temperature. Figs. 4 (a), (b), and (c) demonstrate that the errors for case (3) are much lower than errors for cases (1) and (2). This means multifrequency and two-polarization measurements are both important for the convergence of weighting coefficients. Without either one, the weighting coefficients diverge. Snow density is related to the effective permittivity of the snow medium [10], [15] in the dense medium radiative transfer theory which expresses the effective permittivity as a function of fractional volume of ice particles and ice permittivity. The effective permittivity affects strongly the difference between the vertical and horizontal polarized reflectivities. Since brightness temperature decreases with the increase of reflectivity, the snow density affects the difference between vertical and horizontal brightness temperatures. Thus for the inverse problem, if two-polarization measurements are absent, the neural network lacks information to determine snow density. As a result, the weighting coefficients do not exhibit convergence. Different frequencies will give different ratio of mean-grain radius to wavelength. Volume scattering depends strongly on mean-grain size. For inverse problem, if the multi-frequency measurements are absent, the neural network lacks information to determine mean-grain size. In such a case the weighting coefficients do not converge.

V. Training Based on Dense Media Theory and Testing on SSMI Data

In the previous section, we have demonstrated that neural networks perform very well in the retrieval of parameters from a full multiple scattering dense medium radiative transfer scattering model involving three parameters. The reason for choosing only three medium parameters in this initial study is of the large increase of training time with increase of the number of parameters. (Presently, we are working on a case involving four medium parameters and have good success). Nevertheless, as an illustration, it is important to see what the trained neural network from the previous section of a scattering half space will give when applied to real-life data. The example chosen in this section is that of the Antarctica in winter (July-August) in view of the fact that there is no sharp reflecting boundary in the subsurface of the Antarctica which seems to fit the configuration of half space that is used in previous sections and the winter season fits the dry snow scattering model with zero volumetric water content. However, the real Antarctica is a much more complicated situation than just a half space with three parameters. At least a few more parameters will play important roles in the scattering model. However, these are major points to be considered in the scattering model and not for the neural network. In the following, we first summarize the major parameters that should be added in the future and comment on their relative importance for the Antarctica. Then we illustrate what the trained neural network of three parameters of the previous section will give when applied to the Antarctica data.

1. The foremost important effect not considered in the half space model is the profile-layered structure nature of the subsurface of the Antarctica [27]. The profile includes a gradual increase of snow density from small values at the surface to large values of ice-like situation in the deep subsurface. This layered structure also accounts for the large difference between vertical and horizontal brightness temperatures that are observed in the Antarctica. To incorporate these effects into the model require

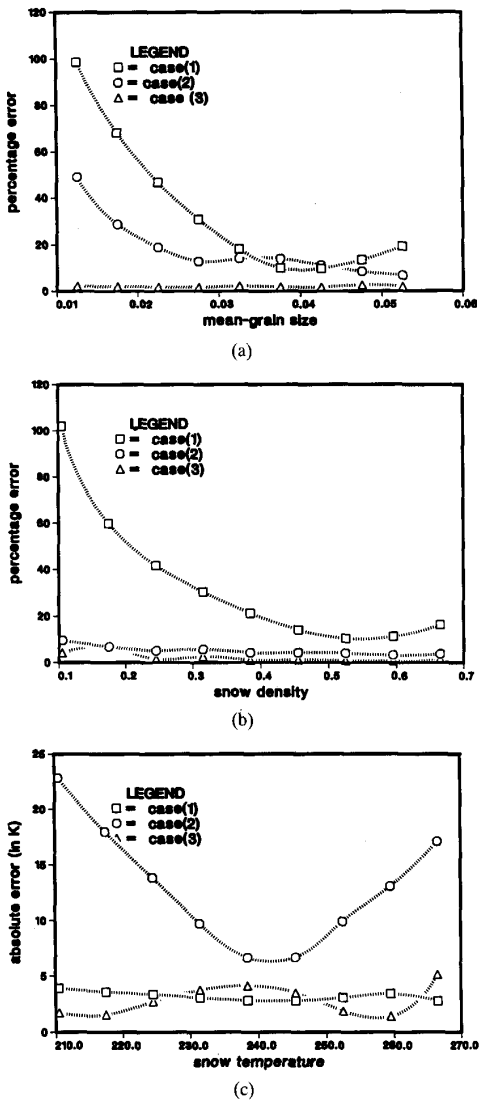


Fig. 4. Testing of the neural network in multifrequency and two-polarization effects. In case (1), the input has multifrequency measurements only. In case (2), the input has two-polarization measurements only. In case (3), the input has both. (a) The percentage error for mean-grain size of ice in snow. (b) The percentage error for snow density. (c) The absolute error for snow temperature.

several more parameters, including the different surface and subsurface values for snow density, grain size, and temperature and the rates of change with depth.

2. Surface roughness parameters are not considered. However, in passive remote sensing of snow over Antarctica, unlike other scattering problems, the air–snow interface in less reflective. Volume scattering, rather than surface scattering, dominates, and surface scattering is less important.
3. Snow-melt is not included. However, in this paper, we only apply the inversion to winter dry snow.
4. Atmospheric effects are not included. This can be included readily in the radiative theory by introducing more parameters.

The dense media model has been used to match experimental data for different snow condition and have been compared well with ground truth information [28]. In this section, we use the neural network trained with passive dense medium theory to invert three snow physical parameters: mean-grain size of ice particles in snow, snow density, and snow temperature, from available SSMI data over Antarctica in July/August which are brightness temperatures at 19 GHz vertical polarization, 19 GHz horizontal polarization, 22 GHz vertical polarization, 37 GHz vertical polarization, and 37 GHz horizontal polarization.

To train the neural network, we assume $0.01\text{cm} \leq a_m(k) \leq 0.055\text{cm}$, $0.1\text{g/cm}^3 \leq d(k) \leq 0.7\text{g/cm}^3$, and 207 °K $\leq Temp(k) \leq$270 °K, which is the same range as in Section IV. We also used $T_{19v}(k), T_{19h}(k), T_{22v}(k), T_{37v}(k), T_{37h}(k)$ are inputs, $a_m(k)$, $d(k)$, $Temp(k)$ are outputs to the neural network. We use $\epsilon_s = (3 + i0.00025)\epsilon_0$ at 19 GHz; $\epsilon_s = (3 + i0.00028)\epsilon_0$ at 22 GHz; $\epsilon_s = (3 + i0.001)\epsilon_0$ at 37 GHz. The model is valid for a large range of a_m and $Temp$. We chose a fairly large range for a_m, and $Temp$ so that any possible ground truth is within that range. The iteration number used here is 10 000. A set of weighting coefficients is obtained after the training and is used to invert the physical parameters of snow. Instead of testing on simulated data as done in Section IV, we use the 5-channel brightness temperatures from SSMI data as inputs to the neural network. Then the outputs of the neural network are considered as the three snow physical parameters: mean-grain size of ice in snow, snow density, and snow temperature of the corresponding geophysical terrain. An advantage of using neural network method here is that the amount of SSMI data is voluminous and neural network can process these data speedily with trained weighting coefficients. In this manner we process 30 000 sets of 5-channel brightness temperature SSMI data over Antarctica region in 10 cpu min on a VAX 3500 workstation and produce contour plots of the three physical parameters: mean-grain size of ice particles in snow, snow density, and snow temperature for the Antarctica region. Fig. 5 shows the results. Figs. 5(a) and (b) show, respectively, latitudes and longitudes of the Antarctica region which we have investigated, with I and J showing the corresponding grid coordinates called polar stereographic coordinates. In Fig. 5(c), we show a contour of vertical brightness temperatures at 19 GHz for that region. In Figs. 5(d), (e), and (f), we show, respectively, the contour plots of mean-grain size of ice particles in snow, snow density, and snow temperature for the same region.

To examine the validity of the inversion, we take the outputs from the network and use the dense media model to calculate the brightness temperatures again and compare them with the corresponding brightness temperatures from SSMI data. For the SSMI data all over the region where the latitude is greater than 70° (Fig. 5 (a)), the rms brightness temperature error is 19 °K. The inversion effort in Figs. 5(a)–(f) is encouraging because we represent the entire Antarctica by only three physical parameters which is an oversimplified picture of the Antarctica. The low physical inverted temperature can be attributed to two reasons. First, the data is taken in winter so that the temperature is low. Secondly, Antarctica has a

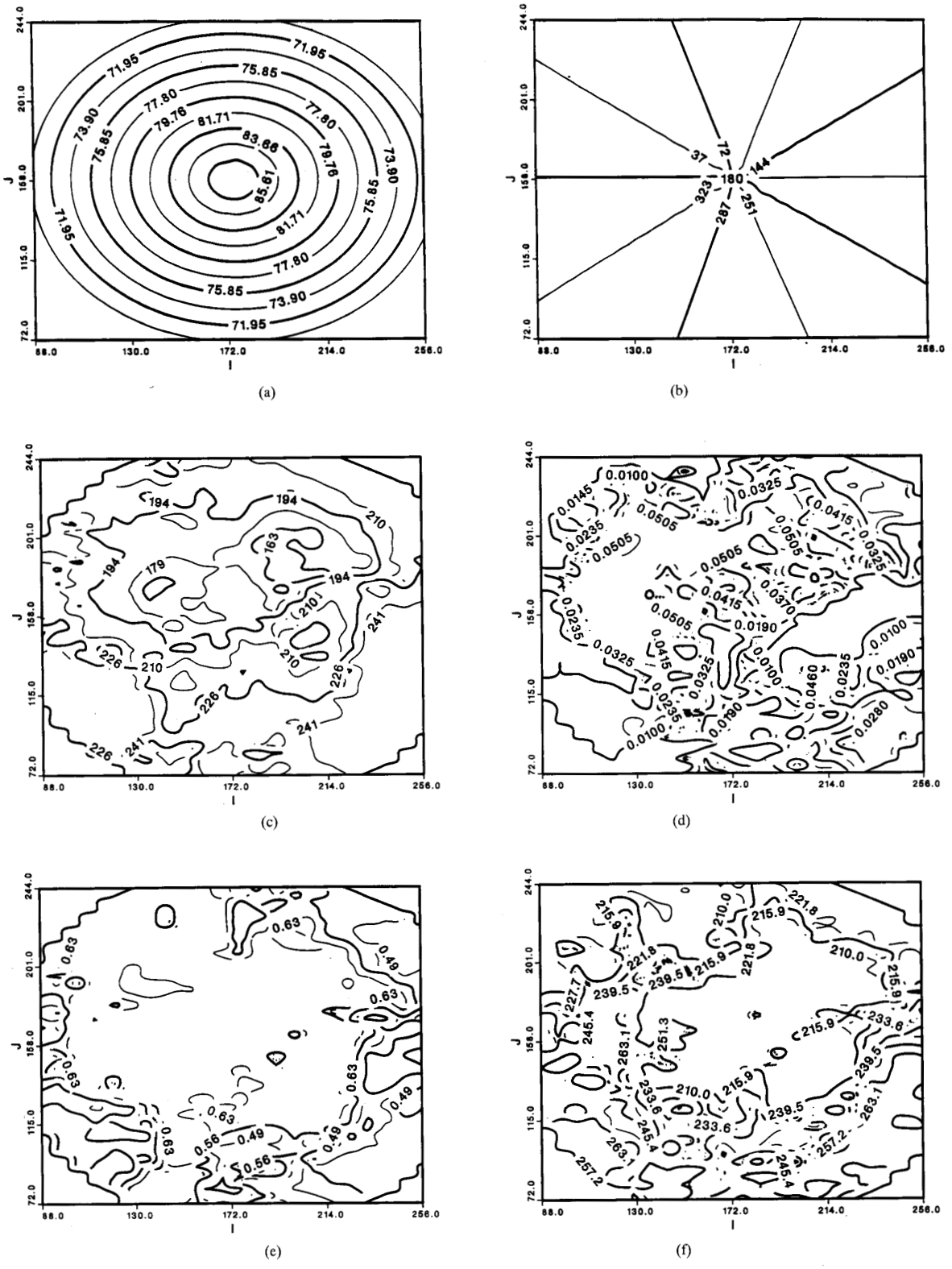


Fig. 5. Inversion of snow parameters from five-channel brightness temperatures of SSMI data. I, J are the polar stereographic coordinates used by SSMI satellites. (a) Corresponding latitudes. (b) Corresponding longitudes. (c) Contour of vertical brightness temperatures at 19 GHz. (d) Contour of mean-grain radius of ice particles in snow in cm. (e) Contour of snow density in g/cm^3 (f) Contour of snow temperature in °K.

profile and layered structure that has been neglected in the present three-parameter model. The layered structure generally is more reflective and decreases the brightness temperature. With that extra reflective effect ignored, the inverted physical temperature can be smaller than it should be.

VI. Conclusion

The inversion of snow parameters from passive microwave remote sensing measurements are performed with a neural network trained with a dense media multiple scattering model. The basic idea is to use the input–output pairs generated by the scattering model to train the neural network. The total training time includes the generation of input–output pairs based on dense media theory and the training of the neural network with the input–output pairs and backpropagation algorithm. The training time can be large. However, once the neural network is trained, it can invert snow parameters speedily from the measurements. We note that the multifrequency and two-polarization measurements are very important for the convergence of the weighting coefficients of the neural network. Without either of them, the weighting coefficients diverge. It is shown that the neural network yields good results for the simulated testing data with absolute percentage errors for mean-grain size of ice particles in snow and snow density less than 10%, absolute error for snow temperature less than 3° K. We also use the neural network with the trained weighting coefficients to invert the SSMI data over the Antarctica region. The available ground truth information at selected sites can also be incorporated in the training by tuning the neural network. The result neural network can interpolate well between the ground truth and the model predictions. This subject is also presently studied. Model refinement and improvement of inversion algorithm are presently being developed.

References

[1] D. L. Phillips, "A technique for the numerical solution of certain integral equations of the first kind," *J. Assoc. Comput. Mach.*, vol. 9, pp. 84–97, 1962.

[2] A. Ishimaru, *Wave Propagation and Scattering in Random Media*, vol. 2. New York: Academic Press, 1978.

[3] G. Backus and F. Gilbert, "Uniqueness in the inversion of inaccurate gross Earth data," *Phil. Trans. R. Soc. London Ser.* A266, pp. 123–192, 1970.

[4] E. R. Westwater and A. Cohen, "Application of Backus-Gilbert inversion technique in determination of aerosol size distributions from optical scattering measurements," *Appl. Opt.*, vol. 12, pp. 1340–1348, 1973.

[5] A. Ishimaru, R. J. Marks II, L. Tsang, C. M. Lam, and D. C. Park, "Particle size distribution determination using optical sensing and neural networks," *Opt. Lett.*, vol. 15, no. 21, pp. 1221–1223, 1990.

[6] L. Atlas, R. Cole, Y. Muthusamy, A. Lippman, G. Connor, D. Park, M. El-Sharkawi, and R. Marks II, "A performance comparison of trained multi-layer perceptrons and trained classification trees," submitted to *Proc. IEEE* Special Issue on Neural Networks, Aug. 1990.

[7] D. E. Rumelhart and J. L. McClelland, Eds., *Parallel Distributed Processing*. Cambridge, MA: MIT Press, 1986.

[8] R. P. Lippmann, "Introduction to computing with neural nets," *IEEE ASSP.*, vol. 4, no. 2, pp. 4–22, Apr. 1987.

[9] L. Tsang, J. A. Kong, and R. T. Shin, *Theory of Microwave Remote Sensing*. New York: Wiley-Interscience, 1985.

[10] B. Wen, L. Tsang, D. P. Winebrenner, and A. Ishimaru, "Dense medium radiative transfer theory: comparison with experiment and application to microwave remote sensing and polarimetry," *IEEE Trans. Geosci. Remote Sensing*, vol. 28, pp. 46–59, 1990.

[11] L. Tsang, "Passive remote sensing of dense nontenuous media," *J. Electromagnetic Waves and Applications*, vol. 1, no. 2, pp. 159–173, 1987.

[12] A. T. C. Chang, B. J. Choudhury, and P. Gloersen, "Microwave brightness of polar firn as measured by Nimbus 5 and 6 ESMR," *J. Glaciology*, vol 25, no. 91, 1980.

[13] R. J. Baxter, "Ornstein–Zernike relation for a disordered fluid," *Australian J. of Physics*, vol. 21, pp. 563–569, 1968.

[14] R. J. Baxter, "Ornstein–Zernike relation and Percus–Yevick approximation for fluid mixtures," *J. of Chemical Physics*, vol. 52, pp. 4559–4562, 1970.

[15] K. H. Ding and L. Tsang, "Effective propagation constants in media with densely distributed dielectric particles of multiple sizes and permittivities," Ch. 3 of *Progress in Electromagnetic Research*, vol. 1., J. A. Kong, Ed. New York: Elsevier, 1989, pp. 241–295.

[16] L. Tsang, "Dense media radiative transfer theory for dense discrete random media with spherical particles of multiple sizes and permittivities," Ch. 5 of *Progress in Electromagnetic Research*, vol. 6. New York: Elsevier, 1992.

[17] A. Ishimaru and Y. Kuga, "Attenuation constant of coherent field in a dense distribution of particles," *J. Opt. Soc. Amer.*, vol. 72, pp. 1317–1320, 1982.

[18] C. Mandt, Y. Kuga, L. Tsang, and A. Ishimaru, "Microwave propagation and scattering in a dense distribution of spherical particles: Experiment and Theory," *Waves in Random Media*, vol. 2, no. 3, pp. 225–234, 1992.

[19] L. Tsang, C. Mandt, and K. H. Ding, "Monte Carlo simulations of extinction rate of dense media with randomly distributed spheres based on solution of Maxwell's equations," *Optics Letters*, vol. 17, no. 5, pp. 314–316, 1992.

[20] R. D. West, *Polarimetric Radar Signatures for Two Layers of Dense Random Media*, M. S. thesis, University of Washington, 1990.

[21] D. Sobajic and Y. Pao, "Artificial neural-net based dynamic security assessment for electric power system," *IEEE Trans. on Power Systems*, vol. 4, pp. 220–228, Feb. 1989.

[22] H. Mori, H. Uematsu, S. Tsuzuki, T. Sakurai, Y. Kojima, and K. Suzuki, "Identification of harmonic loads in power systems using an artificial neural network," *Proc. of 2nd Symp. on Expert Systems Applications to Power Systems*, pp. 371–377, July 1989.

[23] E. H. Chan, "Application of neural-network computing in intelligent alarm processing," *Proc. of PICA*, pp. 246–251, May 1989.

[24] H. Tanaka, S. Matsuda, H. Ogi, Y. Izui, H. Taoka, and T. Sakaguchi, "Design and evaluation of neural network for fault diagnosis," *Proc. of 2nd Symp. on Expert Systems Application to Power Systems*, pp. 378–384, July 1989.

[25] H. Mori and S. Tsuzuki, "Power system topological observability analysis using a neural network model," *Proc. of 2nd Symp. on Expert System Application to Power Systems*, pp. 385–391, July 1989.

[26] R. A. Jacobs, "Increased rates of convergence through learning rate adaptation," *Neural Network*, vol. 1, pp. 295–307, 1988.

[27] S. C. Colbeck, "The layered character of snow covers," *Reviews of Geophysics*, vol. 29, no. 1. pp. 81–96, Feb. 1991.

[28] R. West, D. Winebrenner, and L. Tsang, "Comparison of dense medium radiative transfer theory with extinction, scattering and emission data for snow," presented at the Joint IEEE/URSI meeting, International Geoscience and Remote Sensing Symposium, IGARSS' 91, (URSI 91). June 3–6, 1991, Espoo, Finland.

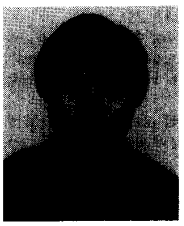
Leung Tsang (S'73–M'75–SM'85–F'90) received the B.S., M.S.E.E., and Ph.D. degrees in 1971, 1973, and 1976, respectively, from the Massachusetts Institute of Technology.

He is presently a professor of electrical engineering at the University of Washington. Between 1980 and 1983 he was with the Department of Electrical Engineering and the Remote Sensing Center at Texas A&M University. He was a research engineer with Schlumberger between 1976 and 1978. His current research interests are in microwave remote sensing, waves in random media, and solid state theory of optoelectronics. He is coauthor of the book *Theory of Microwave Sensing* (Wiley-Interscience, 9185). He is an editor of the *Journal of Electromagnetic Waves and Applications*, on the editorial board of the *Journal of Waves in Random Media*, an associate editor of *Radio Science*, and an associate editor of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.

Dr. Tsang is a member of the Electromagnetics Academy.

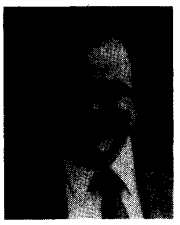
Zhengxiao Chen was born in Changsha, China, on May 7, 1968. He received the B.S.E.E. degree from the University of Southern California in 1989, and the M.S.E.E. degree from the University of Washington in 1990. He is presently pursuing the Ph.D. degree in electrical engineering at the Unversity of Washington.

His main research interests include electromagnetic scattering, microwave remote sensing, and neural networks.



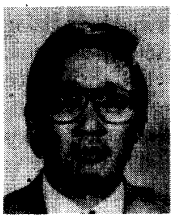
Seho Oh received the B.S. degree in electronics engineering from Seoul National University and the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology, Seoul. He received the Ph.D. degree in electrical engineering from the University of Washington, Seattle, in 1989.

From 1981 through 1986 he was in the Central Research Laboratory of Goldstar Company. His research interests are in the area of signal analysis, artificial neural networks, medical imaging, and pattern recognition. He is the author of over 40 archival and proceedings papers and has been issued two U.S. patents.



Robert J. Marks II (S'71–M'72–S'76–M'77–SM'83) is a professor in the Department of Electrical Engineering at the University of Washington, Seattle. He was awarded the Outstanding Branch Councilor Award in 1982 by IEEE and, in 1984, was presented with an IEEE Centennial Medal. He was Chair of the IEEE Neural Networks Committee and was cofounder and first Chair of the IEEE Circuits and Systems Society Technical Committee on Neural Systems and Applications. He was also first President of the IEEE Council on Neural Networks, in 1992. He was named an IEEE Distinguished Lecturer in 1992. He is cofounder and current President of Multidimensional Systems Corporation in Lynnwood, WA. He has been the Editor-in-Chief of the IEEE TRANSACTIONS ON NEURAL NETWORKS since the beginning of 1992. He was the topical editor for Optical Signal Processing and Image Science for the *Journal of the Optical Society of America* from 1989 to 1991 and a member of the editorial board for the *International Journal of Neurocomputing* from 1989 to 1992. He serves as the Organizational Chair of the IEEE-SP International Symposium on Time-Frequency and Time-Scale analysis (Victoria, B.C., 1991), the Program and Tutorials Chair for the First International Forum on Applications of Neural Networks to Power Systems (Seattle, 1991), and the General Chair of the International Symposium on Circuits and Systems (Seattle, 1995). He has published over 100 journl and proceeding papers in the areas of signal analysis, detection theory, signal recovery, optical computing, signal processing, and artificial neural processing. He is the author of the book *Introduction to Shannon Sampling and Interpolation Theory* (Springer-Verlag, 1991).

Dr. Marks is a Fellow of OSA and a Senior Member of IEEE. He was the cofounder and first President of the Puget Sound Section of the Optical Society of America and was elected that organization's first Honorary Member. He is a member of Eta Kappa Nu and Sigma Xi.



Alfred T. C. Chang (M'88–SM'91) received the B.S. degree in physics in 1964 from National Cheng Kung University, Taiwan, and the M.Sc. and Ph.D. degrees in physics, in 1969 and 1971, respectively, from the University of Maryland, College Park.

Since 1974 he has been employed by the NASA Goddard Space Flight Center, Greenbelt, MD, where he is a research scientist in the Laboratory for Hydrospheric Processes. His special areas of research are in radiative transfer, microwave radiometry, and the development of techniques for determining the properties of snow, soil, ice, and rain by remote sensing.

Dr. Chang is a member of AGU and AMS.