Atmospheric Liquid Water Retrieval Using a Gated Experts Neural Network

E. MOREAU,*,+ C. MALLET,+ S. THIRIA,* B. MABBOUX,# F. BADRAN,*,# AND C. KLAPISZ+

*Laboratoire d'Océanographie Dynamique et de Climatologie, Paris, France
+Centre d'Etude des Environnements Terrestre et Planetaires, Paris, France
#Conservatoire National des Arts et Métiers, Paris, France

(Manuscript received 28 March 2001, in final form 31 July 2001)

ABSTRACT

Gated experts (GE) neural networks have been developed in order to retrieve atmospheric liquid water content over ocean from radiometer data. Gated experts neural networks are statistical models, which can model any general class of function. This paper focuses on the case where the complex transfer functions can be split on different simpler functions in order to improve the accuracy. Two atmospheric quantities are considered: the integrated cloud liquid water (iclw) and the surface rain rate (RR). In the case of iclw, the GE neural network finds two modes, splitting the problem into low and high iclw values. The physical meaning of those modes is discussed. A comparison with a standard regression algorithm and a multilayer perceptron neural network is done on simulated data and an ''indirect comparison'' is done using Special Sensor Microwave Imager (SSM/I) data. In the case of RR, the focus is on the ability of GE neural networks to perform a classification between rainy and nonrainy situations. Tropical Rainfall Measuring Mission (TRMM) data are used for rain-rate validation: rain-rate retrieval from the GE algorithm applied to actual TRMM Microwave Imager (TMI) measurements are compared with collocated precipitation radar (PR) rain rate.

1. Introduction

Nonlinear regression is widely used in geophysics in order to propose nonlinear models relating measured quantities to significant geophysical parameters under study. In many cases, the theoretical direct model that predicts some relationship between the physical parameters and the measurements is strongly nonlinear (Thiria et al. 1993). Thus, the determination of the inverse model (from measurements to physical parameters) leads to very difficult inverse problems. Moreover, the possibility of many different direct models can lead to ambiguity in the inverse problem.

a. The geophysical problem

Atmospheric liquid water content (in cloud or rain) is an important parameter of the atmosphere. Its knowledge is essential for a wide variety of applications, such as meteorological forecasting by means of data assimilation, climate process, hydrological balance, and radio communications (among many others, Marécal et al. 2001, 2000; Hou et al. 2000). This study focused on the use of two particular spaceborne microwave radiometers, the Special Sensor Microwave Imager (SSM/I) (Hollinger et al. 1990) and the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) (Kummerow et al. 1998), that provide brightness temperatures with a very wide spatial coverage (see Table 1). The aim of this study is thus to determine adequate inverse models to compute two atmospheric contents, the integrated cloud liquid water (iclw) and the surface rain rate (RR) from the radiometer signal (brightness temperatures). In this section, after presenting the parameters we want to retrieve and the quantities we can measure with radiometers (brightness temperatures), we will briefly expose the physical basis of the use of multifrequency microwave radiometry for remote sensing.

In the atmosphere, water is present in three different phases: vapor, liquid, or ice. Liquid water is present in clouds and rain. From a physical point of view, the main difference between cloud water and rainwater is the size of the droplets. The boundary between these two atmospheric components is generally considered to be around 100 μm.

In passive microwave remote sensing, a radiometer records the radiant energy emitted naturally by a source, namely the earth's surface and atmosphere. The signal is detected by a downward-looking radiometer on board a satellite. The power of the signal is typically expressed as brightness temperature (TB) and includes information about the polarization state of the signal. The signal results from emission and scattering by rain and ice, emission by clouds, atmospheric gases (such as H_2O and O_2), and the earth's surface.

Corresponding author address: Dr. E. Moreau, CETP, 10-12 av. de l'Europe, 78140 Vélizy, France.
E-mail: emmanuel.moreau@cetp.ipsl.fr

TABLE 1. Main characteristics of the different instruments used in the study.

	Center frequency (f in GHz)						
	10.65	13.8	19.35	21.3	22.23	37.0	85.5
SSM/I							
Polarization			H, V		V	H, V	H, V
Footprint (km × km)			69 × 43		50 × 40	37 × 29	15 × 13
TMI							
Polarization	H, V		H, V	V		H, V	H, V
Footprint (km × km)	59 × 36		30 × 18	23 × 16		16 × 10	7 × 4
TRMM PR							
Horizontal resolution (km^2)		4.3 × 4.3					
Vertical resolution (m)		250					

The advantage of using the microwave portion of the spectrum is that microwave radiation penetrates clouds and interacts strongly with cloud droplets. Remote sensing is a viable approach to iclw estimation only for clear or cloudy sky. In fact, in rainy situations, the TBs are strongly affected by the rain particles and the iclw information contained in the TBs is thus not usable for the frequency considered here.

The brightness temperature is a frequency-dependent complex function of the vertical profiles of atmospheric temperature, pressure, water vapor density, cloud liquid water, and liquid and ice rain. Other parameters, such as the size of hydrometeor droplets and characteristics of the earth surface emission and reflectivity, must also be considered. The relation between these geophysical quantities and measurable radiometric brightness temperatures can be modeled by the radiative transfer equation, which is based on theoretical consideration. This equation takes into account the vertical distribution of the different involved quantities. The influence of the different quantities depends on the frequency.

The difficulties generally encountered when building retrieval algorithms are due to the complexity and the nonlinearity of this radiative transfer equation. Moreover, the presence of hidden parameters (such as water vapor content and ocean surface characteristics) increases the difficulty. Multifrequency measurements are thus necessary to take into account the contribution of the different components. For example, a channel in the water vapor absorption line (near 22 GHz) allows us to take the water vapor contribution into account (Ulaby et al. 1981). Neglecting the noise introduced by the sensor and setting the other variables that affect TB—the so-called ''hidden'' variables in estimation theory—to standard values, the five lower SSM/I channels allow an estimation of the integrated cloud liquid water in a unique way. But at these frequencies the sensitivity of TB to the hidden variables is very important, as will be shown in Fig. 1.

First, we notice that the hidden variables do not affect the TBs in a similar way. The lower frequencies are more sensitive to the variation of the ocean surface than are higher frequencies, explaining the larger dispersion of iclw at all magnitudes at 19 GHz. Thus, the 37-GHz frequency is better correlated to iclw.

Second, we notice that for both the 19- and 37-GHz frequencies, the dispersion depends on the magnitude of iclw. Two modes can be detected, one mode for the lower values of iclw with a large dispersion and one mode for the higher values of iclw, characterized by a saturation phenomena of the signal. The problem is then to be able to split the input space into several parts in

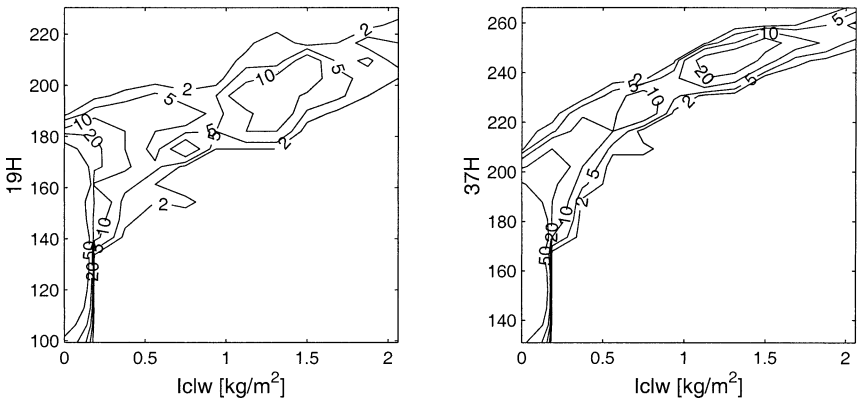

FIG. 1. Brightness temperature (TB) at 19 GHz (left) and 37 GHz (right) in the horizontal polarization vs iclw. The contour lines represent the frequency distribution of TB as a function of iclw. Units of contour lines numbering are in (K kg m^{-2})$^{-1}$. Dataset consists of 1000 points.

a way that estimates as well as possible the transfer functions for each mode.

b. Why gated experts network?

In this study, we propose a general method to solve such inverse problems by using gated experts (GE) networks. The introduction of the mixture of experts concept into the neural network community is relatively recent (Jacobs et al. 1991); the term gated experts, which stands for nonlinearly gated nonlinear experts, was proposed by Weigend et al. (1995) for a specific mixture of experts. In their paper Weigend et al. (1995) demonstrate the efficiency of GE in the context of times series prediction and analysis. The GE networks are a generalization of the well-known multilayer perceptron (MLP) (Rumelhart et al. 1986) widely used for transfer function approximation (Thiria et al. 1993). Like MLP, GE are adaptive, providing a flexible and easy way of modeling a large variety of physical phenomena. Here adaptive means that the method is able to process a large amount of data or deal with new relevant variables. Even if the calibration of the network takes a long time, its use during an operational phase is very efficient. But in addition to MLP, GE provides a more general class of functions. In fact, GE can represent any arbitrary conditional probability distribution in the same way that MLPs can represent arbitrary functions (Bishop 1995). It is particularly interesting for solving difficult inverse problems when multimodal functions are considered; in that case one looks for different possible values as possible predictions of the physical parameter. During the calibration of a GE network, a partition of the measurement space is automatically done, a dedicated MLP being adapted simultaneously on each part. In that sense, it can be said that a GE network proposed different transfer functions, and each one is dedicated to a particular subprocess. In the case of geophysical phenomena presenting different regimes, looking separately at the different subprocess defined by a GE network allows retrieval of these different regimes and study of their behavior.

In a GE network each subprocess, which is classical MLP, is named ''expert.'' And the gating network, which is a particular MLP, allows choice of the regime. So a GE network consists of a nonlinear gating network and several (also nonlinear) competing experts. Each expert is devoted to a part of the input space; it learns a local model from the data. The gating network learns to predict the probability of each expert, given the measurements, and to split the input space.

c. Strategy

In what follows, attention is mainly focused on the retrieval of integrated cloud liquid water from SSM/I in non-rain situations. In a more succinct way, the retrieval of rain rate from TMI is also shown to illustrate the capability of the GE to perform rain rate classification. For iclw retrieval with SSM/I data, numerous algorithms have been developed and validated that allow us to evaluate the performance of the GE network. For RR retrieval, the TMI presents two advantages: its spatial resolution is better than that of SSM/I for this application, and the presence of the precipitation radar (PR) on the same satellite allows an easy validation with actual data.

The aim of the present paper is to show that GE networks are able to model a large class of complex inverse nonlinear models. The method is presented in detail and can easily be extended to a large class of inverse problems. In section 2, the database used to retrieve the liquid water content (in rain or cloud) from space-based passive microwave measurements is presented. Section 3 gives an overall presentation of GE networks. Section 4 describes in detail how they have been used in this work and display the obtained results concerning iclw retrieval. A comparison with those of more traditional methods developed on the same dataset is also presented. A validation with real data is then performed. Section 5 presents the obtained results concerning RR retrieval.

2. The simulated data

Like MLPs, GE networks must be calibrated using a dedicated learning dataset. This learning dataset, which has to be carefully processed, consists of input/output pairs of multifrequency TBs and a corresponding atmospheric parameter (iclw or RR). A large amount of TB data is available in the microwave remote sensing community. However, the scarcity of in situ meteorological data concerning cloud or rain systems necessitates tackling this problem through simulated data.

a. Atmospheric profiles and radiative transfer models

A large set of atmospheric profiles and surface characteristics are used to compute iclw or RR and the corresponding TBs. The atmospheric profiles are obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) model (Tiedtke 1993). From these atmospheric profiles, TBs at the top of the atmosphere are computed through microwave radiation transfer models. These models consider the radiative transfer of polarized radiation through a horizontally infinite plane-parallel atmosphere.

In our learning dataset we keep only the atmospheric profiles obtained over the ocean and between 60°S and 60°N. For historical reasons two sets of ECMWF simulations are used. One set consist of 1.125° × 1.125° resolution, 36-h forecast experiment for 1 August 1992, 1 August 1996, and 1 March 1996. These data are used only for the iclw retrieval, and they are resampled at higher resolution (Gérard et al. 1998) to ensure simulated TBs as similar to spaceborne radiometer measured

TABLE 2. Characteristics of the database for rain-rate retrieval.

	Liquid rain rate (mm h^{-1})	Ice precipitation rate* (mm h^{-1})	Iclw (kg m^{-2})
Mean	2.74	0.78	0.66
Standard deviation	3.47	0.92	0.56
Minimum	0	0	0
Maximum	19.72	5.11	4.09
Number		2458	

* Maximum equivalent ice precipitation rate for a profile.

TABLE 3. Characteristics of the database for cloud liquid retrieval.

	Iclw [kg m^{-2}]	Water vapor [g cm^{-2}]
Mean	0.20	3.56
Standard deviation	0.46	1.64
Minimum	0.00	0.19
Maximum	2.03	7.53
Number	2610	

temperatures as possible. Altogether, these simulations provided 30 000 atmospheric profiles. Then, a radiative transfer model devoted to non-rain situations is used (Prigent et al. 1994) to compute brightness temperatures at the five lower frequencies of the SSM/I radiometer. It computes the absorption of atmospheric gases (Liebe et al. 1993), such as oxygen and water vapor, absorption of cloud liquid water, and the ocean surface emissivity (Guillou et al. 1996). The cloud liquid water is considered to be made up of Rayleigh particles with drop radii less than 100 μm.

The second set of simulations, used for the rain-rate retrieval, was a 0.5° × 0.5° resolution, 36-h ECMWF forecast experiment from 10 and 16 February 1998 and cloud model simulation as described in Viltard et al. (2000). Altogether, these simulations provided 20 000 atmospheric profiles. In rain situations, because of the presence of large droplets, a new radiative transfer model has been developed to take scattering phenomena into account (Moreau et al. 1999). This model is used to compute TBs at the seven higher frequencies of the TMI radiometer.

b. Database

In order to avoid an overrepresentation of the more frequent cases of iclw or RR, a sampling of profiles is performed on the primary database. We thus selected about 10% of the profiles, to create the dataset used to calibrate and validate the GE models. These profiles have been selected in order to represent all types of meteorological situations. For example, for cloudy situations the TBs are more sensitive to water vapor (WV) than to other parameters such as the surface temperature or the surface wind speed. In the final database, all ranges of iclw and WV are thus represented with similar proportions. This selection ensured that the data would be well uncorrelated. Tables 2 and 3 present the main characteristics of the two databases used for the development of iclw and RR algorithms.

The neural network methodology divides each simulated database into three parts named learning, validation, and test set (Bishop 1995). In this study the size of the learning dataset represents half of the total database, and each of the validation and test databases represent one-quarter of the data. The learning dataset is used for calibrating the GE. The validation dataset is used for evaluating the performances of GE during this calibration phase and to choose an optimal GE network. The test dataset is used to measure the performance of the calibrated GE for actual use; therefore, the test dataset should be completely independent of the data used during calibration (learning and validation). Using the test set makes it possible to compare with other algorithms and methods. For a more realistic performance evaluation, we present in section 5 comparison experiments using actual radiometric measurements.

3. Theory of multiexpert

In this section, we briefly introduce the theoretical background of the GE networks. As mentioned in section 1, we are interested in the general problem of estimating conditional probability density functions using GE networks. In order to investigate a particular phenomena involving two multidimensional related variables, $\mathbf{x} \in R^p$ and $\mathbf{y} \in R^q$, a possible approach is to get a relevant statistical set of pairs of observations, which composed the learning dataset, $D = \{(\mathbf{x}_i^{\text{obs}}, \mathbf{y}_i^{\text{obs}}), i = 1 \ldots N^{\text{obs}}\}$ and to study it. More often, the numerical values are approximations of actual values, and it can be assumed that the observed data $(\mathbf{x}^{\text{obs}}, \mathbf{y}^{\text{obs}})$ are realizations of random variables with probability density functions:

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y}/\mathbf{x})p(\mathbf{x}). \quad (1)$$

When the relation between $\mathbf{x}$ and $\mathbf{y}$ is multivalued, the best way to describe the underlying phenomena is to approximate the density probability function $p(\mathbf{y}/\mathbf{x})$.

The basic idea of the underlying modeling is to approximate $p(\mathbf{y}/\mathbf{x})$ by a mixture distribution of K Gaussian density probability functions whose parameters are estimated using a particular class of functions (the GE networks). The main property of Gaussian mixtures is that they can approximate any continuous density function to arbitrary accuracy provided that the mixture has a sufficiently large number K of Gaussians and provided that its parameters are set up correctly using the learning dataset D. In the following (1) is rewritten using $q = 1$, which means that the desired output is a scalar; the methodology is easily generalized to any q. The analytical form of a Gaussian mixture is

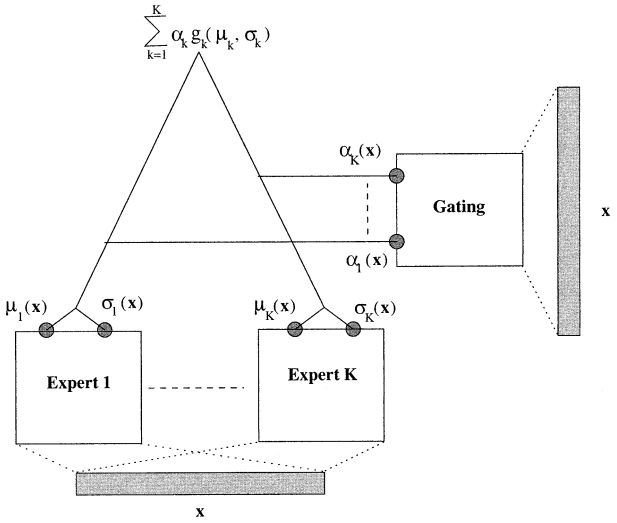

FIG. 2. Architecture of the GE, where **x** denotes the inputs vector (TBs). Each box is a nonlinear neural network that generated its own outputs: $\mu_k(\mathbf{x})$ (retrieved atmospheric parameter) and $\sigma_k(\mathbf{x})$ (associated standard deviation). The gating outputs $\alpha_k(\mathbf{x})$ weight the expert outputs $\mu_k(\mathbf{x})$ and $\sigma_k(\mathbf{x})$. The global output is the density probability function, defined by: $p(y/x) = \Sigma_{k=1}^{K} \alpha_k(x) g_k(y/x)$.

$$p(y/\mathbf{x}) = \sum_{k=1}^{K} \alpha_k(\mathbf{x}) g_k(y/\mathbf{x}), \tag{2}$$

where g_k is a Gaussian density function whose mean is $\mu_k(\mathbf{x})$ and standard deviation $\sigma_k(\mathbf{x})$:

$$g_k(y/\mathbf{x}) = \frac{1}{\sqrt{2\pi}\sigma_k(\mathbf{x})} \exp\left(-\frac{\|y - \mu_k(\mathbf{x})\|^2}{2\sigma_k^2(\mathbf{x})}\right). \tag{3}$$

In (2) the mixing coefficients $\alpha_k(\mathbf{x})$ represent the a posterior probability that the observation **x** has been generated by the distribution g_k. These coefficients are normalized: $\Sigma_{k=1}^{K} \alpha_k(\mathbf{x}) = 1$.

The problem is thus to estimate the $K^*(q + 2)$ parameters of the mixture models: $\mu_k(\mathbf{x})$, $\alpha_k(\mathbf{x})$, and $\sigma_k(\mathbf{x})$ with $k = 1 \ldots K$. This is done by maximizing the likelihood of the set of the observations D. In GE networks the $K^*(q + 2)$ parameters that maximize the likelihood are determined by using the Expectation Maximization algorithm (Jordan and Xu 1995). Let us introduce the opposite of the log-likelihood where the $K^*(q + 2)$ parameters are to be estimated (Bishop 1995):

$$E(\mu_1 \ldots \mu_K, \sigma_1 \ldots \sigma_K, \alpha_1 \ldots \alpha_K) = -\sum_{i=1}^{N^{\text{obs}}} \ln \sum_{k=1}^{K} \alpha_k(\mathbf{x}_i^{\text{obs}}) g_k(y_i^{\text{obs}}/\mathbf{x}_i^{\text{obs}}). \tag{4}$$

Maximizing the likelihood is equivalent to minimizing (4).

In the following we used a particular architecture proposed by Jordan and Jacobs (1995) (see Fig. 2). In this network one can isolate different subnetworks (named expert), linked by a gating network, which compute the mixing coefficients $\alpha_k(\mathbf{x})$. The K subnetworks are MLPs with two outputs, $\mu_k(\mathbf{x})$ and $\sigma_k(\mathbf{x})$. The gating is an MLP whose K output neurons have an exponential transfer function. Each expert k defines a function $F_k(\mathbf{x}, W_k)$ from R^p to R^q, which estimates the Gaussian function g_k, the W_k being the weights of the K different MLPs. The gating network $F_{\text{gate}}(\mathbf{x}, \mathbf{W}_{\text{gate}})$ is a function from R^p to $(R^+)^K$, which estimates the mixing coefficients $\alpha_k(\mathbf{x})$, $\mathbf{W}_{\text{gate}}$ being the weights of the gating MLP. In the experiments presented in sections 4 and 5, we assumed that the K standard deviations $\sigma_k(\mathbf{x})$ of the different Gaussians do not depend on **x** and are estimated during the learning process. The learning phase allows us to determine the weights $\mathbf{W} = \{\mathbf{W}_1, \ldots, \mathbf{W}_K, \mathbf{W}_{\text{gate}}\}$ of the different experts by minimizing the opposite of the log-likelihood $E(\mathbf{W})$. Using the experts and the gating, the cost function becomes

$$E(\mathbf{W}) = -\sum_{i=1}^{N^{\text{obs}}} \ln \sum_{k=1}^{K} F_k(\mathbf{x}, \mathbf{W}_k) F_{\text{gate}}(\mathbf{x}, \mathbf{W}_{\text{gate}}). \tag{5}$$

At the end of the learning phase, each expert becomes specialized and represents a given Gaussian. The outputs of expert k provide an estimation of the conditional mean value $\mu_k = E(y/\mathbf{x}, k)$ and σ_k; the k outputs of the gating network define the probability of choosing one expert given the observation **x**. The value $\Sigma_{k=1}^{K} \alpha_k(\mathbf{x})\mu_k(\mathbf{x})$ represents the global expected value $E(y/\mathbf{x})$. This model assumes that each expert is assigned to a specific work, which has been automatically determined during the learning process. When an observation **x** is presented, it is distributed among the different experts according to the probabilities provided by the gating network. As each expert becomes specialized, it can be used after learning for the study of the phenomena it is representing.

4. Integrated cloud liquid water retrieval

The first application concerns the retrieval of iclw from SSM/I brightness temperatures. In this section we consider the notation used in the GE general presentation (section 3): x is a vector of the five SSM/I brightness temperatures corresponding to the three lower frequencies and its different polarization (Table 1), and y designates the iclw. During the training phase, many experiments have been performed with different numbers of experts, architectures, initial weights, and learning rates. All experiments lead to the same number of experts: i.e., only two experts survived. We selected the best architecture for the expert, this with the smallest number of weights and giving the best results on the validation set. Under these considerations, our GE model is composed of two experts with two hidden layers respectively composed of three and two hidden neurons, and one gating network with two hidden layers respectively composed of three and two hidden neurons. We have found coincidentally that the selected experts and

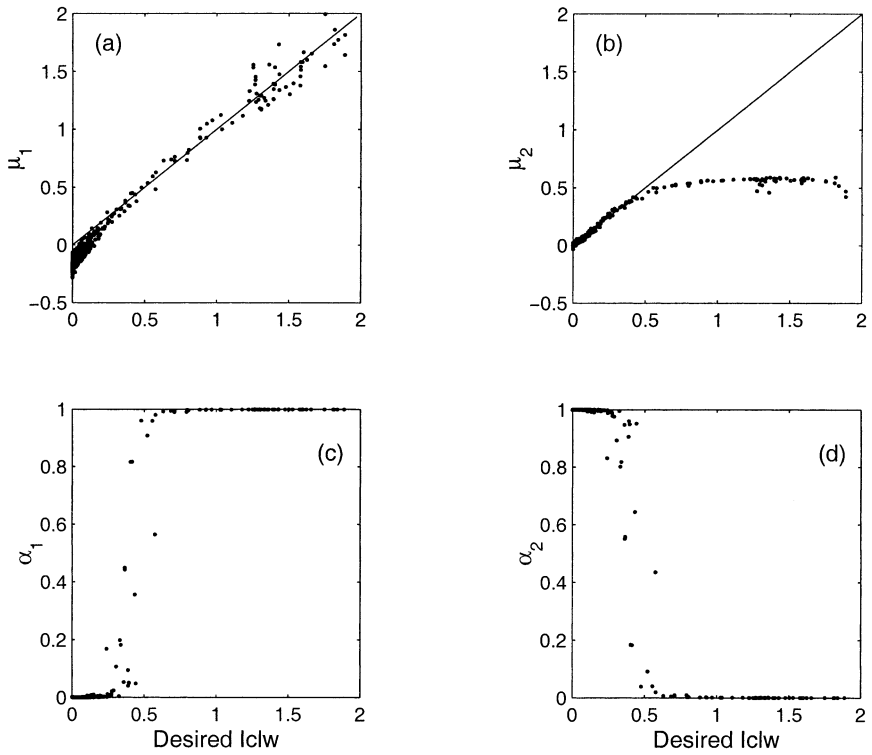

FIG. 3. (top) Output of experts (a) μ_1 and (b) μ_2 vs desired iclw values. (bottom) Mixing coefficients (c) α_1 and (d) α_2 vs desired iclw values. GE_{iclw} algorithm used on learning simulated dataset.

gating subnetworks have the same architecture. In the following this GE model is noted GE_{iclw}.

a. The split between the two experts

The gating network has one output α_k for each expert *k*. Its goal is to estimate the probability that a given input **x** was generated by expert *k*. Figure 3 shows the results obtained on the learning set. Similar results are obtained on the validation set (not shown). Figure 3a shows the output μ_1 of expert 1 and Fig. 3c the corresponding mixture coefficient α_1 vs reference, i.e. desired iclw values. Figures 3b and 3d are the same as Figs. 3a and 3c for expert 2, and Fig. 4 shows the global output of the GE_{iclw}, i.e. the sum of the contributions from experts 1 and 2.

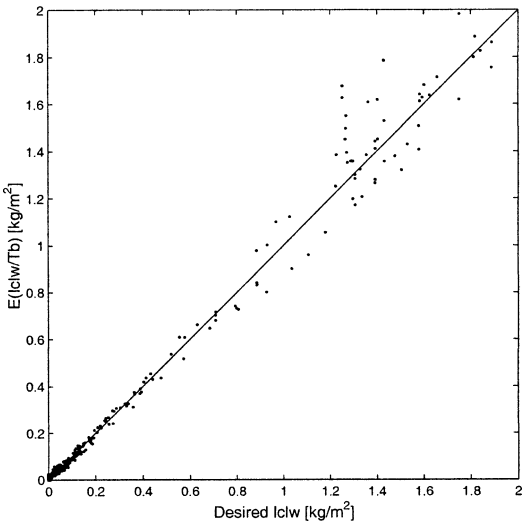

FIG. 4. Global output vs desired iclw values for the GE_{iclw} algorithm used on the test simulated dataset.

In Figs. 3a, 3b, and 4, the falling diagonal corresponds to an output equal to the desired iclw value. As shown in Fig. 4, these points follow the line quite well. Looking simultaneously at the output of the experts and at the probability given by the gating, it is clear that high iclw values are well represented only by expert 1:

$$\alpha_1 = 1, \quad \alpha_2 = 0, \quad \text{and} \quad \mu_1 = E(y/\mathbf{x}, 1)$$

$$\text{for iclw} > 0.6 \text{ kg m}^{-2}.$$

And low values are well represented by expert 2:

$$\alpha_1 = 0, \quad \alpha_2 = 1, \quad \text{and} \quad \mu_2 = E(y/\mathbf{x}, 2)$$

$$\text{for iclw} < 0.3 \text{ kg m}^{-2}.$$

The global output $E(y/\mathbf{x}) = \alpha_1\mu_1 + \alpha_2\mu_2$ is thus equal to μ_1 for high values and to μ_2 for low values. The competition is soft; the gating outputs α_k can take any value between 0 and 1, and for points close to the boundary between the two regimes both experts contribute to

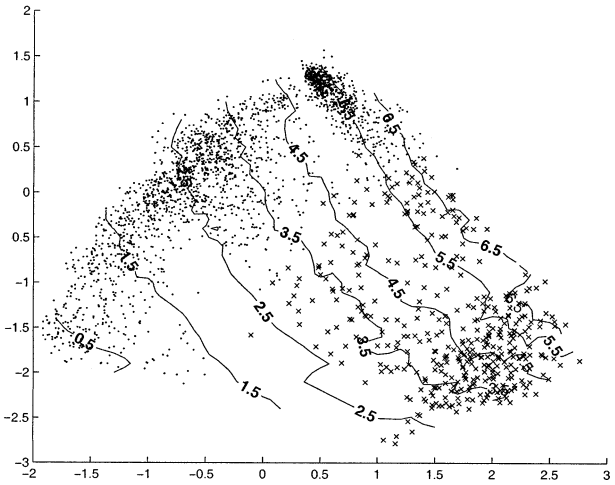

FIG. 5. Separation of experts on the validation data in the reduced input space after a principal component analysis computation. Dots represent $\alpha_1 > 0.5$ (corresponding to expert 1) and crosses represent $\alpha_2 > 0.5$ (corresponding to expert 2). The contours represent the water vapor content (g cm^{-2}).

the global output. The built-in constraint that the gating outputs sum to unity implements the competition between the experts.

Thus the partition of the input space is fashioned in a manner that for points close to the boundary between two regimes the corresponding expert outputs are very close together:

$$\alpha_1 \quad \text{and} \quad \alpha_2 \neq \{0, 1\} \quad \text{and} \quad \mu_1 \approx \mu_2$$
$$\text{for } 0.3 < \text{iclw} < 0.6 \text{ kg m}^{-2}.$$

We found that we do not have a partition in clear (iclw = 0) and cloudy (iclw $\neq$ 0) skies, but in low (iclw $<$ 0.3 kg m^{-2}) and high values (iclw $>$ 0.6 kg m^{-2}). To better understand the physical meaning of this partition, Fig. 5 shows the separation of experts in the reduced input space after principal component analysis computation. The original input space dimension is equal to five, corresponding to the five TBs. The principal component analysis reduced the input space to two components, keeping 93% of the explained variance. The principal component space also exhibits the clear differences between the two modes identified by the two experts. Contours of water vapor are also displayed on the same diagram showing that one of the modes (expert 1) covers a large range of water vapor content (from 0 to 7 kg m^{-2}), while the other mode (expert 2) covers the higher values of the water vapor content (from 3 to 7 kg m^{-2}).

The relation between the brightness temperature at 19 and 37 GHz and the iclw is unique that is there can only be one value for iclw for a given TB and vice-versa when all other atmospheric and surface parameters are fixed. Figure 5 shows that the large variation of water vapor creates a dispersion on the brightness temperature since the ''noise level'' of the brightness temperature is directly correlated to the possible variation of the water vapor. Each expert is specialized to a region characterized by a different noise level. This analysis of the split of the input space deduced from Fig. 5 is also confirmed by the standard deviation associated with each expert. For the expert devoted to the noisy dataset we obtain $\sigma_2 = 0.01$ kg m^{-2}, that is to say 29% of the relative value, and for the other one we obtain $\sigma_1 = 0.12$ kg m^{-2}, only 9% of the relative value.

TABLE 4. Algorithm performances on the validation dataset.

Algorithm	Bias (10^{-3} kg m^{-2})	Standard deviation (10^{-2} kg m^{-2})
GE_{iclw}	−0.33	4.40
MLP	−1.70	3.70
LL	−3.20	5.80

b. Comparisons with other statistical methods

The GE_{iclw} algorithm will be compared to a regression algorithm (denoted LL hereafter) and an MLP. These two algorithms are developed on the same training dataset used for the GE_{iclw} calibration. The form of the regression algorithm is $\text{iclw} = a_0 + \sum_{i=1}^{5} a_i \log[280 - x(i)]$ where $x(i)$ denotes the five lower channels of SSM/I and a_i denotes the regression coefficients (Gerard et al. 1998). The log-linear form linearizes the iclw–TB relationship. This linearization is valid only for an optically thin atmosphere. It should be mentioned that there are other regression-based algorithms that utilize combinations of only two or three SSM/I channels (e.g., Karstens et al. 1994) or a linear combination of some or all SSM/I channels (e.g., Alishouse et al. 1990).

Several MLP are tested for different hidden units. The number of inputs is fixed at five as for the GE. The best performances are obtained with six units in the first hidden layer and four units in the second hidden layer.

The comparison is made on the test data subset, which was not used during the learning phase. The bias and standard deviation of the three algorithms are summarized in Table 4. As shown, the two neural network algorithms have somewhat better performances than the log-linear one. The global root-mean-square error is equal to 0.042 kg m^{-2} for the GE_{iclw} as compared to 0.066 kg m^{-2} for the log-linear algorithm. Globally the performances of the GE_{iclw} and the MLP appear very close.

The performances of GE_{iclw}, MLP, and LL functions of iclw values are shown in Table 5. The two neural network algorithms provide a significant gain in accuracy for all range values of iclw. The relative low accuracy of LL at high iclws is due to the fact that the linearization is only valid for optically thin atmospheres.

Next, we investigate the three algorithms' performances for iclw = 0. This clear sky statistic is summarized in Table 6. For all algorithms, the bias is negligible. The standard deviation for LL is significantly

TABLE 5. Bias and standard deviation in the iclw retrievals on the test dataset containing 558 points for three algorithms: GE_{iclw}, MLP, and LL. The first (last) range of iclw corresponds to the validity domain of expert 1 (expert 2). Units: kg m^{-2}.

Range of iclw (kg m^{-2})	Bias (10^{-2} kg m^{-2})			Std dev (kg m^{-2})		
	0–0.3	0.3–0.6	0.6–2.0	0–0.3	0.3–0.6	0.6–2.0
GE_{iclw}	0.05	0.25	−0.30	0.010	0.035	0.12
MLP	−0.10	−0.40	−0.75	0.012	0.035	0.10
LL	0.30	−3.80	−3.6	0.030	0.058	0.14

TABLE 6. Algorithm performances on the test dataset for the expected clear sky value (iclw = 0 mm h^{-1}). Units: kg m^{-2}.

Algorithm	Min	Max	Avg	Std dev
GE_{iclw}	-3.3×10^{-2}	2.2×10^{-2}	1.8×10^{-3}	0.7×10^{-2}
MLP	-4.5×10^{-2}	3.1×10^{-2}	2.1×10^{-3}	1.1×10^{-2}
LL	-7.3×10^{-2}	8.5×10^{-2}	-4.5×10^{-3}	2.5×10^{-2}

higher than those of GE_{iclw} and MLP. GE_{iclw} shows the lowest minimum and maximum values and the lowest standard deviation of 0.7×10^{-2} kg m^{-2}. The differences between GE_{iclw} and MLP are small. This comparison shows that the GE_{iclw} algorithm gives somewhat better results than the LL algorithm. These results are confirmed by those of Jung et al. (1988), which show that MLP algorithms outperform regression algorithms for iclw retrieval.

In this application, we found that using GE_{iclw} algorithms rather than MLP algorithms did not increase accuracy. The main advantage of GE_{iclw} is its capacity to split a problem into several simpler subproblems, and in providing a physical meaning of each mode. With classical neural networks such as MLP, the link between the neural function and the physical problem is often very hard to find. With GE_{iclw}, this information is partially present in the physical meaning of each mode. The validation of the methodology presented in this paper using SSM/I TBs is presented in the next section.

c. Performances comparison on actual data

Real validation of iclw retrieval algorithms is very difficult to perform. The main reason is the difficulty of measuring this atmospheric parameter in situ. Therefore, we have applied the GE_{iclw} to actual SSM/I brightness temperatures to compare these performances with other more classical algorithms developed for the SSM/I radiometer. The main point we have validated concerns the retrieval of the zero value of iclw. First, because of the lack of in situ measurement a real validation is impossible; however, we can assume that the more probable iclw value in the atmosphere is zero and thus use this assumption to compare the different algorithms. Second, because of the influence of water vapor and surface emissivity in clear sky conditions, the zero value is particularly difficult to measure. Finally, if an algorithm presents many errors for the zero value, we can suppose it will also present many uncertainties for all low values. Some algorithms produce negative values of iclw. The simplest way to correct these errors is to suppose that these negative values correspond to zero values; however, we can conclude that the direct model or the inverse model presents an error near zero and thus we can suspect errors for all low values. The comparison is performed using SSM/I data over the Indian Ocean on 23 September 1995. Figure 6 presents the histogram of iclw retrieved by the three algorithms. The GE_{iclw} and the MLP both present a probability density function with a maximum at the zero value, then the probability density function decreases as the iclw increases. This shape seems more realistic than those exhibited by the LL algorithm (Fig. 6c), where an even probability density function is found from 0 to 0.1 kg m^{-2}. Applied to real data, all the three algorithms produced few negative values.

5. Application to RR retrieval

This second application concerns the retrieval of RR from TMI brightness temperatures. The input vector, **x**, is now a vector of the seven TMI brightness temperatures corresponding to the four higher frequencies (Table 1) and the output scalar, *y*, designates the RR. The

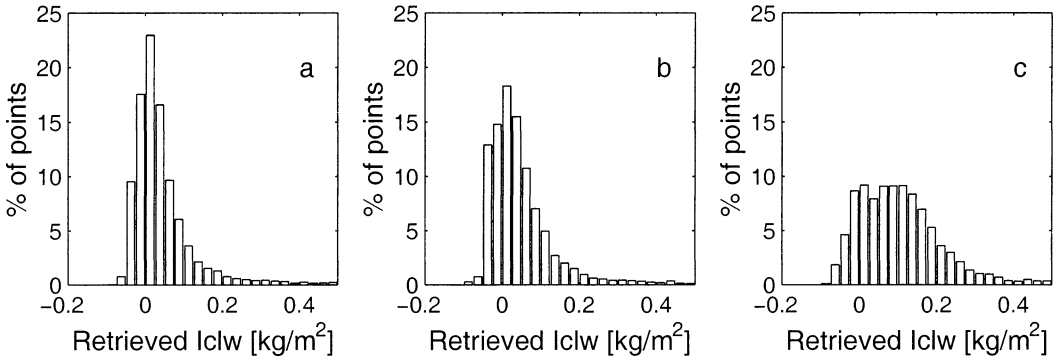

FIG. 6. Normalized frequency distribution of the retrieved iclw from actual SSM/I measurements taken over the Indian Ocean for different algorithms: (a) GE_{iclw}, (b) MLP, and (c) LL.

TABLE 7. Rain-rate retrieval algorithm performances on TRMM data.

		Number of points	Mean/max (mm h^{-1})		GE_{RR}–RR at 19 (mm h^{-1})		2A12–RR at 37 (mm h^{-1})	
			RR at 19	RR at 37	Bias	Std dev	Bias	Std dev
Global	All (RR $\geq$ 0)	9616	1.4/31	1.4/40	0.04	1.40	1.55	5.50
	Rain (RR $>$ 0.2)	3656	3.6/31	3.5/40	0.06	2.30	4.00	9.30
Homogeneous	All (RR $\geq$ 0)	2944	1.7/18	1.7/22	0.46	1.45	2.17	7.14
Stratiform	Rain (RR $>$ 0.2)	940	5.5/18	5.5/22	1.39	2.30	6.73	11.35

10-GHz channel is not used because the plan parallel hypothesis used in the method is incompatible with its very low resolution. Using the same methodology described before, the optimal GE configuration has two experts. Each experts network is composed of two hidden layers with respectively four and two neurons. The gating network is composed of two hidden layers with respectively six and four neurons.

After the learning phase one expert was found to specialize in non-rain conditions (RR = 0) and the other expert to rain situations (RR $>$ 0). For RR retrieval, compared to iclw retrieval, we have the opportunity to perform a extensive validation with real data. That is why the results obtained during the training and test phases on simulated data are not presented here. On board the TRMM satellite, the first spaceborne radar devoted to rain measurement, a PR is also present; thus an abundance of TMI/PR collocated data are available.

We first compare the performances the GE_{RR} model with the standard TMI algorithm (2A12). The 2A12 algorithm [also known as the Gprof algorithm (Olson et al. 1996; Kummerow et al. 1996, 2000)] being used for TMI makes use of the Bayes theorem to relate the observed TBs to the rain rate provided by an a priori database. The rain retrieved from the PR, as described by Ferreira and Amayenc (1999), is taken as the reference. In a second part we show the ability of the GE_{RR} network to work as a classifier to obtain a flag between rain and non-rain situations.

a. Validation data

The validation is performed on four case studies located on the Pacific Ocean, three in August and one in February 1998. The February 1998 case corresponds to a particular situation with heavy stratiform precipitation; the one measured in August 1998 corresponds to low precipitation; and the two others correspond to convective areas with lots of small rain structures.

The difference in spatial resolution between PR and TMI data (see Table 1) requires averaging the RR radar to allow a representative comparison. The 2A12 algorithm resolution is the same as the resolution of the 37-GHz channel (Kummerow et al. 1998). The GE_{RR} algorithm is developed using data simulated with the assumption of horizontally homogenateous atmosphere, and no particular resolution is considered. By default, the resolution corresponds to the lower-frequency channel (19 GHz) resolution used by the GE_{RR} algorithm. The radar RR is thus average at the proper resolution. We denote RR@19 (RR@37) as the radar retrieval RR averaged at the 19-GHz resolution (37-GHz resolution). The radar RR standard deviation is also computed in order to select a horizontally homogeneous situations.

b. Algorithm performances

The performance of the two algorithms (GE_{RR} and 2A12) relative to the radar RR are summarized in Table 7. We observe very good agreement between GE_{RR} and radar precipitation estimation. Only homogeneous and stratiform situations have been simulated because we are not able to simulate other situations with our radiative transfer model (see section 2a). For these situations (bottom of Table 7) RR@19 is about equal to RR@37. The results obtained with all rain-rate values are very good because a lot of points that correspond to non-rain situations (RR = 0) and negligible errors (see next section) are included in this dataset. When we focused on rain situations we observed a negligible overestimation of 1.39 mm h^{-1} and a standard deviation of 2.30 mm h^{-1} that corresponds to a great improvement relative to the TRMM 2A12 algorithm (6.73 mm h^{-1} and 11.35 mm h^{-1}, respectively).

The global results appear to be better. This is due to a compensating effect between the positive bias in homogeneous cases and small negative bias in heterogeneous situations (not shown in Table 7). The resulting bias is closer to zero. The GE_{RR} algorithm not only performs very well, the value provided by the gated expert allows use of GE_{RR} as a rain detector, as explained in the following section.

c. GE as classifier

As one expert is devoted to non-rain conditions (RR = 0) and the second to rain situations (RR $>$ 0), the gated expert acts as a rain detector. The output of the gating network can be interpreted as a flag for the presence of rain. The corresponding classification error matrix is given in Table 8, to allow a comparison between TMI and PR rain detection. The percentage of good detection is very high (36.1% + 57.5% = 93.6%). Bad detection occurs in only 6.4% (4.6% + 1.8%) of the cases. A pixel according to the radar, is designated non-rain when RR@19 is lower than 0.2 mm h^{-1}. Various

TABLE 8. Classification error matrix between rain and non-rain detection with TRMM data.

		RR at 19	
		Rain (RR at 19 $\geq$ 0.2 mm h^{-1})	No rain (RR at 19 $<$ 0.2 mm h^{-1})
GE_{RR}	Rain ($\alpha_1 \geq 0.9$)	36.1%	4.6%
	No rain ($\alpha_1 < 0.9$)	1.8%	57.5%

non-rain rain-rate limits have been tested in the range 0.1–0.5 mm h^{-1}. The percentage of pixels classified as non-rain by the radar and as rain by the GE_{RR} increases relative to the rain-rate limit selected. In the same way, the percentage of pixels classified as rain by the radar and as non-rain by the GE_{RR} decreases. Globally no significant differences occur relative to the percentage of good and bad detection.

6. Conclusions

The aim of this paper was to show that GE neural networks are able to improve the integrated cloud liquid water and the surface rain-rate retrieval over ocean with respect to classical algorithms. The problem is to infer atmospheric quantities from spaceborne microwave data. The transfer function from the radiometer measurements (TB) to the atmospheric water is a very complex nonlinear function. The several modes of the transfer function are found by the gating network. Gated experts neural networks improve the accuracy of the computation by allowing the computations to be done by the adequate expert. Gated experts constitute a powerful tool for modeling a large class of complex transfer functions. In fact, the GE are able to solve many problems encountered in retrievals of physics properties. A major advantage of GE is their ability to extract information in a noisy environment, whatever the noise. We show that the GE_{iclw} algorithm has two experts and that each one is adapted to the noisy level of the data. More improvements can be done, using GE neural networks in which the K standard deviations $\sigma_k(\mathbf{x})$ depend on the input $\mathbf{x}$. Those new algorithms are currently under development.

Gated experts neural networks are also able to perform classification. The GE_{RR} algorithm, in addition to improving rain-rate retrieval performances, allows classification of rain and non-rain situations. A comparison with the rain detection of the PR shows a very high percentage of good detection (93.6%). Gated experts neural networks, like classical neural networks, present many other advantages. It is very easy to take into account new parameters even if their dependence cannot be established in an analytical form. In these examples, it would be easy to investigate how sensitive the solution is to the introduction of additional variables, such as the neighbor radiometric pixel, in order to take into account the spatial context.

Acknowledgments. The authors would like to acknowledge Dr. Viltard for valuable discussions and for providing TRMM data. Dr. Crepon is acknowledged for helpful comments and useful discussions. We also thank three anonymous reviewers for their comments.

REFERENCES

Alishouse, J. C., S. A. Snyder, J. Vongsathorn, and R. R. Ferraro, 1990: Determination of oceanic total precipitable water from the SSM/I. *IEEE Trans. Geosci. Remote Sens.*, **28,** 811–816.

Bishop, C. M., 1995: *Neural Networks for Pattern Recognition.* Oxford University Press, 481 pp.

Ferreira, F., and P. Amayenc, 1999: Impact of adjusting rain relations of rain profiling estimates from the TRMM precipitation. Preprints, *29th Conf. on Radar Meteorology,* Montreal, Quebec, Canada, Amer. Meteor. Soc., 643–646.

Gérard, E., and L. Eymard, 1998: Remote sensing of integrated cloud liquid water: Development of algorithms and quality control. *Radio Sci.,* **33,** 433–447.

Guillou, C., S. J. English, C. Prigent, and D. C. Jones, 1996: Passive microwave airborne measurements of the sea surface response at 89 and 157 GHz. *J. Geophys. Res.,* **101** (C2), 3775–3788.

Hollinger, J. P., J. L. Pierce, and G. A. Poe, 1990: SSM/I instrument evaluation. *IEEE Trans. Geosci. Remote Sens.*, **28,** 781–789.

Hou, A. Y., D. V. Ledvina, A. M. da Silva, S. Q. Zhang, J. Joiner, R. M. Atlas, G. J. Huffman, and C. D. Kummerow, 2000: Assimilation of SSM/I-derived surface rainfall and total precipitable water for improving the GEOS analysis for climate studies. *Mon. Wea. Rev.,* **128,** 509–537.

Jacobs, R. A., M. I. Jordan, S. J. Nowlan, and G. E. Hinton, 1991: Adaptive mixtures of local experts. *Neural Comput.,* **3,** 79–87.

Jordan, M. I., and R. A. Jacobs, 1995: Hierarchical mixtures of experts and EM algorithm. *Neural Comput.,* **6,** 181–214.

——, and L. Xu, 1995: Convergence results for the Em approach to mixtures of experts, architectures. *Neural Networks,* **8,** 1409–1431.

Jung, T., E. Ruprecht, and F. Wagner, 1998: Determination of cloud liquid water path over the oceans from Special Sensor Microwave/Imager (SSM/I) data using neural networks. *J. Appl. Meteor.,* **37,** 832–844.

Karstens, U., C. Simmer, and E. Ruprecht, 1994: Remote sensing of cloud liquid water. *Meteor. Atmos. Phys.,* **54,** 157–171.

Kummerow, C., B. S. Olson, and L. Giglio, 1996: A simplified scheme for obtaining precipitation and vertical hydrometeor profiles from passive microwave sensors. *IEEE Trans. Geosci. Remote Sens.,* **34,** 1213–1232.

——, W. Barnes, T. Kozu, J. Shiue, and J. Simpson, 1998: The Tropical Rainfall Measuring Mission (TRMM) sensor package. *J. Atmos. Oceanic Technol.,* **15,** 809–817.

——, and Coauthors, 2000: The status of the Tropical Rainfall Measuring Mission (TRMM) after two years in orbit. *J. Appl. Meteor.,* **39,** 1965–1982.

Liebe, H. J., G. A. Hufford, and M. G. Cotton, 1993: Propagation modeling of moist air and suspended water/ice particles below 1000 GHz. *Proc. AGARD 52d Specialists' Meeting of Panel on Electromagnetic Wave Propagation,* Palma de Mallorca, Spain, AGARD, 3-1–3-10.

Marécal, V., and J. F. Mahfouf, 2000: Variational retrieval of temperature and humidity profiles from TRMM precipitation data. *Mon. Wea. Rev.,* **128,** 3853–3866.

——, E. Gérard, J. F. Mahfouf, and P. Bauer, 2001: The comparative impact of the assimilation of SSM/I and TMI brightness temperatures in the ECMWF 4D-Var system. *Quart. J. Roy. Meteor. Soc.,* **127,** 573, 1123–1143.

Moreau, E., C. Mallet, and C. Klapisz, 1999: Effects of aspherical ice and liquid hydrometeors on microwave brightness temperatures. *Microwave Radiometry and Remote Sensing of the*

Earth's Surface and Atmosphere, P. Pampaloni and S. Paloscia, Eds., VSP, 291–298.

Olson, W. S., C. D. Kummerow, G. M. Heymsfield, and L. Giglio, 1996: A method for combined passive-active microwave retrievals of clouds and precipitation profiles. *J. Appl. Meteor.*, **35,** 1763–1789.

Prigent, C., A. Sand, C. Klapisz, and Y. Lemaitre, 1994: Physical retrieval of liquid water contents in a North Atlantic cyclone using SSM/I data. *Quart. J. Roy. Meteor. Soc.*, **120,** 1179–1207.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams, 1986: *Parallel Distributed Processing.* Vol. 1. MIT Press, 547 pp.

Thiria, S., C. Mejia, F. Badran, and M. Crepon, 1993: A neural network for modeling nonlinear transfer functions: Application for wind retrieval from spaceborne scatterometer data. *J. Geophys. Res.*, **98,** 22 827–22 841.

Tiedtke, M., 1993: Representation of clouds in large-scale models. *Mon. Wea. Rev.*, **121,** 3040–3061.

Ulaby, F. T., R. K. Moore, and A. K. Fung, 1981: *Fundamentals and Radiometry*. Vol. 1, *Microwave Remote Sensing: Active and Passive*, Artech House, 456 pp.

Viltard, N., C. Kummerow, W. S. Olson, and Y. Hong, 2000: Combined use of the radar and the radiometer of TRMM to estimate the influence of drop size distribution on the rain retrieval. *J. Appl. Meteor.*, **39,** 2103–2114.

Weigend, A. S., M. Mangeas, and A. N. Srivastava, 1995: Nonlinear gated experts for time series: Discovering regimes and avoiding overfitting. *Int. J. Neural Syst.*, **6,** 373–399.