JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 106, NO. D14, PAGES 14,887–14,907, JULY 27, 2001

A new neural network approach including first guess for retrieval of atmospheric water vapor, cloud liquid water path, surface temperature, and emissivities over land from satellite microwave observations

F. Aires and C. Prigent

Department of Applied Physics, Columbia University, NASA Goddard Institute for Space Studies New York

W. B. Rossow

NASA Goddard Institute for Space Studies, New York

M. Rothstein

Science Systems and Application, Inc., NASA Goddard Institute for Space Studies, New York

Abstract. The analysis of microwave observations over land to determine atmospheric and surface parameters is still limited due to the complexity of the inverse problem. Neural network techniques have already proved successful as the basis of efficient retrieval methods for nonlinear cases; however, first guess estimates, which are used in variational assimilation methods to avoid problems of solution nonuniqueness or other forms of solution irregularity, have up to now not been used with neural network methods. In this study, a neural network approach is developed that uses a first guess. Conceptual bridges are established between the neural network and variational assimilation methods. The new neural method retrieves the surface skin temperature, the integrated water vapor content, the cloud liquid water path and the microwave surface emissivities between 19 and 85 GHz over land from Special Sensor Microwave Imager observations. The retrieval, in parallel, of all these quantities improves the results for consistancy reasons. A database to train the neural network is calculated with a radiative transfer model and a global collection of coincident surface and atmospheric parameters extracted from the National Center for Environmental Prediction reanalysis, from the International Satellite Cloud Climatology Project data, and from microwave emissivity atlases previously calculated. The results of the neural network inversion are very encouraging. The theoretical RMS error of the surface temperature retrieval over the globe is 1.3 K in clear-sky conditions and 1.6 K in cloudy scenes. Water vapor is retrieved with a theoretical RMS error of 3.8 kg m^{-2} in clear conditions and 4.9 kg m^{-2} in cloudy situations. The theoretical RMS error in cloud liquid water path is 0.08 kg m^{-2}. The surface emissivities are retrieved with an accuracy of better than 0.008 in clear conditions and 0.010 in cloudy conditions. Microwave land surface temperature retrieval presents a very attractive complement to the infrared estimates in cloudy areas: time record of land surface temperature will be produced.

1. Introduction

Even after 20 years of global microwave satellite observations, the use of microwave data over land for the retrieval of atmospheric and surface parameters is still very limited. While the ocean surface has a low microwave emissivity ~0.5 that produces good contrast of atmospheric phenomena against a low brightness temperature background, the land surface emissivities are usually close to unity, making atmospheric features much more difficult to identify against a higher brightness temperature background. In addition, the land surface emissivities are not only variable in space and time but also very complex to model since they are modulated by vegetation, topography, flooding, and snow, among other factors. Until recently, no estimates of the

Paper number 2001JD900085.
0148-0227/01/2001JD900085$09.00

microwave land emissivities over the globe were available at a spatial resolution compatible with satellite observations.

Only a few efforts have been directed toward the estimation of atmospheric parameters and surface temperature from microwave observations over land. *Jones and Vonder Haar* [1990] proposed a method to retrieve cloud liquid water path, with a more recent development by *Greenwald et al.* [1997], using Special Sensor Microwave Imager (SSM/I) observations. They estimate the surface emissivity with the help of collocated visible and infrared satellite observations to determine clear-sky conditions before analyzing cloudy scenes, but they only present results from a few cases over the central Unated StatesS. *Njoku* [1995] concluded from simulations that land surface temperature could be retrieved from multichannel microwave observations with an accuracy of 2 to 2.5 K. *MacFarland et al.* [1990] investigated the correlation between SSM/I observations and surface air temperature, and *Basist et al.* [1998] proposed a method to retrieve near-surface air temperatures from SSM/I.

Microwave land surface emissivities over the globe have been recently estimated from SSM/I observations by removing the contributions of the atmosphere, clouds, and rain using ancillary satellite data [*Prigent et al.*, 1997, 1998]. The correspondences between the geographical patterns and seasonal variations of the estimated microwave emissivities are compatible with geographic variations of large-scale topography, vegetation type, flooding, and snow cover extent. The standard deviations of the day-to-day variations of the retrieved emissivities within a month are typically about 0.012 for all the SSM/I frequencies, which is an estimate of the precision of these estimates. Assuming that these emissivities are constant over a month, a nonlinear iterative variational assimilation was developed to retrieve simultaneously the surface and atmospheric parameters (surface temperature Ts, integrated water vapor WV, and cloud liquid water path LWP) over land from SSM/I observations [*Prigent and Rossow*, 1999]. The theoretical estimate of the error of the surface temperature retrieval has a mean standard deviation of 1.6 K, does not strongly depend on surface type, and is not very sensitive to the presence of thin clouds. The sensitivity of SSM/I to water vapor is very low, except in the most arid areas where the microwave surface emissivities are low for the horizontal polarization; so the results do not improve on the first guess values. With an estimated accuracy of $\sim$0.1 kg m^{-2}, the SSM/I retrieval does not properly characterize the thinner clouds (the majority), but the cloud structures with higher liquid water content are well delineated.

A further improvement in this variational assimilation scheme could be obtained by also retrieving the seven surface emissivities as they undergo small day-today changes induced by variations of the soil moisture, the vegetation density, or the snow cover. However, in this case, 10 variables would have to be retrieved (Ts, WV, LWP plus the seven emissivities Em_i, where i represents the seven channels of SSM/I: 19 GHz V, 19 GHz H, 22 GHz V, 37 GHz V, 37 GHz H, 85 GHz V, and 85 GHz H, where V is for vertical polarization and H is for horizontal polarization) from the seven SSM/I brightness temperatures and additional information would be needed to solve the problem. The monthly mean emissivity values previously computed could be used as first guess (or, using more specifically the variational assimilation formalism, the background) estimates of the surface emissivity and the first guess matrix of error covariances could be calculated. There are several options: The covariance matrix could be calculated globally for a given month, estimated for a given type of surface, or even calculated for each single pixel considering all the monthly mean emissivities for this pixel. The inversion scheme would then rely very heavily on the representativeness of such covariance matrices, giving an important weight to the statistical description of the emissivity relationships. Given this difficulty with the retrieval of the surface emissivities with a variational assimilation method, another inversion approach is considered.

Neural network techniques have already proved very successful in the development of computationally efficient inversion methods for satellite data and for geophysical applications [*Escobar et al.*, 1993; *Aires et al.*, 1998; *Chevallier et al.*, 2000]. They are well adapted to solve nonlinear problems and are especially designed to capitalize on the inherent statistical relationships among the retrieved parameters. Such an approach has been used for retrieving columnar water vapor and liquid water or wind speed using SSM/I observations [*Stogryn et al.*, 1994; *Krasnopolsky et al.*, 1995; *Krasnopolsky et al.*, 2000]. In these works, the problem is better defined than over land. Note that variational assimilation techniques, as usually implemented, do not account for correlations among the retrieved parameters but between the first guess error of variables. However, for many ill-conditioned problems, the use of a first guess estimate is very important to regularize the inversion process, and the first guess error covariance matrix is also essential in three-dimensional/four-dimensional (3-D/4-D) variational assimilation schemes since it controls the impact of the measurements on the retrieved parameters [*Thépaut et al.*, 1993]. Up to now, neural network techniques have not used such a priori information (i.e., a specific state-dependent first guess estimate), which was a major handicap of this technique compared to the classical variational assimilation method.

In this study, a neural network approach is developed that includes the use of a first guess to retrieve the surface skin temperature Ts, the integrated water vapor content WV, the cloud liquid water path LWP, and the microwave land surface emissivities Em_i between 19 and 85 GHz from SSM/I observations. Section 2 shows that the neural network with first guess and vari-

ational assimilation approaches share important theoretical concepts and highlights some of the technical differences. A simulated database is carefully designed to train and test the neural network with special attention to its statistical representativeness on a global basis (section 3). It is derived from a global collection of coincident surface and atmospheric parameters extracted from the National Center for Environmental Prediction (NCEP) reanalysis, from the International Satellite Cloud Climatology Project (ISCCP) data [*Rossow and Schiffer*, 1991], and from the microwave emissivity atlases previously calculated. Results are presented in section 4, and section 5 concludes this study, highlighting the merits of the neural network inversion technique.

2. Neural Network and Variational Assimilation Techniques

2.1. Inverse Problems

Let y be a physical forward-model output (the radiative transfer function in the atmosphere for the following application) so that

$$y^o = y(x) + \eta, \tag{1}$$

where x are the physical variables (Ts, WV, LWP, and the seven Em_i in this study), y^o are the observations (seven brightness temperatures TB observed by SSM/I in this study), and η are the observation (or model) uncertainties (instrumental noise on SSM/I in this study). Note here that y and x are vectors representing multivariate observations and multiparameter data. The inverse problem consists in retrieving the physical variables x given the observation y^o. There exist two main approaches to solve this problem.

In the first one called "localized" inversion, an inverse process is used for each observation to find an estimate $\hat{x}$ of the physical variables x by minimizing a distance

$$D(y(\hat{x}), y^o). \tag{2}$$

This distance is dependent on the a priori information available on the probability distribution functions of the variables involved. If the observations y^o are assumed to be Gaussian distributed with zero mean and without other a priori information, the Mahalanobis distance [*Crone and Crosby*, 1995] is optimal

$$\frac{1}{2}[y(\hat{x}) - y^o]^t < y \cdot y^t >^{-1} [y(\hat{x}) - y^o], \tag{3}$$

where $< y \cdot y^t >$ is the covariance matrix of the observable quantities without measurement noise y. This is preferable than using noisy observations because in that way we compare two quantities $y(\hat{x})$ and y^0 more directly in their structure, that is, less sensitive to the noise which pollutes the comparison. This procedure has to be applied to find an optimum solution for each observation separately and can require significant computational resources.

The second approach consists in estimating a transfer function g_W, with parameters W, that is a global model for y^{-1}. The parameters W are the results of the minimization of a cost function

$$\int D(\hat{x}, x) P(x, \eta), \tag{4}$$

where $\hat{x} = g_W(y^o) = g_W(y(x) + \eta)$ and P is the joint probability distribution of the physical variables x and the noise η. The distance $D(\hat{x}, x)$ is integrated over the physical states and over the observation noise, so that the transfer function g_W is optimized globally over the range of x and the noise. In practice, to minimize the previous criterion, a database is created, composed of a statistically representative sample of coincident variables x and observations y^o and the estimation of the parameters W is made once and for all using this data set. These schemes are called "global" inversions. After this preliminary step for the estimation of W, the inversion of an observation is very fast since it involves only the direct use of the transfer function g_W.

The distances used for localized and global inversion schemes involves different variables. The first one works on the brightness temperature space, the second one on physical variable space. The optimum solution in (4) gives an estimation $\hat{x}$ that is close to the true solution x while the distance in (2) specifies that the brightness temperatures $y(\hat{x})$, associated with the estimated solution $\hat{x}$, are close to the brightness temperatures $y(x)$, associated to the real solution x.

Inverse problems are often ill-posed since the existence and the uniqueness or the stability of the solution is not always known [*Vapnik*, 1997]. This is especially the case when the "forward" model $y(x)$ is not linear; in our case the radiative transfer is not linear. To regularize the inversion process, all a priori information available should be used to constrain the solution, the inverse model, or the noise model. For example, a first guess x^b for the solution can be used, a "virtual" measurement in contrast to the observations y^o which are "direct" measurements. To our knowledge, this has not been tried in neural network inversion schemes.

In this study, a localized technique, the variational assimilation approach, and a global technique, the neural network approach, are theoretically compared. Both can solve nonlinear problems. Notations are summarized in the notation list. The variational assimilation is described in Appendix A, using the same formalism and notation in order to facilitate the comparison with the neural network technique.

2.2. New Neural Network Inversion Scheme Using First Guess Information

2.2.1. Nonlinear model: The multilayer perceptron neural network. The multilayer perceptron (MLP) network is a nonlinear mapping model composed of parallel processors called "neurons." These processors are organized in distinct layers: The first

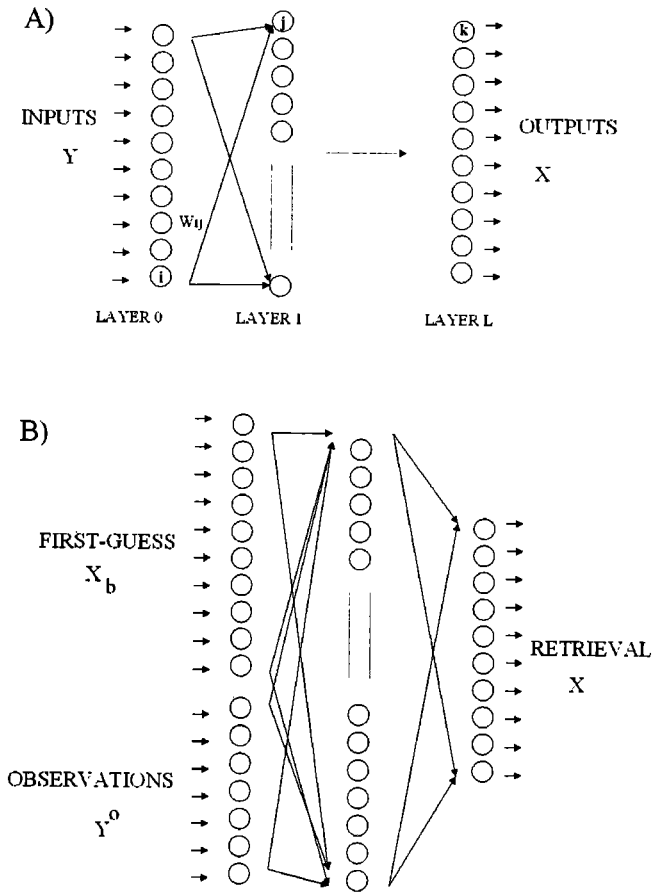

Figure 1. Neural network architectures: (a) classical MLP, and (b) MLP with first guess. Variable y^o is the observation, and x_b is the first guess for x, the retrieved variables.

layer S_0 represents the input $Y = (y_i \ ; \ i \in S_0)$ of the mapping. The last layer S_L represents the output mapping $X = (x_k \ ; \ k \in S_L)$. The intermediate layers S_m $(0 < m < L)$ are called the "hidden layers." These layers are connected via neural links (Figure 1): Any neurons, i and j, in two consecutive layers are connected with a synaptic weight w_{ij}.

Each neuron j executes two simple operations. First, it makes a weighted sum of all of its inputs z_i: This signal is called the activity of the neuron

$$a_j = \sum_{i \in \text{Inputs}(j)} w_{ij} \cdot z_i. \tag{5}$$

Then it transfers this signal to its output through a so-called "activation function," often a sigmoid function such as $\sigma(a) = \tanh(a)$. The output z_j of neuron j in the hidden layer is then given by

$$z_j = \sigma(a_j) = \sigma\left(\sum_{i \in \text{Inputs}(j)} w_{ij} \ z_i\right). \tag{6}$$

Generally, for regression problems, the output units have no activation function. For example, in a one-hidden-layer MLP, the kth output x_k of the network is defined as

$$x_k(Y) = \sum_{j \in S_1} w_{jk} \ \sigma(a_j) = \sum_{j \in S_1} w_{jk} \ \sigma\left(\sum_{i \in S_0} w_{ij} \ y_i\right). \tag{7}$$

This equation is the only computation required in the operational mode (once the synaptic weights have been determined by the training procedure). A bias term for each neuron has been deliberately omitted to simplify the notation, although if it is used in the neural network. It has been demonstrated [*Hornik et al.*, 1989; *Cybenko*, 1989] that any continuous function can be represented by a one-hidden-layer MLP with sigmoid functions σ.

2.2.2. Optimization algorithm: Backpropagation of errors. Given a neural architecture (functions used as activation functions σ, number of layers, neurons, and connections), all the information of the network is contained in the set of all synaptic weights $W = \{w_{ij}\}$. The learning algorithm is an optimization technique that estimates the optimal network parameters W by minimizing a cost function $C_1(W)$, approaching as closely as possible the desired function. The criterion usually used to derive W is the mean square error in network outputs

$$C_1(W) = \frac{1}{2} \sum_{k \in S_2} \int \int D_E(\hat{x}_k(Y; W), x_k)^2 \ P(Y, x_k) \ dx_k dY, \tag{8}$$

where D_E is the Euclidean distance between x_k, the kth desired output component, and $\hat{x}_k$, the kth neural network output component, and S_2 is the output layer of the neural network. Other contrast measures can be used if a priori information is available. $P(Y, x_k)$ is the joint probability distribution function of Y and x_k. This criterion is just the integrated distance between $\hat{x}$ and x introduced in (4).

In practice, the probability distribution function $P(Y, x_k)$ is sampled in a data set $\mathcal{B} = \{(Y^e, {x_k}^e), e = 1, \ldots, N\}$ of N input/output couples, and $C_1(W)$ is then approximated by the classical least squares criterion:

$$\tilde{C}_1(W) = \frac{1}{2N} \sum_{e=1}^{N} \sum_{k \in S_2} D_E(\hat{x}_k(Y^e; W), {x_k}^e)^2. \tag{9}$$

The error back-propagation algorithm [*Rumelhart et al.*, 1986] is used to minimize $\tilde{C}_1(W)$. It is a gradient descent algorithm that is very well adapted to the MLP hierarchical architecture because the computational cost is linearly related to the number of parameters. Traditional gradient descent algorithms use all the samples of the data set $\mathcal{B}$ to compute a mean Jacobian of the criterion $\tilde{C}_1(W)$ in (9). These algorithms are called deterministic gradient descent. The major inconvenience of this approach is that the descent can be trapped in local

minima. In the present application, a stochastic gradient descent algorithm is adopted: It uses the gradient descent formula iteratively for a unique random sample in the data set. With some constraints not discussed here, the stochastic character of this optimization algorithm theoretically allows the optimization technique to reach the global minimum of the criterion instead of a local minimum [*Duflo*, 1996].

2.2.3. Introduction of the first guess. When an inverse problem is ill-posed, the solution can be nonunique and/or unstable. The use of a priori first guess information is important to suppress ambiguities: The chosen solution is then constrained so that it is physically more coherent. Statistically, this regularization avoids local minima during the learning process and it speeds it up.

Introduction of a priori first guess information as part of the input to the neural network is proposed. This idea is simple and general. However, the inputs of the network are no longer homogeneous (i.e., different types of variables, which changes their dynamics), but this problem can be solved as described below. First, the neural transfer function becomes

$$\hat{x} = g_W(x^b, y^o), \tag{10}$$

where $\hat{x}$ is the retrieval (i.e., retrieved physical parameters), g_W is the neural network g with parameters W, x^b is the first guess for the retrieval of physical parameters x, $y^o = y(x) + \eta$ are the observations, and η the observation noise.

The learning algorithm consists of estimating the parameters W of the neural network that minimize the mean least squares error criterion. The term "mean" depends on the probability distribution functions of the physical problem. In this experiment the least squares criterion is of the following form

$$C_2(W) = \frac{1}{2}\int\int\int D_E(g_W(x^b, y^o), x)^2 \; P(x, y^o, x^b) \tag{11}$$

$$C_2(W) = \frac{1}{2}\int\int\int D_E(g_W(x+\varepsilon, y+\eta), x)^2 \; P(x)P_\eta(\eta)P_\varepsilon(\varepsilon), \tag{12}$$

where $P(x)$ is the probability distribution function of the physical variables x that depends on the natural variability. $P_\eta(\eta)$ is the probability distribution function of the observation noise η. $P_\varepsilon(\varepsilon)$ is the probability distribution function of the first guess error $\varepsilon = x^b - x$.

The quality criterion in (11) is very similar to the quality criterion of variational assimilation in (A2). The differences are that in neural network criterion we minimize a difference (i.e., least squares approach for the Euclidian distance), and in the variational assimilation approach, we maximize a conditional expectation (i.e., maximum likelihood approach, which is close to least squares). Furthermore, neural network criterion in (11) involves the distribution $P(x)$. This is due to the fact that the neural network simulates the inverse of the radiative transfer equation globally, once and for all, and uses the distribution $P(x)$ for this purpose. The neural network model is then valid for all observations (i.e., global inversion). The variational assimilation model has to compute an estimator for each observation (i.e., local inversion).

To minimize this criterion, we create a data set $\mathcal{B} = \{(x^e, y^{oe}, x^{be});\ e = 1, \ldots, N\}$ that samples as well as possible all the probability distribution functions in (11). Then the practical criterion used during the learning stage is given by

$$\tilde{C}_2(W) = \frac{1}{2N}\sum_{e=1}^{N} D_E(g_W(x^{be}, y^{oe}), x^e))^2. \tag{13}$$

First, to sample the probability distribution function $P(x)$, we select geophysical states (x^e) that cover all natural combinations and their correlations and by calculating $y^e = y(x^e)$ with the physical model (the radiative transfer model in this case). Alternatively, we could obtain these relationships from a "sufficiently large" set of colocated and coincident values of y and x. For sampling P_η, we need a priori information about the measurement noise characteristics; a physical noise model could be used, but if all we have is an estimation of the noise magnitude, then we have to assume Gaussian distributed noise η that is not correlated among the measurements. To sample the first guess variability with respect to state x (i.e., sampling $P(x^b|x)$), there are two situations. If a first guess data set $\{x^{be};\ e = 1, \ldots, N\}$ exists, then x^{be} can be used directly. If such a data set is not available, we have to determine $P(\varepsilon)$ (as it is done in variational assimilation technique), the distribution of errors in the first guess, $\varepsilon = x^b - x$, and use $x^b = x + \varepsilon$ as input to the network. The balance between reliance on the first guess and the direct measurements is then made automatically and optimally by the neural network during the training.

Table 1 summarizes the specific features of the neural network scheme with first guess and the variational assimilation inversion technique.

3. Generation of a Database to Train the Neural Network

To constrain the problem (the problem is then better posed), we use the clear/cloudy flag information provided by the ISCCP data set to train two neural networks: One for clear scenes (NN1) and one for cloudy scenes (NN2). This specialization of the NNs facilitate the training of the neural network models. They both retrieve simultaneously the surface temperature Ts, the integrated water vapor content WV, and the seven SSM/I surface emissivities Em_i. In addition to these parameters, NN2 retrieves the cloud liquid water

Table 1. Comparison of the Variational Assimilation and Neural Network Inversion Schemes

	Variational Method	Neural Network Inversion
Observation / measurement	y^o	y^o
First guess a priori information	x^b	x^b
Retrieved variable	x	x
Direct model used	radiative transfer model, used during the inversion	radiative transfer model, used during the construction of $\mathcal{B}$, if no collocated data set exists
Inverse model	linearized locally	nonlinear, global
Model	$y(x) = y(x_n) + H(x_n)(x - x_n)$	nonlinear: $x = g_w(x^b, y^o)$
Quality criterion	$\int\int x\ P(x\|y^o, x^b)$	$\frac{1}{2}\int\int\int D_E(g_W(x^b, y^o), x)^2\ P(x, y^o, x^b)$
Data set of observations	used to estimate the first guess error covariance matrix B	used to sample the pdfs
Direct model errors	assumed to be Gaussian: with error covariance matrix F	already sampled in the data set, if $\mathcal{B}$ is simulated by a RT model
First guess error	assumed to be Gaussian: with error covariance matrix B	no constraint, simulated using true and first guess solution data sets
Observation error	assumed to be Gaussian: with error covariance matrix E	no constraint, depends on instrument, supposed Gaussian in this study, E
Inversion type	local inversion: inversion process for each observation	global inversion: estimation of the inverse model once and for all

path LWP. Two sources of information are used for this purpose: (1) seven SSM/I brightness temperatures (observations), and (2) a priori information of the state of the surface and atmospheric variables from ancillary data sets. In this study the experimental configuration is similar to the one used by *Prigent and Rossow* [1999]. A collection of SSM/I observations colocated and coincident with independent measurements of the parameters to be retrieved (Ts, WV, LWP, and the Em_i) is not available. However, other estimates of Ts and LWP are available every 3 hours from ISCCP, NCEP provides WV analysis every 6 hours, and the land surface microwave emissivities are available as monthly estimates. As a consequence, brightness temperatures simulated by the radiative transfer model are used in the database instead of observations. These radiative transfer results are obtained using the selected values of T_s, WV, LWP, and Em_i. To the extent that these data sets provide a proper global distribution of the surface and atmospheric parameters, including their correlations, the neural network represents a global fit of the inverse radiative transfer model (i.e., transfer function).

The SSM/I instrument on the Defense Meteorological Satellite Program polar satellites senses atmospheric and surface emissions at 19.35, 22.235, 37.0, and 85.5 GHz with both horizontal and vertical polarizations, except for 22.235 GHz which is vertical polarization only [*Hollinger et al.*, 1987]. An instrument evaluation has been performed by *Hollinger et al.* [1990], and an intersensor calibration has been completed by *Colton and Poe* [1999]. The radiometric noise is supposed to be Gaussian distributed, so it is entirely defined by its noise covariance matrix $E =< \eta \cdot \eta^t >$. Errors in channels are supposed to be uncorrelated, and the standard deviation of each channel error is estimated to be 0.6 K. So matrix E is defined as $0.6 \times I_{7\times7}$, where $I_{7\times7}$ is the 7×7 identity matrix.

3.1. A Priori First Guess Information and Related Background Errors

3.1.1. Water vapor first guess derived from the NCEP reanalysis. The temperatures and relative humidities for eight levels up to 300 mbar (middle and lower troposphere) are available from the NCEP reanalysis data set. The NCEP reanalysis project is described by *Kalnay et al.* [1996]. It uses various data collections but excludes SSM/I derived information. These profiles are available every 6 hours at a spatial resolution of 2.5° in latitude and longitude. For each location the atmospheric profile has been adjusted for consistency with the topography (truncation or downward extrapolation of the atmospheric profiles depending on the topography differences between the NCEP reference elevation and ours). The integrated water vapor WV is used as first guess a priori information. The first guess error is taken to be 0.4 times the NCEP WV first guess, similar to the WV error values obtained when using the error covariance of each humidity level as given by *Eyre et al.* [1993]. The mean temperature of the first atmospheric layer Ta derived from NCEP is also used as a priori information in the retrieval.

3.1.2. Use of the ISCCP data set. In the ISCCP data, cloud parameters and related quantities are retrieved from visible (VIS ~0.6 μm wavelength) and infrared (IR ~11 μm wavelength) radiances provided by the set of polar and geostationary meteorological satellites [*Rossow and Schiffer*, 1999]. The ISCCP data set

is used in this study to discriminate between clear and cloudy scenes (selecting NN1 or NN2) and to give estimates of the cloud top temperature and surface skin temperatures. The pixel level data set (the DX data set) is selected for its spatial sampling of about 30 km and its sampling interval of 3 hours [*Rossow et al.*, 1996].

3.1.2.1. Surface temperature first guess: ISCCP provides the surface skin temperature first guess retrieved from IR radiances under clear conditions. The IR emissivity of the surface is always close to 1 and varies with the land surface type as in the Goddard Institute for Space Studies general circulation model. Instead of selecting the closest-in-time DX image to derive the surface temperature, a linear interpolation between two ISCCP surface temperature estimates to the precise time of the SSM/I overpass is calculated to account for the diurnal cycle. If the ISCCP DX scenes are cloudy, a clear-sky compositing procedure is conducted within the ISCCP process to derive an estimation of the surface temperature (see *Rossow and Garder* [1993] for more details). The error associated with the surface temperature is estimated to be 4 K [*Rossow and Garder*, 1993].

3.1.2.2. Cloud a priori information: First the ISCCP data discriminates between clear and cloudy scenes. Over the ocean it has been shown that the VIS and IR observations have a better ability than the microwave measurements to detect clouds [*Lin and Rossow*, 1994]. Given that the sensitivity of the microwave to clouds over land is much lower than over ocean, when a pixel is considered clear by the ISCCP procedure, the LWP is fixed to zero. Two neural networks are used, one for clear scenes another for cloudy scenes. The ISCCP cloud flag directs the retrieval to one network or the other.

For cloudy scenes, the cloud top temperature derived from IR measurements is added to the retrieval process as additional information to account for the changes in the emission temperature of the cloud and in the cloud liquid water absorption coefficient. In contrast to the ocean case, clouds induce only small variations in the microwave radiation over land and additional cloud information improves their retrieval. *Prigent and Rossow* [1999] showed that the ability to estimate liquid water path depends essentially on the contrast between the radiance emitted by the cloud and the radiance that emanates from the surface: consequently, the accuracy of LWP retrieval varies widely with the cloud condition, especially with cloud top temperature [*Prigent and Rossow*, 1999]. Thus knowledge of the cloud top temperature helps retrieve cloud liquid water path, and the ISCCP DX cloud top temperature derived from IR measurements is used as a priori information. If the ISCCP DX cloud top temperature is ≥ 260 K, it is assumed that the cloud is composed of liquid water [*Lin and Rossow*, 1997] and the location of the cloud temperature is dictated by ISCCP.

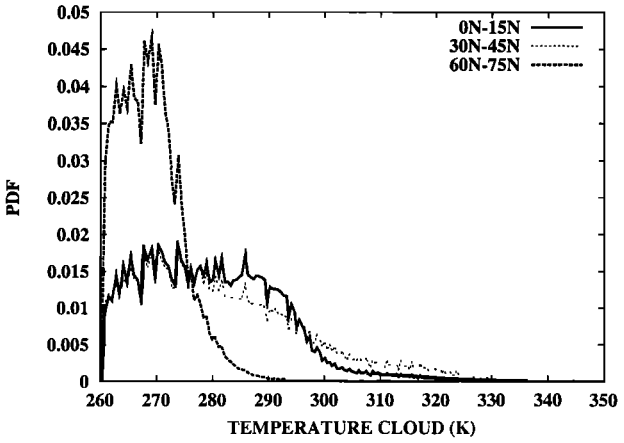

Figure 2. Distribution of liquid water cloud top temperature (Tc) in June for three different latitude zones: tropical (0°N-15°N), mid latitude (30°N-45°N), and high-latitude (60°N-75°N) zones.

If the ISCCP cloud top temperature is < 260 K, the higher portion of the cloud is probably formed of ice and the contribution of the ice in the cloud is not considered. There is a possibility that this ice cloud obscures a liquid cloud or becomes a liquid cloud below. An analysis has been performed on global ISCCP DX data to estimate the statistical distribution of cloud top temperature of the liquid clouds over land for each 15° latitude zone for each month. Examples of these distributions are given in Figure 2. Assuming that the distribution of liquid cloud top temperatures is not modified in the presence of an overlying ice cloud, the first guess liquid cloud top temperature is then stochastically drawn from the distribution of liquid cloud top temperature of the corresponding month and latitude zone to maintain random error characteristics.

3.1.3. Microwave emissivities. First guess a priori information for the microwave emissivities at each location are derived from the monthly mean emissivities previously estimated by *Prigent et al.* [1997]. The standard deviation of day-to-day variations of the retrieved emissivities within a month σ_ϵ have been calculated for each channel and for each location and are used as estimates of first guess errors for these quantities.

3.2. Radiative Transfer Model

A direct radiative transfer (RT) model adapted to the SSM/I channels is used to create the learning and testing databases required for the neural network inversion. The MPM 93 gaseous absorption model of *Liebe et al.* [1993] is adopted for all the SSM/I frequencies. In this model, H_2O and O_2 lines are added up to 1 THz, assuming a Van Vleck Weisskopf line shape. For oxygen, this function is modified by *Rosenkranz* [1992] to include line coupling in the 60 GHz band. An empirical H_2O continuum is added, derived from laboratory measurements [*Liebe et al.*, 1993]. A revised gaseous

absorption model, validated up to submillimeter wavelengths, has been recently developed [*Pardo et al.*, 2001] but is not used in this study because the differences at SSM/I frequencies are negligible.

Cloud absorption is calculated using the Rayleigh approximation which is valid for most non precipitating liquid water clouds at SSM/I frequencies. The cloud temperature is assumed to be equal to the air temperature at the same level. The dielectric properties of liquid water are taken from *Manabe et al.* [1987]. Scattering by large particles is not considered, meaning that convective clouds and rain are not represented in the database. The surface contribution is calculated using the monthly mean emissivities previously calculated [*Prigent et al.*, 1997, 1998] and assuming specular reflection at the surface.

The consistency of the radiative transfer model has been checked. Observed brightness temperatures and simulated Tbs using the ISCCP Ts and LWP, the NCEP WV, and the monthly Em_i have been compared for 2 months of SSM/I data globally over snow and ice-free pixels: For all channels the bias is smaller than 0.5 K even for cloudy cases. Thus the training data set generated with this radiative transfer model and sources of global data accurately represents the distribution of these parameters that SSM/I observes.

3.3. Statistical Analysis of the Training Database

The training database generated by the RT model applied to the ISCCP, NCEP, and monthly Em_i data sets contains the variables to be retrieved (Ts, WV, LWP, and the seven Em_i), the seven simulated brightness temperatures Tb, and a priori information on the cloud top temperature Tc and the temperature of the lowest layer of the atmosphere Ta. An error is associated with most variables that are used as first guesses. The database is produced from data collected for January and June 1993 over land between 60°S and 80^{circ}N. Snow- or ice-covered pixels are not considered: The snow and ice information comes from the NOAA operational analysis. Of 1,391,671 samples collected, 55% of them correspond to cloudy scenes.

Figure 3 shows the "global" distributions of some of the variables in the training database. The distributions are non-Gaussian and some of them are truncated. For example, the liquid water path distribution has its peak frequency at the lowest values and obviously cannot be negative. When retrieving such a variable with the usual form of the variational assimilation, the assumptions of a Gaussian distribution and positive values introduces biases in the retrieval. A change of variables is sometimes performed to alleviate this problem of non-Gaussian distributions [*Phalippou*, 1996]. The clear and cloudy distributions of the brightness temperatures are very similar for all frequencies and for both polarizations indicating the difficulty of detecting clouds over land. These distributions have been used to control the quality of the data: Values lying beyond the first zero frequency on either side of the mode values in the distributions have been suppressed.

A good representation of the physical probability distributions in the learning data set is fundamental to describe all complex relationships of the variables, especially when the distributions are non-Gaussian. The data set needs to sample the real probability density function (pdf) if we want the neural network to be optimal for this natural variability. Another possible approach would be to use uniform distributions in order to give the same statistical weight to each atmospheric condition, even to extreme events, which would allow the neural network to have the same level of accuracy for all atmospheric situations. In this work we have chosen to optimize accuracy for the most frequently occurring events.

Table 2 presents the linear correlations between the variables, in the global database for clear and cloudy cases, separately, to illustrate the inter dependence of the variables in the database. These correlations do not distinguish direct dependence between variables from indirect ones due to intermediate variables: Variables that are not physically related can be statistically correlated via a third variable. This illustrative calculation assumes linear relationships between the variables but nonlinear relationships are more likely. Note that the neural network technique can exploit these nonlinear correlations between variables to improve the retrievals, whereas the usual forms of the variational assimilation approach neglect all correlations.

The seven Tb are strongly correlated with each other, especially for a given polarization. This fact is often ignored in design of simpler retrieval methods and in the estimation of their errors (cf. discussion by *Lin and Rossow* [1994]). Note that the Tb in this table are calculated by the physical model, but similar results (not shown) are obtained with observed Tb. The seven Em_i are also highly correlated (correlation $\geq$0.7): Since the seven emissivities are not independent variables, use of their statistical relationships will definitively help constrain their retrieval.

At a given frequency, correlation of the surface temperature with brightness temperature is higher for the vertical polarization than the horizontal one. This can be explained by less variability in the emissivities for the vertical polarization than for the horizontal one. Correlations between surface temperature and brightness temperatures at vertical polarization are similar at all SSM/I frequencies, which was not anticipated. At 22 and 85 GHz, water vapor absorption was expected to impede a direct relation between surface contributions and top of the atmosphere measurements, and as a matter of fact, derivatives of the brightness temperatures with surface temperature (sensitivities) are smaller at 22 and 85 GHz than at 19 and 37 GHz [*Prigent and*

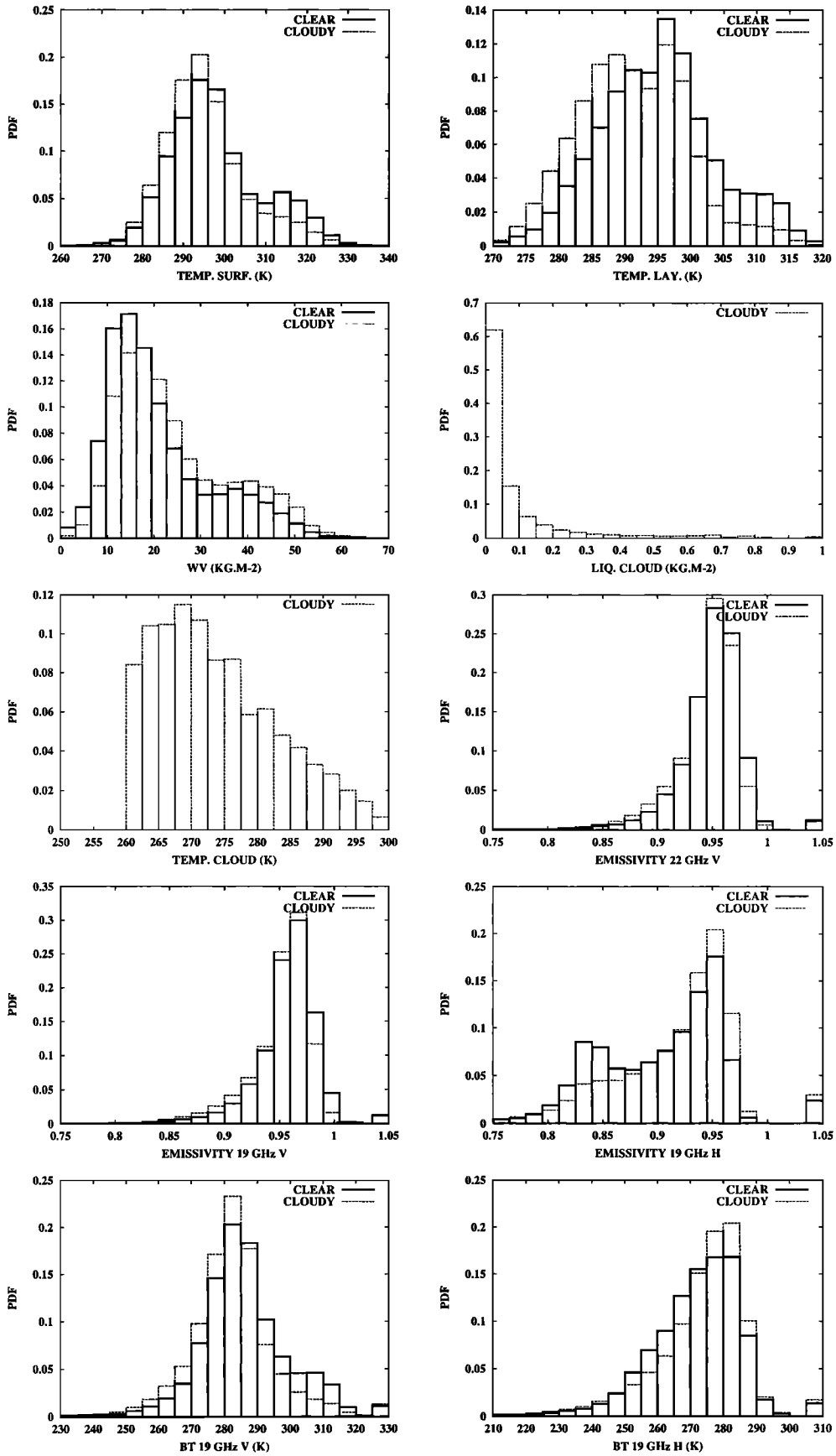

Figure 3. Probability distribution functions of variables in the training database.

Table 2. Global Linear Correlations of Physical Variables and the Calculated Microwave Brightness Temperatures[a]

	Ts	*WV*	*Ta*	*Tc*	*LWP*	*Tb1*	*Tb2*	*Tb3*	*Tb4*	*Tb5*	*Tb6*	*Tb7*	*Em1*	*Em2*	*Em3*	*Em4*	*Em5*	*Em6*	*Em7*
Ts	1.00	0.15	0.80	...	...	0.85	0.43	0.89	0.88	0.49	0.87	0.59	-0.04	-0.30	-0.08	-0.15	-0.34	-0.34	-0.38
WV	0.20	1.00	0.36	...	...	0.09	0.44	0.16	0.15	0.45	0.39	0.61	-0.17	0.24	-0.16	-0.19	0.21	-0.06	0.19
tlay	0.70	0.58	1.00	...	...	0.74	0.37	0.79	0.76	0.42	0.81	0.57	0.03	-0.24	-0.01	-0.08	-0.29	-0.26	-0.32
Tc	0.27	0.24	0.40	1.00	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
LWP	-0.13	0.08	-0.10	-0.09	1.00	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Tb1	0.83	0.23	0.69	0.28	-0.11	1.00	0.52	0.99	0.98	0.55	0.84	0.55	0.49	-0.07	0.44	0.35	-0.12	-0.02	-0.20
Tb2	0.46	0.42	0.43	0.16	-0.02	0.71	1.00	0.53	0.54	0.98	0.62	0.87	0.26	0.73	0.25	0.22	0.68	0.23	0.58
Tb3	0.86	0.30	0.76	0.33	-0.12	0.98	0.69	1.00	0.99	0.57	0.89	0.61	0.39	-0.10	0.35	0.27	-0.15	-0.06	-0.21
Tb4	0.87	0.25	0.70	0.30	-0.08	0.98	0.68	0.98	1.00	0.60	0.91	0.64	0.40	-0.08	0.37	0.32	-0.10	0.03	-0.15
Tb5	0.53	0.43	0.47	0.18	0.03	0.73	0.98	0.73	0.74	1.00	0.69	0.93	0.21	0.65	0.22	0.22	0.40	0.25	0.57
Tb6	0.86	0.44	0.80	0.41	-0.15	0.86	0.64	0.92	0.92	0.71	1.00	0.83	0.12	-0.03	0.11	0.10	-0.04	0.04	-0.04
Tb7	0.65	0.60	0.67	0.33	-0.03	0.72	0.83	0.77	0.77	0.90	0.88	1.00	0.02	0.43	0.04	0.06	0.43	0.18	0.42
Em1	0.01	-0.02	0.11	0.05	-0.03	0.56	0.56	0.46	0.46	0.50	0.23	0.26	1.00	0.36	0.99	0.93	0.33	0.54	0.24
Em2	-0.21	0.19	-0.09	-0.05	0.04	0.20	0.75	0.13	0.13	0.67	0.04	0.38	0.66	1.00	0.38	0.40	0.98	0.52	0.89
Em3	-0.03	-0.03	0.07	0.03	-0.02	0.51	0.54	0.42	0.43	0.49	0.19	0.25	0.99	0.67	1.00	0.96	0.37	0.62	0.32
Em4	-0.11	-0.14	0.05	-0.03	0.00	0.41	0.46	0.32	0.37	0.44	0.12	0.20	0.93	0.64	0.95	1.00	0.43	0.76	0.41
Em5	-0.26	0.12	-0.16	-0.08	0.05	0.14	0.70	0.08	0.10	0.65	0.01	0.37	0.63	0.97	0.66	0.68	1.00	0.60	0.95
Em6	-0.30	-0.19	-0.25	-0.13	0.05	0.07	0.27	0.01	0.09	0.29	-0.01	0.12	0.61	0.55	0.69	0.81	0.65	1.00	0.72
Em7	-0.32	0.06	-0.22	-0.11	0.06	0.03	0.57	-0.01	0.03	0.56	-0.02	0.33	0.53	0.87	0.60	0.66	0.94	0.79	1.00

[a]Upper (lower) triangle of the table is for the cloud-free (cloudy) case.

Rossow, 1999]. Other authors have also observed large correlations at 85 GHz between surface air temperatures and *Tbs* at 85 GHz. *MacFarland et al.* [1990] investigated the correlation between SSM/I observations and surface air temperature and concluded that 22 and 85 GHz measurements, depending on the surface type, are the most sensitive to the land surface temperature. *Basist et al.* [1998] also proposed a method to retrieve near-surface air temperature from SSM/I that relies heavily on the 85 GHz channels. These results can be explained by two factors. First, for a given polarization the surface emissivities at 19 GHz are more variable than at other frequencies because of higher sensitivity to surface properties like soil moisture or vegetation water content and structure. These emissivity variations are not correlated with surface temperatures fluctuations as indicated by the correlation coefficients between *Ts* and emissivity at 19 GHz (see Table 2). Second, the absorption at 22 and 85 GHz actually attenuates the effects of emissivity fluctuations, enhancing the relationship between brightness temperatures and surface temperature. The global correlation coefficients in Table 2 may not be representative on a local scale. However, correlation coefficients for *Ts* have been calculated for three ranges of atmospheric water vapor amount and emissivities and no significant differences were observed in the coefficients.

Global correlations between atmospheric water vapor and brightness temperatures are relatively low especially for vertical polarization because of large surface emissivities reducing the contrast between atmospheric and surface emissions. Even for horizontal polarization, global correlations never exceed 0.6. However, these global values mask large local differences. Correlation coefficients calculated for different ranges of emissivities and water vapor amounts show that the results are very different, especially for the 85 and 22 GHz channels depending on water vapor amount. As a consequence, the use of simple algorithms based on linear regressions or on global correlation statistics will not be adequate.

Cloud liquid water path is not correlated with the *Tbs* (correlation coefficients lower than 0.2), whatever the channel. This casts doubts on the ability of the SSM/I observations for accurate liquid water retrieval over land, unless additional observations or a priori information are added to the retrieval process.

The neural network analyzes all the local statistical relationships in the database and benefits from them, even when the relationships are highly nonlinear. These relationships represent nonlinear correlations among the physical variables, among the observations (brightness temperatures), among the first guess errors, and between the variables and the observations. All of these correlations constitute additional information which the neural network can exploit to improve its retrieval if such correlations are properly represented in the training data set.

In contrast, the variational assimilation scheme, in its classical implementation, does not take into account statistical information about correlation between the variables (see Appendix A). It only uses the matrices of first guess and observation error covariances. In most formulations it is assumed that the first guess errors are uncorrelated (i.e., diagonal covariance matrices). These matrices are often calculated locally and may not be representative at a global scale. At the European Centre for Medium-Range Weather Forecasts (ECMWF), for instance, the matrices of error covariance for water vapor and temperatures have been estimated from observations in the London area only [*Eyre et al.*, 1993; *Gadd et al.*, 1995]. Presently, first guess error (or, more precisely, background error, in the variational assimilation terminology) covariance matrices are estimated by ECMWF with a certain latitude dependence assuming that the difference between forecasts at different ranges valid at the same time are representative of short-range forecast error [*Rabier et al.*, 1998; *Derber and Boutier*, 1999]. Still, no cross correlation between the first guess error of temperature, specific humidity, and ozone is used, whereas situation-dependent first guess error covariance matrices with different length scales at different locations are desirable.

To represent such complex relationships between the variables, matrices of error covariances would have to be calculated for a set of situations that describe the correlated time and space variabilities of the parameters. Use of localized covariance matrices corresponds to the linearization of a complex, nonlinear function, possibly producing continuity problems. Because of its nonlinear capacity, the neural network approach avoids these difficulties by adapting itself to the statistical variabilities and correlation relationships of the physical variables, provided that the database satisfactorily describes the variety of the situations to be analyzed.

4. Results From the Neural Network Inversions

Two neural networks have been trained, one for clear pixels (NN1) the other one for cloudy pixels (NN2), both using a priori first guess information. The ISCCP cloud flag discriminates between clear and cloudy pixels. The architecture of the network NN1 is a MLP with 17 inputs coding the seven SSM/I observations, y^0, and the first guess, x_b (Ts, Ta, WV, and 7 Em_i), 30 neurons in the hidden layer, and 9 neurons in the output layer coding the retrieval, x (Ts, WV, and 7 Em_i). The number of neurons in the hidden layer is estimated by a heuristic procedure that monitors the generalization errors of the neural network as the configuration is varied. The network NN2 has one additional input, the cloud top temperature Tc, and one additional retrieval, the liquid water path (LWP). The input variables and their associated standard deviation errors are summarized in Table 3. The full matrix of the error covariances is calculated at the end of the training phase (not shown here). This matrix gives the statistical structure of the

Table 3. RMS Error Results for First Guess and Retrievals

	Observation or First Guess Errors	NN1 Clear Without First Guess	NN1 Clear With First Guess	NN2 Cloudy Without First Guess	NN2 Cloudy With First Guess
TbSSMI 19 GHz V (K)	0.600	...	...	...	...
TbSSMI 19 GHz H (K)	0.600	...	...	...	...
TbSSMI 22 GHz V (K)	0.600	...	...	...	...
TbSSMI 37 GHz V (K)	0.600	...	...	...	...
TbSSMI 37 GHz H (K)	0.600	...	...	...	...
TbSSMI 85 GHz V (K)	0.600	...	...	...	...
TbSSMI 85 GHz H (K)	0.600	...	...	...	...
Ta[a](K)	3.000	...	...	...	...
Tc[b](K)	2.000	...	...	...	...
Ts[b](K)	4.000	3.470	1.340	3.310	1.570
LWP[b](kg.m-2)	...	...	...	0.090	0.080
WV[a](kg.m-2)	40.00[c]	5.330	3.830	6.860	4.900
Em 19 GHz V	0.016	0.012	0.004	0.012	0.006
Em 19 GHz H	0.018	0.011	0.004	0.012	0.006
Em 22 GHz V	0.018	0.013	0.005	0.013	0.006
Em 37 GHz V	0.015	0.012	0.004	0.012	0.006
Em 37 GHz H	0.018	0.011	0.005	0.013	0.006
Em 85 GHz V	0.020	0.015	0.006	0.016	0.009
Em 85 GHz H	0.023	0.016	0.008	0.018	0.010

[a]NCEP.
[b]ISCCP.
[c]In %.

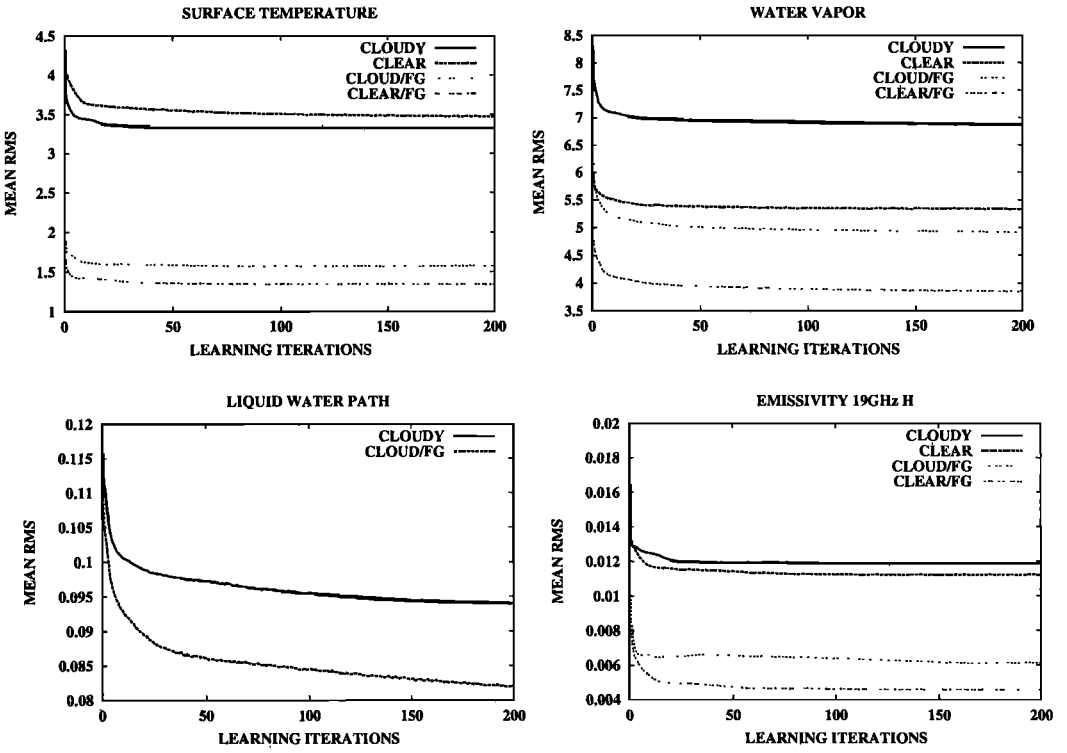

Figure 4. Time evolution of the RMS retrieval errors during the learning phase of the neural network for network with (FG) and without first guess inputs.

errors and is of great importance in the assimilation of retrieved products in a Numerical Weather Prediction scheme.

For each variable the distribution of the first guess error is a Gaussian truncated at 2 standard deviations. In contrast to the variational assimilation, where only Gaussian distributions can be used, the neural network method can use any distribution shape. However in the present study, no in situ data are available to calculate the distribution of the first guess errors, so Gaussian noise is introduced independently for each variable to generate the first guess. In the operational mode with real first guesses, the technique will use the structure of the first guess error correlations and the results should be even better.

Figure 4 presents the learning curves of the neural network for clear and cloudy situations with and without first guesses. To measure the impact of the introduction of the first guess information in a neural network inversion scheme, two additional networks have been trained without first guesses, one for clear conditions and another for cloudy scenes. The architectures of the networks without first guess are similar in structures, except that there are only seven inputs, coding the SSM/I observations y^0. For each retrieved variable, the RMS error decreases from the first guess RMS error to a stable value after several iterations. The networks with first guess input show substantially better fits to the training data set (see Table 3 for retrieved variable RMS global errors for the networks with and without first guess). The continuity between the two networks NN1 and NN2 at low liquid cloud content ($LWP \leq 0.005$) has been checked and is satisfactory: Mean differences are -0.5 K in Ts and 0.79 kg m^{-2} in WV.

Results for 1 day are displayed on Plate 1, for Ts, WV, LWP, and Em at 19 GHz horizontal polarization. For the same retrieved variables, Figure 5 shows the distributions of the retrieval errors, separately for three ranges of Em and for clear and cloudy scenes, since different sensitivities to the retrieved parameters are expected depending on the surface and cloud characteristics. The surface types classified by monthly mean Em at 19 GHz in the horizontal polarization are related to the vegetation density [*Prigent et al.*, 2001]: Surfaces with 19 GHz $Em < 0.9$ for the horizontal polarization generally correspond to desert-like areas; zones of dense vegetation show 19 GHz emissivities > 0.95. Cloudy scenes are divided into two groups according to their LWP estimated by ISCCP. The results for each variable are discussed next.

4.1. Surface Temperature

The SSM/I observations have a good ability to measure the surface skin temperature with a RMS error of 1.3 K in clear areas and 1.6 K in cloudy cases. This RMS error represents a significant improvement over the first guess RMS of 4 K, which was based in part on the inability of infrared instruments to measure surface temperature under cloudy conditions. The accuracy of

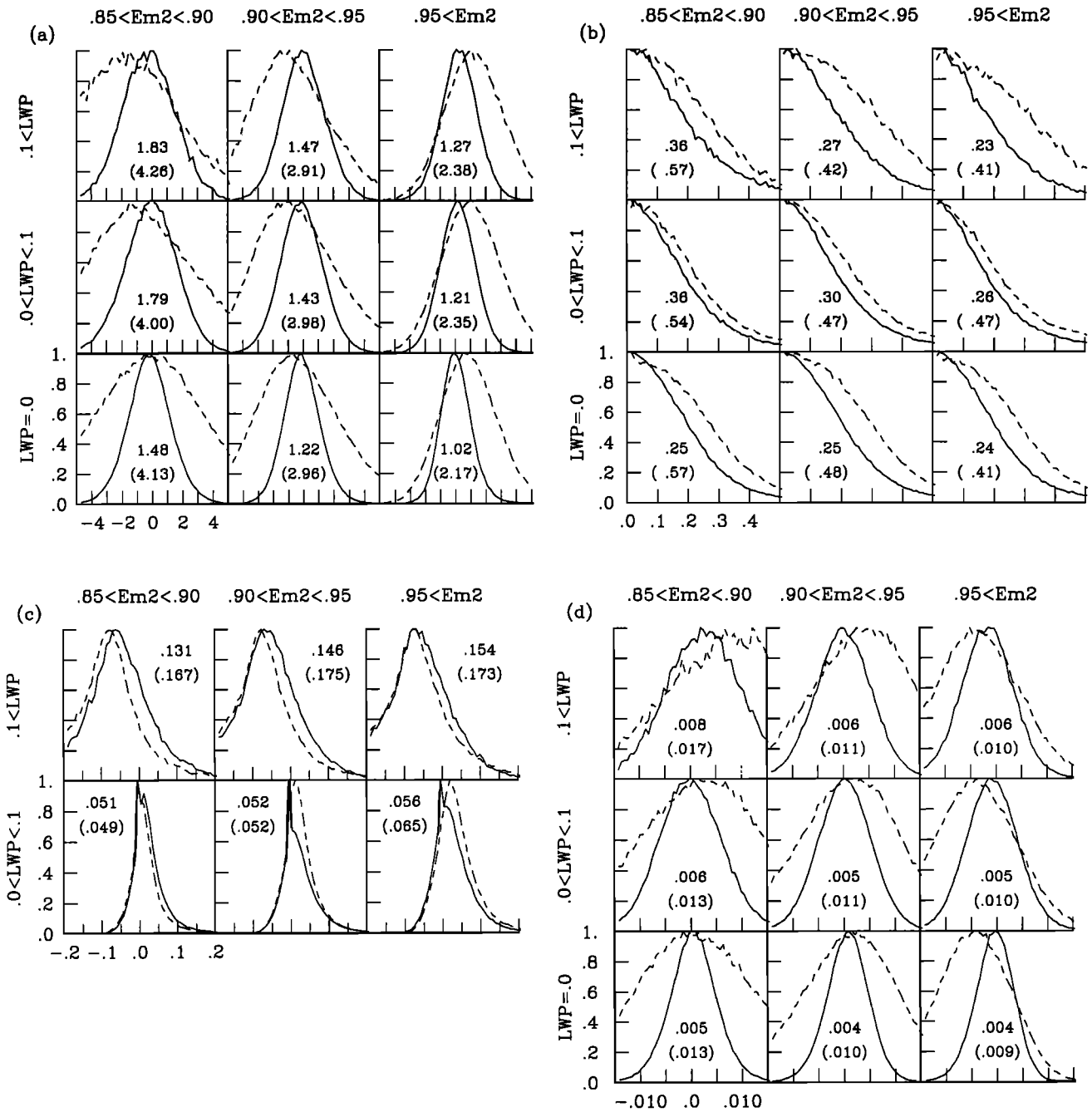

Figure 5. Normalized histograms of the errors (new minus observed) for (a) Ts in Kelvin, (b) WV in relative error, (c)LWP in kg m^{-2}, and (d) Em 19 GHz horizontal polarization. Results are presented for three cloud conditions (clear-sky, ISCCP LWP lower than 0.1 kg m^{-2}, and higher than 0.1 kg m^{-2}) and for three ranges of E_m at 19 GHz horizontal polarization. Solid lines indicate the errors with first guess and dashed lines without first guess. The RMS errors are indicated with first guess and without first guess (in parentheses).

this retrieval is not affected much by the presence of clouds (Figure 5), and it increases slightly with increasing surface emissivity because of the increased contribution of the surface to the observed brightness temperatures. Examination of Plate 1 shows that the surface temperature fields exhibit very realistic gradients and there are no spurious structures related to variations in the emissivity fields. We checked that the Ts fields around the major rivers are not contaminated by rapid changes in the emissivity fields. Without a first guess solution containing estimates of the emissivities, the RMS error is much larger, especially for low surface emissivities. Most of the improvement is not expected to emanate from the surface temperature first guess itself: Correlation between surface temperature and brightness temperature is large enough to warrant a

good sensitivity to surface temperature variations. The benefit is most likely related to a good first guess in surface emissivity and the fact that the neural network can exploit the spectral dependence of the first guess emissivities to provide a more accurate estimate of both the emissivities and the surface temperature.

4.2. Water Vapor

WV is retrieved with a relative error of $\sim$ 30% for both clear and cloudy situations, when using a first guess. This is a small improvement over the first guess RMS error of 40%. The errors are not significantly different in the presence of clouds. With the variational assimilation method [*Prigent and Rossow*, 1999], the retrieval errors were found to increase with decreasing emissivities and to increase in presence of clouds as expected from the sensitivity of the radiative transfer to the various parameters. As observed in Table 2, the correlation between the brightness temperatures and WV is rather low (maximum of $\sim$0.6 globally), and the neural network scheme is likely to exploit water vapor correlation with another variable to extract water vapor information when direct correlation between Tbs and WV is not sufficient. It is worth mentioning that the neural network is trained to minimize the absolute WV error difference and not the relative error in WV. Changes could be made to minimize the relative error if this option was preferred.

4.3. Cloud Liquid Water Path

For LWP the RMS error is 0.08 kg m^{-2} globally, which is an improvement over variational assimilation retrieval and over few-channel methods. As expected, the error is larger in areas of high emissivities where the contrast between the land surface and the cloud is smaller. Even in areas of low emissivities ($0.85 \leq Em2 \leq 0.9$), the accuracy of the retrieval is not suitable for detection of majority of clouds. The cloud flag from ISCCP is of importance in this case to direct the retrieval toward the appropriate neural network. However, cloud structures with large liquid water path can be detected whatever the surface type; Plate 1 shows several extended and thick clouds that are also present on the ISCCP images (not shown). Plate 1 does not show any evidence of LWP errors (discontinuities) related to strong emissivity gradients.

4.4. Land Surface Emissivities

When using a first guess, the neural network technique shows a good aptitude for retrieving land surface emissivities with an RMS error lower than 0.008 (0.010) globally for all channels, in clear conditions (cloudy conditions, respectively). This is an improvement over the first guess errors. Unaided by the first guess estimate, the neural network technique does not perform so well. The first guess provides the emissivity spectral relationship and the retrieval exploits from it. The emissivity field presented on Plate 1 shows very consistent structures. The gradient between vegetated and arid areas is correctly retrieved, as well as specific hydrological structures like the Amazon or the Congo Rivers. The possibility of retrieving daily land surface emissivity with low RMS errors will enable an interesting range of studies from analysis of the effect of soil moisture or dew deposition to the postdetection of rainfall events (maybe even snowfall events can be recognized, but this idea needs further study). Detailed analysis and understanding of the emissivity variations will also benefit the microwave retrieval of WV and T profiles over land: until recently a fixed emissivity of 0.95 was used for Microwave Sounder Unit retrievals over land, and there is an urgent need for more accurate emissivity estimates [*English*, 1999].

4.5. Quality Control of the Retrieved Products

The quality of the retrieved products can be tested by comparing the brightness temperatures at the input of the neural network with brightness temperatures that are calculated with the retrieved products as inputs. Table 4 gives the RMS differences for each channel for the two neural networks with first guess (clear and cloudy). The mean RMS difference for all channels is 0.70 for clear conditions and 0.82 for cloudy scenes. These differences are of the order of the noise of each channel (0.60 K), showing that the retrieval scheme performs well. Further investigation reveals that the errors are larger for cases that are infrequent in the learning database (large cloud liquid water content or low emissivity areas for instance). Within the learning phase the synaptic weights W (equation (9)) are calculated so that the neural network functions in an optimal way on a global basis. For cases that are underrepresented in the database the retrieval will not be as accurate. This conclusion stresses the importance of the selection of an adequate training database, depending on the purpose of the retrieval. In this study, for instance, another alternative would be to have the same accuracy in the re-

Table 4. RMS Differences Between the SSM/I Brightness Temperatures at the Input of the Neural Network and the SSM/I Brightness Temperatures Computed From the Retrieved Products at the End of the Inversion Process[a]

	NN1 Clear With First Guess	NN2 Cloudy With First Guess
TbSSMI 19 GHz V (K)	0.55	0.70
TbSSMI 19 GHz H (K)	0.78	0.99
TbSSMI 22 GHz V (K)	0.61	0.68
TbSSMI 37 GHz V (K)	0.59	0.74
TbSSMI 37 GHz H (K)	0.82	0.96
TbSSMI 85 GHz V (K)	0.87	1.06
TbSSMI 85 GHz H (K)	1.13	1.31

[a]The first guess RMS error for each input brightness temperature is set to 0.60 K.

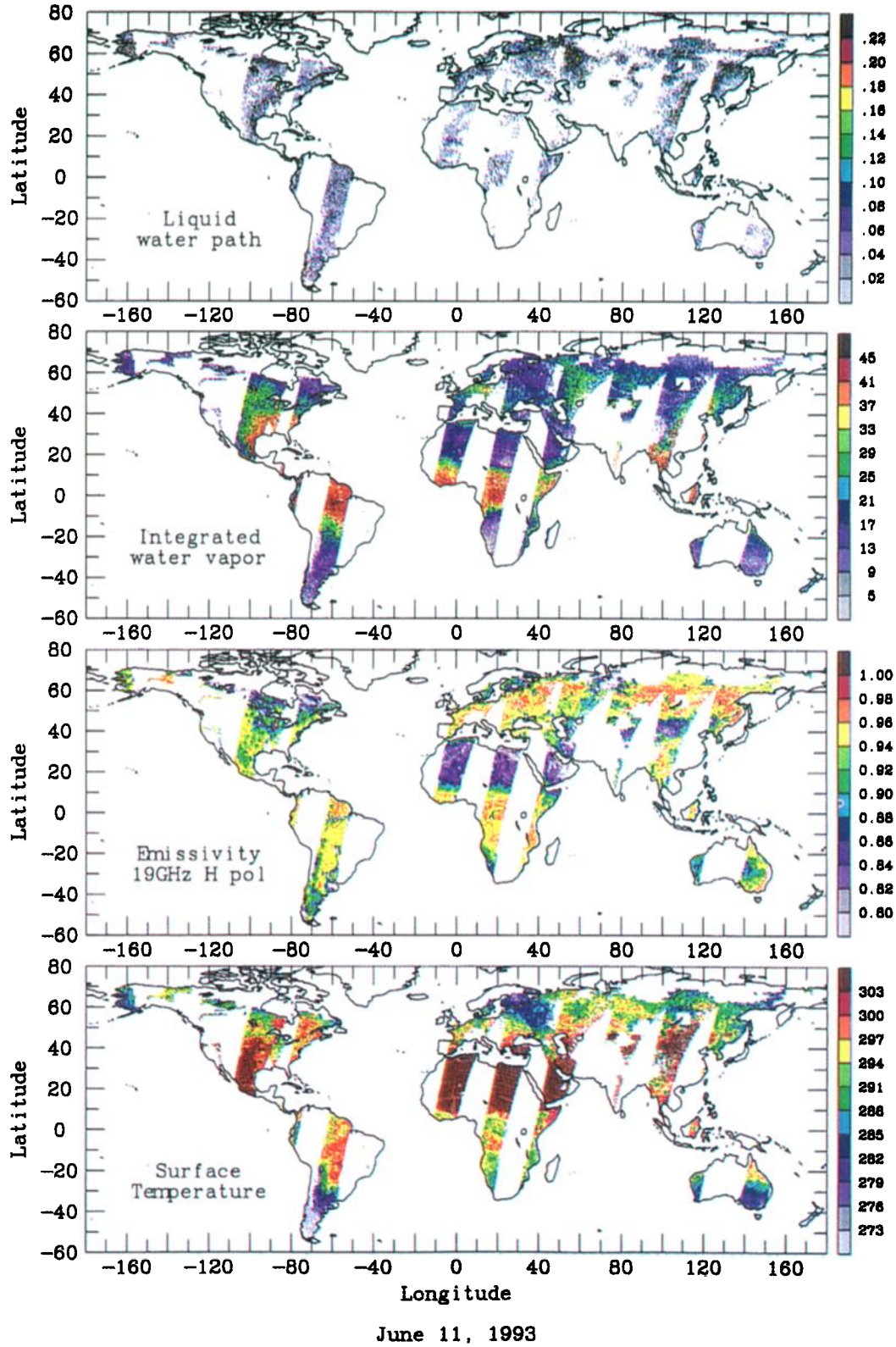

Plate 1. Retrieved fields (from the top to the bottom) of Ts in K, Em 19 GHz horizontal polarization, WV in kg m^{-2}, and LWP in kg m^{-2} for June 11, 1993, from SSM/I observations with the F10 and F11 satellites.

trieval, regardless of the surface emissivity. In this case, instead of the natural emissivity distributions shown on Figure 3, uniform distributions of the emissivities would be chosen. Also, instead of only cloudy-clear neural networks, we could have clear, cloudy (small water path), and cloudy (large water path) neural networks. This idea could be extended to precipitation cases.

4.6. Analysis of the Neural Network Sensitivities

An interesting capability of the neural network technique is that the adjoint model of the neural network is directly provided [*Aires et al.*, 1999]. The computation of this adjoint model (or neural Jacobians or neural sensitivities) is analytical and very fast. Since the neural network is nonlinear, these Jacobians are dependent on the situation x. For example, the neural Jacobians in our example of (7) (a MLP network with one hidden layer) are

$$\frac{\partial x_k}{\partial y_i} = \sum_{j \in S_1} w_{jk} \sigma' \left(\sum_{i \in S_0} w_{ij} y_i \right) w_{ij}, \qquad (14)$$

where σ' is the derivative of the activation function σ. For a more complex MLP network with more hidden layers, there exists a back-propagation algorithm that efficiently computes the neural Jacobians. The neural Jacobian concept is a very powerful tool since it allows for a statistical estimation of the multivariate and nonlinear sensitivities between input and output variables in the model under study (F. Aires and W.B. Rossow, Inferring instantaneous, multivariate and nonlinear sensitivities for the analysis of feedback processes in a dynamical system: Lorenz model case study, submitted to *Journal of Atmospheric Sciences*, 2000).

Table 5 gives the mean neural Jacobian values for the variables x_k and y_i for the neural network NN1 with first guess. The neural Jacobians are normalized by the standard deviation of the respective variables $((\partial x_k/\partial y_i) \times (std(y_i)/std(x_k)))$ to enable comparison of the sensitivities between variables with different variation characteristics. These values indicate the relative contribution of each input in the retrieval of a given output parameter. The numbers correspond to global mean values which may mask rather different behavior in various regions of the globe.

Figure 6 presents some of the normalized neural Jacobians for the surface temperature and the water vapor for three ranges of Em at 19 GHz H polarization. Depending on the surface emissivity, the sensitivity of Ts to different inputs changes from larger sensitivity to 19 GHz vertical polarization for high emissivities to larger sensitivity to the 85 GHz observations and the first guess information at low emissivities (Figure 6a). For WV retrieval, very different regimes are observed for low and high water vapor amounts (Figure 6b), from larger sensitivity to the 85 GHz channel horizontal polarization for high water vapor amount to smaller sensitivity for low water vapor contents. The same trend is observed at 22 GHz. We have already commented on the differences between local and global correlations in section 3.3. In contrast to a linear regresssion-type algorithm that fits a mean state mapping between inputs and outputs, the neural network can adapt itself to the different local situations by using optimally all of the input parameters. This means that the neural sensitivities are localized (depending on the situation) and are multivariate. The statistically normalized neural sensitivities are a multivariate and local generalization of the correlations between the inputs and the outputs.

5. Concluding Remarks and Perspectives

A neural network inversion scheme, including first guess information, is developed and applied to the retrieval of atmospheric water vapor, cloud liquid water,

Table 5. Global Mean Neural Sensitivities for NN1 (Clear Sky Condition)

	Tsurf	$Vap-int$	$Em1$	$Em2$	$Em3$	$Em4$	$Em5$	$Em6$	$Em7$
Tsurf	0.17	-0.13	-0.17	-0.11	-0.16	-0.19	-0.10	-0.12	-0.06
$Vap-int$	-0.04	0.33	0.04	0.00	0.04	0.03	-0.02	-0.04	-0.08
$Tb1$	0.21	0.18	0.58	0.02	0.47	0.13	-0.21	-0.19	-0.17
$Tb2$	0.14	0.32	-0.04	0.88	-0.17	-0.38	0.09	-0.22	-0.30
$Tb3$	0.09	-0.78	0.05	-0.09	0.16	-0.24	-0.09	-0.57	-0.24
$Tb4$	0.21	-0.04	0.17	-0.30	0.10	0.72	0.05	0.50	-0.03
$Tb5$	0.28	-0.95	-0.35	0.19	-0.26	0.04	0.79	-0.22	0.64
$Tb6$	0.25	-0.20	-0.38	-0.13	-0.30	-0.09	-0.28	0.89	0.04
$Tb7$	-0.21	2.30	0.03	-0.22	0.08	-0.17	-0.03	-0.21	0.36
$Em1$	-0.12	0.06	0.14	0.08	0.15	0.15	0.07	0.13	0.07
$Em2$	-0.12	-0.02	0.13	0.11	0.14	0.15	0.10	0.15	0.10
$Em3$	-0.09	0.05	0.11	0.06	0.14	0.12	0.06	0.14	0.08
$Em4$	-0.10	0.02	0.11	0.07	0.12	0.14	0.08	0.14	0.07
$Em5$	-0.12	-0.05	0.12	0.10	0.14	0.16	0.11	0.16	0.12
$Em6$	-0.05	-0.05	0.06	0.05	0.08	0.08	0.05	0.17	0.11
$Em7$	-0.05	-0.15	0.06	0.06	0.09	0.09	0.08	0.20	0.19
$tlay$	-0.03	0.07	0.00	0.00	-0.01	-0.01	-0.01	-0.06	-0.03

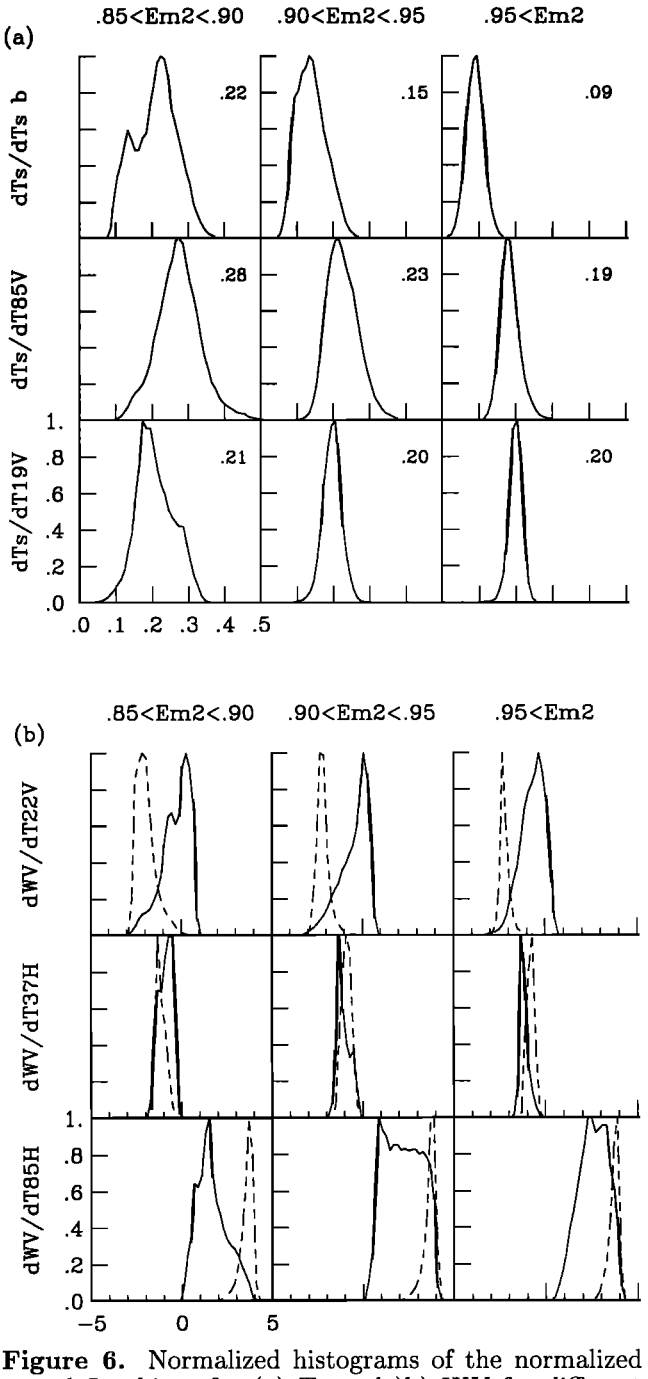

Figure 6. Normalized histograms of the normalized neural Jacobians for (a) Ts and)b) WV for different input variables. Results are presented for three cloud conditions (clear-sky, ISCCP LWP lower than 0.1 kg m^{-2}, and higher than 0.1 kg m^{-2}) and for three ranges of Em at 19 GHz horizontal polarization. In Figure 4a the mean values of the Jacobian are indicated for each subclass. In Figure 4b solid lines indicate the results for low WV amounts, whereas dashed lines represent large WV amounts.

surface temperature, and surface emissivities over land from SSM/I observations. Such a neural network approach has several similarities to the variational assimilation technique but has some advantages in both theoretical and practical aspects. Although their practical implementation can appear very different, this study shows that the two techniques share very basic aspects. They are both statistical inversion methods (analytical inversion methods could not deal with uncertainties in the inputs, and the physical models to invert in remote sensing are generally too complex) that minimize a quality criterion, using a priori first guess information and a radiative transfer (physical) model (Table 1). Theoretical advantages of the neural network scheme include its ability to perform a global inversion and to handle high nonlinearity and non-Gaussian variables. Only limited a priori hypotheses are required, which is why this technique is so flexible. This flexibility allows the neural network to exploit complex relationships among the observations and among the retrieved quantities that can vary with situation. Furthermore, the radiative transfer calculations are only needed once to generate the training data set. Even this is not needed if colocated and coincident in situ measurements are available, but this approach, which is like a traditional empirical analysis, cannot be generalized outside the observed variability of the data set. In principle, our model-based approach can be general. In application, inversion of new observations only involves simple and rapid calculations of two matrix products and one pass through the logistic function σ of the neural network (equation (7)). When processing large volumes of global observations, this is a very important asset compared to the variational assimilation. On the other hand, variational assimilation techniques have been designed to handle two-, three- and four-dimensional data (the 2-, 3- or 4-D-Var assimilation schemes used in Numerical Weather Prediction), a possibility that has not yet been investigated with neural network techniques.

In this study, we have developed a neural network scheme that includes first guess information. Its potential has been tested in the complex and ill-conditioned problem of inversion of SSM/I microwave observations over land. A database to train the neural network is derived from a global collection of coincident surface and atmospheric parameters, extracted from the NCEP reanalysis, from the ISCCP data, and from microwave emissivity atlases previously calculated. The introduction of the first guess information into the neural network has a considerable impact on the results compared to the network without first guess.

The theoretical RMS error of the surface temperature retrieval is 1.3 K in clear-sky conditions and 1.6 K in cloudy scenes over the globe. Microwave land surface temperature retrieval presents a very attractive complement to the infrared estimates in cloudy areas. By combining both measurements as we have done, a complete (clear and cloudy days) time record of land surface temperature can be produced. Water vapor is retrieved with a theoretical RMS error of 3.8 kg m^{-2} in clear conditions and 4.9 kg m^{-2} in cloudy situations. The theoretical RMS error in liquid water path is 0.08 kg m^{-2}. The surface emissivities are retrieved with an ac-

curacy of better than 0.008 in clear conditions and 0.010 in cloudy conditions, both improvements on the original first guess.

Validation of the retreived variables is on the way, using independent measurements. Validation of the surface skin temperature is a particularly challenging task. The surface skin temperature is not a routinely measured variable. Only the surface air temperature is available from in situ measurements and differences between surface skin and surface air temperatures are a complex function of the surface characteristics, the local time, and the solar flux

The analysis methodology presented here and compared to the better-known variational assimilation technique provides an illustration of a more general approach to the analysis of high-volume, multiwavelength satellite observations that may have great potential. The common practice of isolating one variable at a time from such data sets breaks correlations among the measurements and among the retrieved quantities. The variational assimilation approach goes a step further by obtaining simultaneous retrievals of many quantities from multiple measurements; however, as usually implemented, the variational assimilation still does not account for correlations of variables. The neural network approach is not only able to accommodate strongly nonlinear relationships but also is able to benefit from the correlations to improve the retrievals. The neural network approach also requires much less computation than the variational assimilation approach. That the two methods are conceptually close, as we have shown, puts the neural network approach on the same theoretical foundation as the better-studied variational assimilation methods. However, the fact that a simple neural network has been shown to provide a statistical fit to any function suggests that what the trained network is doing is simulating (statistically) the equations of the physical model, in this case an inverse radiative transfer model. Thus, despite use of statistical methods, the analysis represents a physical model of the relationship of the observations and physical quantities. Consequently, the quality of the results depends on two key factors: the accuracy of the physical model used to calculate the training data set (note that model errors can be accounted for in the retrieval design) and the completeness of the sample of parameter correlations represented by the training data set.

The introduction of first guess information into a neural network scheme is expected to improve applications of this technique and to promote new developments, as a possible alternative to variational assimilation methods, for inversion of geophysical satellite observations for a broad range of applications. Another particularly interesting feature of the neural network technique is its ability to merge information coming from different instruments. This feature of neural networks has been used, for example, by *Prigent et al.* [2001]. This strategy would be an excellent way to use the synergy of all instruments in such missions as Tropical Rainfall Measurement Mission or in the next generation Earth Observing Satellite satellites.

Appendix A: One-Dimensional Variational Assimilation Scheme

This method is described by *Rodgers* [1976] and by *Eyre* [1989]. The unified notation of *Ide et al.* [1997] is adopted. Depending on the approach used to derive the inversion formula, the technique has many names: minimum variance method, least squares fitting, best linear unbiased estimator, variational assimilation, expectation maximization, or maximum probability estimator. All these techniques make the same assumptions (local linearization of the physical model, Gaussian distributions of the variables) so the resulting formulae are the same for each technique.

The Bayesian estimator associated with the Newtonian optimization algorithm is adopted with first guess (or background, in the variational assimilation terminology) information. First, we assume that a current estimate x_n of the geophysical parameters to be retrieved exists, and we compute the next estimate x_{n+1}, x_0 being the first guess x^b. The Newtonian method consists of expanding the modeled observation vector y as a Taylor series about the present value x_n:

$$y(x) = y(x_n) + H(x_n)(x - x_n), \tag{A1}$$

with $H(x)$ representing the partial derivatives of $y(x)$ with respect to the elements of x. In a linear inversion approach this linearization of the physical model equation is done only once about the first guess x^b and there are no iterations. The nonlinear approach of the Newtonian optimization algorithm consists in linearizing about the estimate x_n, which is improved at each optimization step n. This assumes that the first guess $x_0 = x^b$ is sufficiently close to the true solution in order to avoid local minima. A bad first guess can result in an inaccurate solution.

The regression of x given y^o and x^b,

$$\hat{x} = E[x|y^o\ x^b] = \int\int x\ P(x|y^o\ x^b)dy^o dx^b, \tag{A2}$$

is equivalent to the maximum likelihood estimator x that maximizes

$$P(x|y^o\ x^b), \tag{A3}$$

the conditional probability of vector x given measurement y^o and first guess x^b. Let $\hat{x}$ be a estimation of the physical variables. For all measurements y^o,

$$\begin{aligned} &E\left[(x-\hat{x})^2|y^o\ x^b\right] \\ &= E\left[\{x - E[x|y^o\ x^b] + E[x|y^o\ x^b] - \hat{x}\}^2|y^o\ x^b\right] \\ &= E\left[(x - E[x|y^o\ x^b])^2|y^o\ x^b\right] + \left(E[x|y^o\ x^b] - \hat{x}\right)^2 \\ &\geq E\left[(x - E[x|y^o\ x^b])^2|y^o\ x^b\right]. \end{aligned} \tag{A4}$$

The regression is also the best unbiased estimator in the least squares sense.

Using the product rule, we can rewrite the conditional probability in (A3) as

$$P(x|y^o\ x^b) = \frac{P(x, y^o, x^b)}{P(y^o, x^b)} = \frac{P(y^o\ x^b|x)P(x)}{P(y^o, x^b)}, \quad \text{(A5)}$$

which is nothing else than the Bayes theorem [*Loredo*, 1990]. It is often assumed, even if it is not always the case, that y^o and x^b, the direct and virtual (first guess) measurements, are independent. In that case, we can expand the corresponding joint probability distribution functions using the product rule

$$P(x|y^o\ x^b) = \frac{P(y^o|x)P(x^b|x)P(x)}{P(y^o)P(x^b)}. \quad \text{(A6)}$$

We want to maximize this probability with respect to x. If the probability distribution $P(x)$ of the physical variables x is available, it is possible to use it in the general context of Bayesian estimation. If this pdf is Gaussian, this would correspond to the addition of a term $\frac{\partial}{\partial x}[x-\bar{x}]^t\bar{B}^{-1}[x-\bar{x}]$ in (A8), where $\bar{x}$ is the mean state of the physical variables and $\bar{B}$ is the covariance matrix of the physical variables. This approach is not used in general in variational assimilation.

If no a priori information on the distribution $P(x)$ is available, this distribution is considered to be uniformly distributed (i.e., no information), so this term can be neglected during the maximization process. The two probabilities $P(y^o)$ and $P(x^b)$ are not dependent on x so they can also be neglected. The maximum likelihood estimator is then obtained at the minimum of minus the log of the two remaining probabilities. Assuming that the minimum is unique, the optimal solution is characterized by

$$-\frac{\partial \log\,[P(y^o|x)P(x^b|x)]}{\partial x} = 0. \quad \text{(A7)}$$

These probabilities need to be rewritten in order to extract the independent random variables involved in the model. Note $P(y^o|x) = P(y^o|y(x))$ since the theoretical radiative transfer function y is not a stochastic function. So $P(y^o|x) = P_\eta(y^o - y)$, where P_η is the probability distribution function of the instrumental noise and the forward model error. Furthermore, $P(y^o|x) = P_\eta[H(x_n)(x - x_n) + (y(x_n) - y^o)]$ using relation (A1). Also $P(x^b|x) = P_\varepsilon(x^b - x)$ where P_ε is the probability distribution function of the first guess error ε.

Assuming that the errors in the observations, the direct radiative transfer model, and the a priori first guess information are unbiased, uncorrelated, and have Gaussian distributions, expression (A7) is equivalent to

$$0 = \frac{\partial}{\partial x}[H(x_n)(x - x_n) + (y(x_n) - y^o)]^t(E + F)^{-1}$$
$$[H(x_n)(x-x_n)+(y(x_n)-y^o)]+\frac{\partial}{\partial x}[x-x^b]^tB^{-1}[x-x^b], \quad \text{(A8)}$$

where B is the first guess error covariance matrix $< (x - x^b) \cdot (x - x^b)^t >$ (estimated with a data set of couples $(x^e, x^{b^e}), e = 1, \ldots, N$); E is the observation error covariance matrix; and the covariance matrix F represents the radiative transfer model errors. We recognize in the first term of this criterion a Mahalanobis distance between $y(x)$ and y^o introduced in (2) and (3).

Relation (A8) is equivalent to

$$\Leftrightarrow H(x_n)^t(E + F)^{-1}[H(x_n)(x - x_n) + (y(x_n) - y^o)] + B^{-1}[x - x^b] = 0. \quad \text{(A9)}$$

We then expand the last term involving the first guess information by introducing the variable x_n,

$$\Leftrightarrow H(x_n)^t(E + F)^{-1}[H(x_n)(x - x_n) + (y(x_n) - y^o)] + B^{-1}[x - x_n] + B^{-1}[x_n - x^b] = 0 \quad \text{(A10)}$$

$$\Leftrightarrow [H(x_n)^t(E + F)^{-1}H(x_n) + B^{-1}][x - x_n] = -[H(x_n)^t(E + F)^{-1}(y(x_n) - y^o) + B^{-1}[x_n - x^b]]. \quad \text{(A11)}$$

Finally, the optimization steps for defining the maximum likelihood estimator are of the form

$$x_{n+1} = x_n - [H(x_n)^t(E + F)^{-1}H(x_n) + B^{-1}]^{-1} [H(x_n)^t(E + F)^{-1}(y(x_n) - y^o) + B^{-1}[x_n - x^b]], \quad \text{(A12)}$$

and the error covariance of the nth step is given by

$$A(x_n) = (B^{-1} + H(x_n)^T(E + F)^{-1}H(x_n))^{-1}. \quad \text{(A13)}$$

Matrix $A(x_n)$ is a theoretical estimate of the error covariance of the solution at step n [*Rodgers*, 1976]. For the final vector $\hat{x}$ and the corresponding $H(\hat{x})$, the matrix A is not a measure of the absolute accuracy of the retrieval but an estimate of the error covariance of the retrieval, valid if all assumptions made for formula (A12) are valid. Within the iterative process, some of the geophysical variables to be retrieved are generally constrained to physically meaningful values (essentially, they should not turn negative in our case).

In the retrieval process, the balance between the information coming from the virtual (first guess) and the direct measurements is implicitly controlled by the covariance matrix of the uncertainties associated with the two pieces of information. If these matrices are not sufficiently precise, or if the variability of the matrices with atmospheric situations is not sufficiently sampled, an "empirical" weight has to be determined.

Notation

x	vector of physical variables to retrieve.
$\hat{x}$	estimate of x.
x^b	first guess a priori information for x.
x_n	nth estimate of x in variational assimilation method.
ε	$= x^b - x$, first guess error.

$y(x)$ radiative transfer function for the physical variable x (also a vector).
y^o SSM/I brightness temperature observations.
η SSM/I instrumental noise.
P generic probability measure.
$P_\eta(\eta)$ probability distribution function of η.
$P_\varepsilon(\varepsilon)$ probability distribution function of ε.
$H(x)$ derivative of y with respect to x.
$A(x)$ covariance matrix of retrieval error estimates in variational assimilation method.
B $=< \varepsilon^t \cdot \varepsilon >$, covariance matrix of the first guess errors.
E $=< \eta^t \cdot \eta >$, covariance matrix of the measurement errors.
F covariance matrix of the radiative transfer model errors.
$E[\cdot]$ expectation operator.
a_i activity of neuron i.
σ sigmoid function of the neural network.
z_i output of the neuron i.
w_{ij} synaptic weight between neuron i and neuron j.
g_W neural network model, or transfer function for our application.
W $= \{w_{ij}\}$, the set of the parameters of the neural network.
y_i neural network input value on neuron i.
x_k neural network output value on neuron k.
$\mathcal{B}$ data set sampling the probability distribution functions.
D generic distance.
D_E Euclidean distance.
$C_1(W)$ theoretical quality criterion for classical neural network learning phase.
$\tilde{C}_1(W)$ practical quality criterion for classical neural network learning phase.
$C_2(W)$ theoretical quality criterion for classical neural network learning phase with first guess information.
$\tilde{C}_2(W)$ practical quality criterion for classical neural network learning phase with first guess information.

Acknowledgments. The authors would like to thank Frédéric Chevallier and Jean-Noel Thépaut for their helpful comments and Alain Chédin for fruitful discussions. This work was partly supported by special NASA Climate and Radiation Program funding provided by Robert J. Curran and now by Donald Anderson.

References

Aires, F., R. Armante, A. Chédin, and N.A. Scott, Surface and atmospheric temperature retrieval with the high resolution interferometer IASI, *Proc. Am. Meteorol. Soc., 98,* 181–186, 1998.

Aires, F., M. Schmitt, N.A. Scott, and A. Chédin, The weight smoothing regularization for Jacobian stabilization, *IEEE Trans. Neural Networks, 10*(6), 1502–1510, 1999.

Basist, A., N.C. Grody, T.C. Peterson, and C.N. Williams, Using the Special Sensor Microwave /Imager to monitor land surface temperatures, wetness, and snow cover, *J. Atmos. Sci., 37,* 888–911, 1998.

Chaboureau, J.-P., A. Chédin, and N.A. Scott, Remote sensing of the vertical distribution of atmospheric water vapor from the TOVS observations. Method and validation, *J. Geophys. Res., 103,* 8743–8752, 1998.

Chevallier, F., J.-J. Morcrette, F. Chéruy, and N.A. Scott, Use of a neural network-based longwave radiative transfer scheme in the ECMWF atmospheric model, *Q. J. R. Meteorol. Soc., 126,* 761-776, 2000.

Colton, M.C., and G.A. Poe, Intersensor calibration of DMSP SSM/I's: F-8 to F-14, 1987-1997, *IEEE Trans. Geosci. Remote Sens., 37,* 418-439, 1999.

Crone, L., and D. Crosby, Statistical applications of a metric on subspaces to satelite meteorology, *Technometrics, 37,* 324-328, 1995.

Cybenko, G., Approximation by superpositions of a sigmoidal function, *Math. Control Signals Syst., 2,* 303–314, 1989.

Derber, J., and F. Boutier, A reformulation fo the background error covariance in the ECMWF global data assimilation system, *Tellus, 51A,* 195–221, 1999.

Duflo, M., Algorithmes stochastiques, in *Mathématiques et Applications*, Springer-Verlag, New York, 1996.

English, S.J., Estimation of temperature and humidity profile information from microwave radiances over different surface types, *J. Appl. Meteorol., 38,* 1526-1541, 1999.

Escobar, J., A. Chédin, F. Chéruy, and N.A. Scott, Réseaux de neurones multicouches pour la restitution de variables thermodynamiques atmosphériques à l'aide de sondeurs verticaux satellitaires, *C. R. Acad. Sci. Paris, 317*(2), 911–918, 1993.

Eyre, J.R., Inversion of cloudy satellite sounding radiances by nonlinear optimal estimation, I, Theory and simulation for TOVS, *Q. J. R. Meteorol. Soc., 115,* 1001–1026, 1989.

Eyre, J.R., G.A. Kelly, A.P. MacNelly, E. Anderson, and A. Persson, Assimilation of TOVS radiance information through one-dimensional variational analysis, *Q. J. R. Meteorol. Soc., 119,* 1427–1463, 1993.

Gadd, A.J., B.R. Barwell, S.J. Cox, and R.J. Renshaw, Global processing of satellite radiances in a numerical weather prediction system, *Q. J. R. Meteorol. Soc., 121,* 615–630, 1995.

Greenwald, T.J., C.L. Combs, A.S. Jones, D.L. Randel, and T.H. Vonder Haar, Further developments in estimating cloud liquid water over land using microwave and infrared satellite measurements, *J. Appl. Meteorol., 36,* 389–405, 1997.

Hollinger, J.P., R. Lo, G. Poe, R. Savage, and J. Pierce, Special Sensor Microwave/Imager user's guide, Nav. Res. Lab., Washington, D. C., 1987.

Hollinger, J.P., J.L. Pierce, and G.A. Poe, SSM/I instrument evaluation, *IEEE Trans. Geosci. Remote Sens., 28,* 781-790, 1990.

Hornik, K., M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, *Neural Networks, 2,* 359–366, 1989.

Ide, K., P. Courtier, M. Ghil, and A.C. Lorenc, Unified notation for data assimilation: Operational, sequential and variational, *J. Meteorol. Soc. J., 75,* 181-189, 1997.

Jones, A. S., and H. T. Vonder Haar, Passive microwave remote sensing of cloud liquid water over land regions, *J. Geophys. Res., 95,* 16,673–16,683, 1990.

Kalnay, E.. et al. The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc., 77,* 437-470, 1996.

Krasnopolsky, V.M., L.C. Breaker, and G.H. Gemmill, A neural network as a nonlinear transfer function model for retrieving surface wind speeds from the special sensor microwave imager, *J. Geophys. Res., 100,* 11,033–11,045, 1995.

Krasnopolsky, V.M., G.H. Gemmill, and L.C. Breaker, A neural network multiparameter algorithm for SSM/I ocean retrievals: Comparisons validations, *Remote Sens. Environ.*, *73*, 133–142, 2000.

Liebe, H.J., G.A. Hufford, and M.G. Cotton, Propagation modeling of moist air and suspended water/ice particles at frequencies below 1000 GHz, paper presented at Specialist Meeting of the Electromagnetic Wave Propagation Panel, Adv. Group for Aerosp. Res. and Dev., Palma de Mallorca, Spain, 1993.

Lin, B., and W.B. Rossow, Observations of cloud liquid water path over oceans: Optical and microwave remote sensing methods, *J. Geophys. Res.*, *99*, 20,907–20,927, 1994.

Lin, B., and W.B. Rossow, Precipitation water path and rainfall rate estimates over oceans using special sensor microwave imager and International Satellite Cloud Climatology Project data, *J. Geophys. Res.*, *102*, 9359-9374, 1997.

Loredo, T.J., From Laplace to supernove SN 1987A: Bayesian inference in astrophysics, in *Maximum Entropy and Bayesian Methods*, edited by P.F. Fougere, pp. 81–142, Kluwer Acad., Norwell, Mass., 1990.

MacFarland, J.M., R.L. Miller, and C.M.U. Neale, Land surface temperature derived from the SSM/I passive microwave brightness temperatures, *IEEE Trans. Geosci. Remote Sens.*, *28*, 839–845, 1990.

Manabe, T., H.J. Liebe, and G.A. Hufford, Complex permittivity of water between 0 and 30 THz, in *12th International Conference on Infrared and Millimeter Waves*, pp. 229-230, IEEE Press, Piscataway, N.J., 1987.

Njoku, E.G., Surface temperature estimation over land using satellite microwave radiometry, in *Passive Microwave Remote Sensing of Land-Atmosphere Interaction*, edited by B.J. Choudhury, et al., pp. 509–530, Vsp, Utrecht, Netherlands, 1995.

Pardo, J.R., J. Cernicharo, and E. Serabyn, Atmospheric Transmission at Microwaves (ATM): An Improved Model for mm/submm applications, *IEEE Trans. on Antennas and Propagation*, in press, 2001.

Phalippou, L., Variational retrieval of humidity profile, wind speed and cloud liquid water path with SSM/I: Potential for numerical weather prediction, *Q. J. R. Meteorol. Soc.*, *122*, 327–355, 1996.

Prigent, C., and W.B. Rossow, Retrieval of surface and atmospheric parameters over land from SSM/I: Potential and limitations, *Q. J. R. Meteorol. Soc.*, *125*, 2379–2400, 1999.

Prigent, C., W.B. Rossow, and E. Matthews, Microwave land surface emissivities estimated from SSM/I observations, *J. Geophys. Res.*, *102*, 21,867–21,890, 1997.

Prigent, C., W.B. Rossow, and E. Matthews, Global maps of microwave land surface emissivities: Potential for land surface characterization, *Radio Sci.*, *33*, 745–751, 1998.

Prigent, C., A. Aires, W.B. Rossow, and E. Matthews, Joint characterization of the vegetation by satellite observations from visible to microwave wavelengths: A sensitivity analysis, *J. Geophys. Res.*, in press, 2001.

Rabier, F., A. McNally, E. Andersson, P. Courtier, P. Undén, J. Eyre, A. Hollingsworth, and F. Bouttier, The ECMWF implementation of three-dimensional variational assimilation (3D-Var), II, Structure functions, *Q. J. R. Meteorol. Soc.*, *124*, 1809-1829, 1998.

Rodgers, C.D., Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation, *Rev. Geophys.*, *14*, 609-624, 1976.

Rosenkranz, P.W., Absorption of microwave by atmospheric gases, in *Atmospheric Remote Sensing by Microwave Radiometry*, edited by M.A. Janssen, Wiley-Interscience, New York, 1992.

Rossow, W. B., and L. C. Garder, Validation of ISCCP cloud detections, *J. Clim.*, *6*, 2370–2393, 1993.

Rossow, W.B., and R.A. Schiffer, ISCCP cloud data products, *Bull. Am. Meteorol. Soc.*, *72*, 2–20, 1991.

Rossow, W.B., and R.A. Schiffer, Advances in understanding clouds from ISCCP, *Bull. Am. Meteorol. Soc.*, *80*(11), 2261–2287, 1999.

Rossow, W.B., A.W. Walker, D.E. Beuschel, and M.D. Roiter, International Satellite Cloud Climatology Project (ISCCP): Document on new cloud datasets, Natl. Aeronaut. and Space Admin., Goddard Inst. for Space Stud., New York, 1996.

Rumelhart, D.E., G.E. Hinton, and R.J. Williams, Learning Internal Representations by Error Propagation, in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. I, Foudations*, edited by D.E. Rumelhart, J.L. McClelland, and the PDP Research Group, pp. 318–362, MIT Press, Cambridge, Mass., 1986.

Stogryn, A.P., C.T. Butler, and T.J. Bartolac, Ocean surface wind retrievals from special sensor microwave imager data with neural networks, *J. Geophys. Res.*, *99*, 981–984, 1994.

Thépaut, J.-N., R.N. Hoffman, and P. Courtier, Interactions of dynamics and observations in a four-dimensional variational assimilation, *Mon. Weather Rev.*, *121*(112), 3393–3414, 1993.

Vapnik, V., *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1997.

F. Aires and C. Prigent, Department of Applied Physics, Columbia University, NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025. (faires@giss.nasa.gov; cprigent@giss.nasa.gov)

W. B. Rossow, NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025. (wrossow@giss.nasa.gov)

M. Rothstein, Science Systems and Application, Inc., NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025. (mrothstein@giss.nasa.gov)

(Received July 31, 2000; revised December 21, 2000; accepted January 7, 2001.)