Radio Science, Volume 33, Number 5, Pages 1373-1386, September-October 1998

Neural networks for the retrieval of water vapor and liquid water from radiometric data

Fabio Del Frate and Giovanni Schiavon
Dipartimento di Informatica, Sistemi e Produzione, Università Tor Vergata, Rome, Italy

Abstract. This paper investigates the potentiality of neural networks for the retrieval of integrated water vapor and integrated liquid water from data simulating the measurements of different ground-based microwave radiometers, with the optional addition of a laser ceilometer. The reliability of the neural network algorithms was evaluated comparing their performance with that achievable by linear regression techniques with the same data sets. The obtained results showed that the neural inversion provides a more accurate estimation of the parameters to be retrieved, especially in the case of strong nonlinearities. In addition, neural networks succeed in exploiting the information given by the ceilometer significantly better than linear regression. The advantages shown by the neural inversion led the investigation further, aiming at the optimization of the architecture of the net and focusing on the number of processing units and connections. An optimum range for the choice of the number of neurons to be inserted in the net has been determined, and ineffective connections have been removed via pruning algorithms. A fault tolerance analysis was also performed, which pointed out other interesting properties of the neural retrieval procedure. In conclusion, the results of this simulation indicate that neural networks, if compared with linear regressions, are more suitable for cases with stronger nonlinearities and are more flexible and robust algorithms. Moreover, their performance can improve when the number of units and connections is optimized.

1. Introduction

Atmospheric water vapor is an important quantity in several areas, including weather prediction, very long baseline interferometry, geodetic metrology, and radar altimetry. Cloud liquid is important in weather modification, detection of conditions relevant to aircraft icing, and in studies of the influence of clouds on solar radiation. Both quantities have a high temporal and spatial variability, which makes it difficult to estimate them from surface humidity data. Under nonprecipitating conditions, the thermal radiation of the troposphere at microwave frequencies originates primarily from oxygen, water vapor, and liquid water (clouds). This allows the estimation of the integrated water vapor (V) and of the integrated liquid water (L) by means of ground-based microwave radiometer measurements [*Hogg et al.*, 1983; *Peter and Kämpfer*, 1992]. Futhermore, the microwave technique is potentially useful in applications, such as nowcasting or mesoscale forecasting, requiring a high time resolution or the collection of many measurements in data-sparse regions of the oceans. Dual-channel radiometers have been operating for several years to measure V and L. The retrievals have been generally carried out by the linear statistic method [*Westwater and Strand*, 1968], and the accuracies are satisfactory for several applications.

However, a first source of error seems to be the dependence of the liquid absorption coefficient on temperature [*Westwater*, 1978]. One way of overcoming this difficulty can be the use of radiometers with more than two channels so that the thermal state of the atmosphere can also be taken into account from the multifrequency data [*Askne et al.*, 1985]. Also, a laser ceilometer can be considered to enhance the measurements vector [*Han and Westwater*, 1995]. Moreover, the errors of liquid and, in particular, of vapor retrievals appear to increase dramatically with increasing amounts of liquid, i.e., when the nonlinear character of the inversion problem becomes preeminent [*Westwater and Guiraud*, 1980].

Recently, artificial neural networks have been recognized as being useful for retrieval operations in remote

Paper number 98RS02133.
0048-6604/98/98RS-02133$11.00

sensing of the atmosphere [*Churnside et al.*, 1994; *Cabrera and Staelin*, 1995; *Li et al.*, 1997; *Del Frate and Schiavon*, 1998]. The use of neural networks in statistical estimation is often effective because they can simultaneously address nonlinear dependencies and complex statistical behavior. Indeed, it has been shown that a multilayer perceptron [*Beale and Jackson*, 1990] with a single hidden layer and nonlinear activation functions is capable of approximating any real-valued continuous function, provided a sufficient number of units within this hidden layer exists [*Hornik et al.*, 1989]. While a general multilayer perceptron can form arbitrary mappings, in practice, finite network size demands that only an approximation to the optimal solution can be achieved.

In this paper we consider a multilayer perceptron for determining V and L from simulated radiometric data. The main objective is to obtain a nonlinear regression surface that represents the desired mapping and to compare its performance with that of a classical linear technique. A second goal, encouraged by this comparison, is the optimization of some of the most significant parameters of the neural algorithm such as the number of units in the hidden layer of the net or the number of its connections, where, for this last step, a pruning procedure can be used. The properties of robustness of the proposed technique are also examined. We simulated, for each channel of the radiometer, an undetected fault, and we analyzed the effects on the accuracy of the retrieval.

2. Basic Principles and the Linear Approach

For a nonscattering atmosphere in local thermodynamic equilibrium and for a given frequency in the microwave region, the brightness temperature T_B measured by an upward looking radiometer is given by the following expression [*Janssen*, 1993]:

$$T_B = \int_0^\infty T(s)\alpha(s)e^{-\int_0^s \alpha(s')ds'}ds + T_{\cos}e^{-\int_0^\infty \alpha(s')ds'} \tag{1}$$

where $T(s)$ (kelvins) and $\alpha(s)$ (per kilometer) are the atmospheric physical temperature and the absorption coefficient at the spatial coordinate s, respectively, while $T_{\cos}$ (kelvins) is the brightness temperature contributed by extraterrestrial sources (cosmic background). The absorption coefficient $\alpha(s)$ is, in general, the sum of the absorption contributions of dry air, water vapor, and clouds,

$$\alpha(s) = \alpha_d(s) + \alpha_v(s) + \alpha_l(s) \tag{2}$$

The water vapor and the cloud absorption coefficients are, in turn, approximately proportional to the water vapor density $\rho_v(s)$ and the liquid density $\rho_l(s)$, so that we can write

$$\alpha_v(s) = k_v\rho_v(s) \tag{3}$$

and

$$\alpha_l(s) = k_l\rho_l(s) \tag{4}$$

where k_v and k_l are the specific (or mass) absorption coefficients of water vapor and liquid, respectively. On the other hand, V and L are related to $\rho_v(s)$ and $\rho_l(s)$ by

$$V = \int_0^\infty \rho_v(s)ds \tag{5}$$

$$L = \int_0^\infty \rho_l(s)ds \tag{6}$$

The integration of (2) over the entire atmosphere, with the introduction of the optical depth $\tau(s)$, given by

$$\tau(s) = \int_0^s \alpha(s')ds' \tag{7}$$

and of the opacity $\tau \doteq \tau(\infty)$, gives

$$\tau = \tau_d + k_vV + k_lL \tag{8}$$

where τ_d is the path integral of $\alpha_d(s)$. The definition of the mean radiating temperature (T_{mr})

$$T_{\mathrm{mr}} = \frac{\int_0^\infty T(s)\alpha(s)e^{-\tau(s)}ds}{\int_0^\infty \alpha(s)e^{-\tau(s)}ds} \tag{9}$$

allows one to derive the opacity τ from (1) by

$$\tau = -\ln\left(\frac{T_{\mathrm{mr}} - T_B}{T_{\mathrm{mr}} - T_{\cos}}\right) \tag{10}$$

Therefore it turns out that a relation exists between the measured brightness temperatures and integrated vapor and liquid and that the inversion is feasible.

In the linear statistical approach the estimation of V and L is commonly obtained operating in two steps [*Westwater*, 1993]. In the first step an estimate of the mean radiating temperature is performed. For each frequency the corresponding T_{mr} can be obtained using a linear regression (LR)

$$T_{\mathrm{mr}} = a_0 + \sum_{n=1}^{N_f} a_nT_{Bn} \tag{11}$$

where N_f is the number of the available radiometric channels and T_{Bn} is the brightness temperature measured

by the nth channel. In the second step the opacity τ, at each frequency, is evaluated using (10). Afterward, the estimate of V and L is performed through the following LRs:

$$V = b_0 + \sum_{n=1}^{N_f} b_n \tau_n \tag{12}$$

$$L = c_0 + \sum_{n=1}^{N_f} c_n \tau_n \tag{13}$$

The coefficients a_i, b_i, and c_i in (11), (12), and (13) are obtained by applying the linear regression analysis [*Hogg et al.*, 1983] to a statistically significant data set. The data sets used in this study will be described in section 3.2.

The linear statistical approach provides unsatisfactory results in the presence of clouds with high liquid content when the nonlinearities of the inversion problem cannot be neglected anymore. Under this point of view the behavior of the neural technique is completely different. In fact, in this case, data are processed by units that are inherently nonlinear.

3. Simulation Setup

3.1. The Neural Net Algorithm

An artificial neural network (NN) may be viewed as a mathematical model composed of many nonlinear computational elements, named neurons, operating in parallel and massively connected by links characterized by different weights. A single neuron computes the sum of its inputs, adds a bias term, and drives the result through a generally nonlinear activation function to produce a single output termed the activation level of that neuron. NN models are mainly specified by the net topology, neuron characteristics, and training or learning rules [*Lippman*, 1987].

The term topology refers to the structure of the network as a whole: the number of its input, output, and hidden units and how they are interconnected. For this study, multilayer perceptrons (MLP) have been considered, which have been found to have the best suited topology for classification and inversion problems [*Hsu et al.*, 1992]. These are feedforward networks where the input flows only in one direction to the output, and each neuron of a layer is connected to all neurons of the successive layer but has no feedback to neurons in the previous layers. Two layers of weights have been used in our case.

The individual neuron is the elemental building block of each layer, and it is mainly characterized by its activation function. The most common activation function is the nonlinear sigmoid function, also used in our simulations, defined as follows:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{14}$$

Such an activation function yields values in the range [0,1]. Since the output units of the mapping network must generally produce an estimate of a parameter with an arbitrary range, the range restriction described above must be removed. A preprocessing of the input variables is commonly performed as well. In particular, this is useful if different input variables have typical values which differ significantly and do not reflect the relative importance in determining the required outputs. For these reasons we scaled the input and output values so that they were between 0.01 and 0.99.

A MLP is designed to approximate an unknown input-output relation by determining the weight or strength of each connection via learning rules. These rules indicate how to pursue minimization of the error function measuring the quality of the network's approximation on the restricted domain covered by a training set (i.e., a set of input-output examples). The ultimate goal, however, is to minimize the error for all possible examples related through the input-output relation, namely, to generalize outside of the training set. In our case the net was trained using the back propagation algorithm [*Rumelhart et al.*, 1986], which uses a gradient search technique and iteratively adjusts the weight coefficients in the network to minimize an error function equal to the mean square difference between the desired and the actual net output. For determining when the training procedure had to be stopped, we considered the "early stopping" algorithm [*Bishop*, 1995]. According to this algorithm, the performance of the net during the training (learning) phase is evaluated either on the training set or on a different independent validation set. In the training set the overall error in the retrieval of the correct output keeps on decreasing with the training, approaching a value of convergence. Conversely, the error on the validation set reaches a minimum value, after which it will start increasing if we continue the training. At this point the learning phase must be interrupted.

The software simulations were performed by means of the Stuttgart neural network simulator (SNNS), developed at the University of Stuttgart (Germany) [*Zell et al.*, 1995], which proved to be a high-level, flexible, and reliable software package. The complete training process seldom took more than 1 hour of a medium speed CPU.

3.2. Data

Two statistically independent ensembles (training and evaluation) of atmospheric profiles were generated [*Schiavon et al.*, 1993] starting from the midlatitude summer standard atmosphere [*Damosso et al.*, 1983]. For temperature, at ground level a random fluctuation with a Gaussian distribution has been superimposed on its standard value. Random fluctuations were also added to the temperature values at different heights, but with a Γ distribution [*Soong*, 1982]. These fluctuations were added to a projected adiabatic profile to avoid unphysical situations. Ground-based thermal inversions with Γ-distributed thickness and strength were included, too.

The profiles of water vapor have been constructed by adding random irregularities with a Gaussian distribution to the standard atmosphere profile. Liquid water simulating fog and/or clouds was generated whenever the relative humidity at a given height was larger than a selected threshold. Both the thickness of the fog layer and the base height of the clouds were Γ-distributed, and the same distribution controlled the statistical generation of cloud thickness. The liquid content of clouds has been assumed to be proportional to their thickness [*Decker et al.*, 1978]. A mixture of liquid and ice has been assumed below 0°C, while only ice is present below −30°C. To better examine a more critical situation, we only considered profiles with clouds.

These synthetic atmospheric profiles were used to compute the brightness temperatures as would be measured at the various frequencies by a ground-based radiometer aiming at zenith. This was made possible by using *Liebe et al.*'s [1993] millimeter-wave propagation model (MPM), which allows one to compute the complex refractivity N of the atmosphere, the imaginary part of which gives the absorption coefficient α required in (1). The calculation of N consists of several additive parts, namely, a frequency independent term plus various spectra of refractive dispersion and absorption. The spectra consist of 44 O_2 and 34 H_2O line contributions; nonresonant spectra of dry air; continuum contributions from the H_2O spectrum above 1 THz, which are formulated as wing responses of a pseudoline centered at 1.8 THz; and refractivity formulation for suspended water droplets and ice crystals, whose effects are treated by means of the Rayleigh approximation. Inputs to the model are barometric pressure, temperature, and relative humidity, which are profiles data, and density of suspended water droplets and ice crystals, which has been evaluated from relative humidity profiles according to *Decker et al.*'s [1978] model.

The use of a propagation model combined with a profile generation algorithm allows one to carry out sensitivity studies between brightness temperature and atmospheric components [*Dawson*, 1994; *Li et al.*, 1997]. It also allows one to cover the full range of variability of atmospheric parameters within established bounds for a good training of the retrieval algorithm, which could be difficult to obtain with measured data.

To simulate noise in the radiometric channels, random fluctuations with Gaussian distribution and varying standard deviations have been added to the brightness temperature data. The results presented here refer to three levels of rms radiometric noise, i.e., 0.5, 1.0, and 2.0 K, which are representative of the typical characteristics of existing radiometers.

Sets of actual radiosonde-measured profiles were considered as well. They belong to various locations in Italy and were used to validate some of the results obtained for the synthetic atmospheres.

The input quantities of the net simulate data measured by three types of instruments. The first one is a multifrequency radiometer which was considered to simulate the retrieval capabilities of an actual new generation radiometer for remote sensing of the atmosphere designed and developed under a project carried out by the European Space Agency (ESA) [*Battistelli et al.*, 1995]. The instrument measures brightness temperatures at the following seven frequencies: 22.23, 23.87, 31.65, 51.25, 52.85, 53.85, and 54.85 GHz. Indeed, although the lowest (22.2 GHz) water vapor absorption line is fairly weak, brightness temperatures around this frequency contain some information on the amount and height distribution of the tropospheric water vapor. On the other side, measurements in the 30- to 40-GHz atmospheric window carry information on the amount of liquid water present in nonprecipitating clouds. Finally, the brightness at frequencies within the 60-GHz oxygen absorption complex depends on the temperature of the lower layers of the troposphere. As a consequence, a radiometric system operating at these seven channels is expected to display a profiling capability, which could have beneficial results both for remote sensing of the troposphere and for propagation studies.

We then simulated measurements of a typical instrument used for the estimation of V and L. This is a dual-channel radiometer whose work frequencies are at 23.87 and 31.65 GHz. The lower frequency, on one shoulder of the water vapor absorption line, yields measurements that are nearly independent from the height distribution of water vapor, while the upper frequency is particularly sensitive to the liquid.

Table 1. Input and Output Statistics for the Two Data Sets

Parameter	Simulated Data		Radiosonde Data	
	Mean	s.d.	Mean	s.d.
	Inputs, K			
$T_{B_{22.23}}$	106.2	18.5	90.2	20.5
$T_{B_{23.87}}$	100.8	18.9	86.2	20.4
$T_{B_{31.65}}$	101.5	25.7	89.5	25.7
$T_{B_{51.25}}$	209.1	23.1	196.1	23.7
$T_{B_{52.85}}$	246.7	12.5	236.9	13.6
$T_{B_{53.85}}$	275.7	4.7	269.2	6.1
$T_{B_{54.85}}$	288.3	3.2	284.1	4.6
	Outputs, cm			
V	4.06	0.891	3.08	0.836
L	0.21	0.084	0.18	0.083

A third instrument that we took into account is a laser ceilometer. This instrument is designed to measure cloud-base height detecting the time (and thus the corresponding distance) needed for a short pulse of light to traverse the atmosphere from the transmitter of the ceilometer to a backscattering cloud base and back to the receiver of the ceilometer. We assumed a vertical resolution of 15 m.

In the remainder of this paper we consider data related to four systems: the two radiometers with and without the ceilometer. More emphasis will be put on the multifrequency radiometer because the additional pieces of information that the 50-GHz-band channels provide could be beneficial to the retrieval of both V and L [*Schiavon et al.*, 1993]. In particular, cloud liquid water absorp-

Table 2. RMS Errors in the Retrieval of V and L With a Multifrequency Radiometer (Synthetic Profiles)

Retrieval	V RMS Error, cm			L RMS Error, cm		
	Radiometric Noise, K			Radiometric Noise, K		
Method	0.5	1.0	2.0	0.5	1.0	2.0
	Input: T_B					
NN1	0.0636	0.1053	0.1782	0.0056	0.0071	0.0091
NN2	0.0650	0.1041	0.1810	0.0061	0.0076	0.0096
LR	0.1033	0.1344	0.1973	0.0106	0.0116	0.0128
	Input: τ					
NN1	0.0619	0.1032	0.1804	0.0055	0.0073	0.0091
NN2	0.0834	0.1181	0.1902	0.0060	0.0075	0.0095
LR	0.0740	0.1134	0.1854	0.0066	0.0083	0.0102

Abbreviations are NN1, two nets, each with a single output; NN2, single net with two components in the output vector; and LR, linear regression.

tion depends on its physical temperature, and therefore a temperature profiling capability can especially improve the retrieval of L. The inclusion of the ceilometer measurement, which precisely locates the cloud and therefore determines its temperature if information on temperature profile is present, can further enhance the estimation of liquid.

The range for the values of L varies from 0.1 to 0.5 cm. This is an interesting interval for comparing the performance of NN and LR algorithms. In fact, it spans a range in which on one side the statistical regression can still be reckoned a valid technique, but on the other side the presence of nonlinearities starts to be meaningful, even though it is bounded by the nonprecipitation condition. In Table 1 the statistics of the brightness temperatures and of V and L considered in the simulation are reported either for the set where profiles have been statistically generated or for the set based on radiosonde data. The two sets have quite similar characteristics. The discrepancies between the mean values are mainly due to the fact that the synthetic set, as already described, was generated starting from a midlatitude summer profile, while the radiosonde set included some profiles measured in other seasons.

4. Retrieval Results

In Table 2 we report the rms error of the retrieval of V and L computed in different cases for the set of the statistically generated profiles. For either NN or LR we considered as input quantities the brightness temperatures and the optical depths obtained by the simulated measurements of the seven-frequency radiometer. We also considered the three already mentioned levels of noise for the radiometric channels. As far as the retrieval with the NN is concerned, parameter inversion can be performed either by using one single net with two components in the output vector (NN2 type net) or by using two different nets, each with a single output for either V or L (NN1 type nets). Results are reported for both types of net configuration.

Table 2 shows that, generally, the performance of the NN is better than that of LR, particularly for noise levels of 0.5 and 1.0 K. It can be observed that in the case of LR the role played by τ, whose computation involves the estimation of T_{mr}, is crucial for the retrieval of the atmospheric parameters. In fact, even if a more complicated processing of the input data is required, a decrease of the rms error up to 35% is obtained; this does not occur if the estimation is performed using the NN, which seems to be able to catch directly from the brightness temperature all the information necessary to retrieve the desired quantities. Hence, in this latter case, the computation of τ is not needed.

Table 3. RMS Errors in the Retrieval of V and L With a Multifrequency Radiometer (Radiosonde Profiles)

	RMS Error, cm	
	V	L
	Input: T_B	
NN1	0.146	0.0091
LR	0.149	0.0142
	Input: τ	
NN1	0.142	0.0090
LR	0.147	0.0099

Radiometric noise is of 1.0 K standard deviation. Abbreviations are NN1, two nets, each with a single output, and LR, linear regression.

It can also be noted that NN1 type nets perform generally better than NN2 type. In fact, although atmospheric radiation at millimeter waves depend on both atmospheric water vapor and cloud liquid, the dependence relation between each retrieved quantity and the input vector (brightness temperature or opacity) is different. If we consider both quantities in the output vector (NN2 type net), the coefficients determined for the net during the training phase do have to represent the two dependence relations. Putting only one element in the output vector of the two NN1 type nets allows one to calculate coefficients that are more specific for the single retrieval problem, therefore generally performing a more accurate approximation of the sought input-output function. For this reason, in the following we shall limit our analysis to NN1 type nets.

The general trend of the preceding results is validated by radiosonde data. In Table 3 the rms errors obtained in this case are reported, where a level noise of 1.0 K for the seven radiometric channels is considered.

To further examine the rms errors obtained with NN and LR, some scatterplots are reported in Figures 1 and 2. A comparison between retrieved values (on the vertical axis) and the true values (horizontal axis) obtained integrating the corresponding profiles is shown. In the restricted case (profiles with more liquid) the NN augments its improvement in the accuracy of the estimation with respect to LR. For liquid a decrease of the rms error of 14%

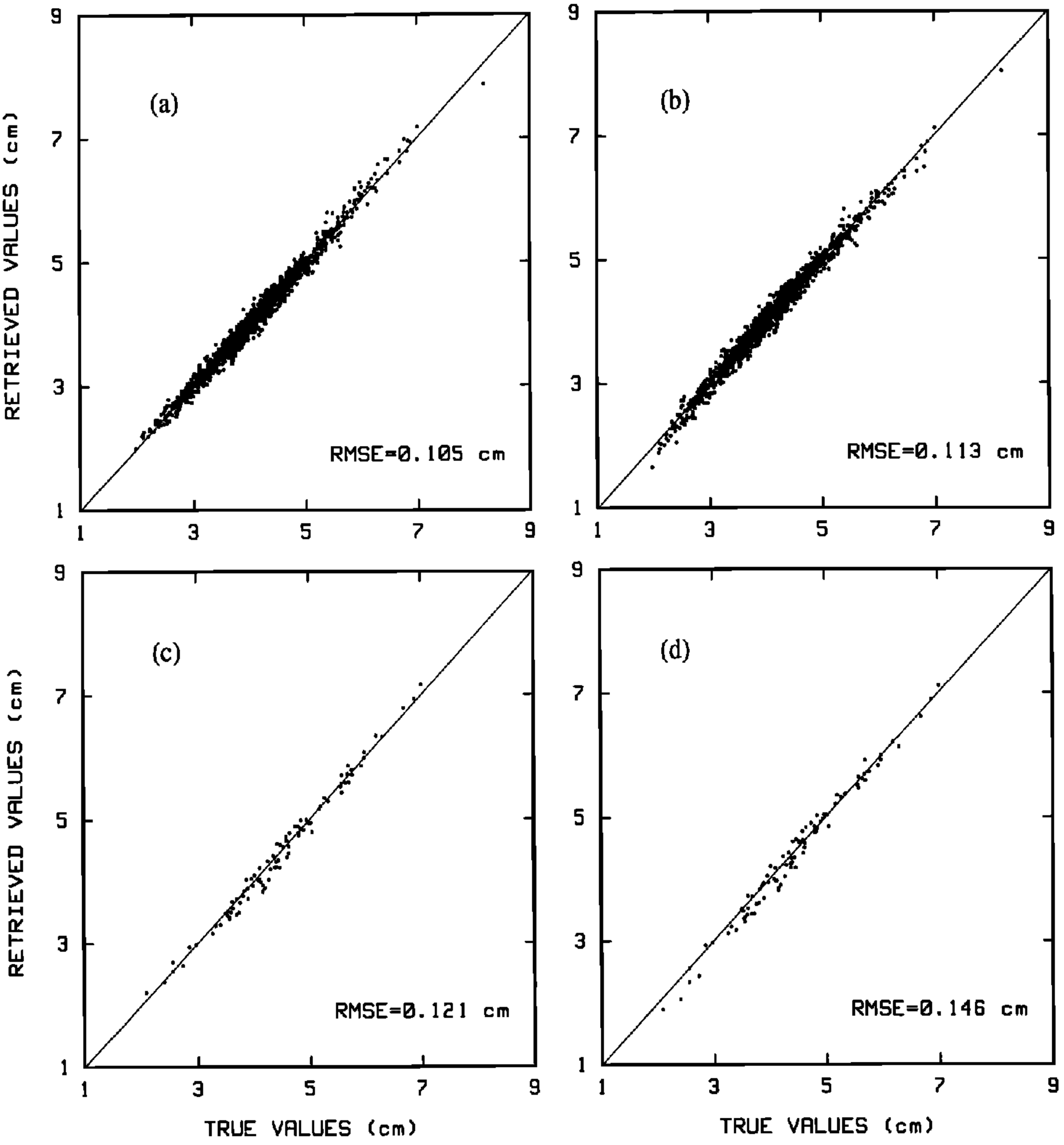

Figure 1. Retrieved versus true values of integrated vapor in the general case (Figures 1a and 1b) and for the case of profiles with an amount of liquid water greater than 0.35 cm (Figures 1c and 1d). Input quantities are the brightness temperatures for the neural networks (Figures 1a and 1c) and the opacities for the linear regression (Figures 1b and 1d). Radiometric noise is of 1.0 K standard deviation.

in the general case is improved to 31% in the restricted case. For water vapor, in the general case the improvement displayed by NN is 8%, while in the restricted case it is 17%. This confirms that the stronger the nonlinearities are in the inversion problem (which correspond to a larger amount of liquid), the better the behavior of NN is compared with LR. The same analysis, either for L or V, has also been carried out considering profiles measured

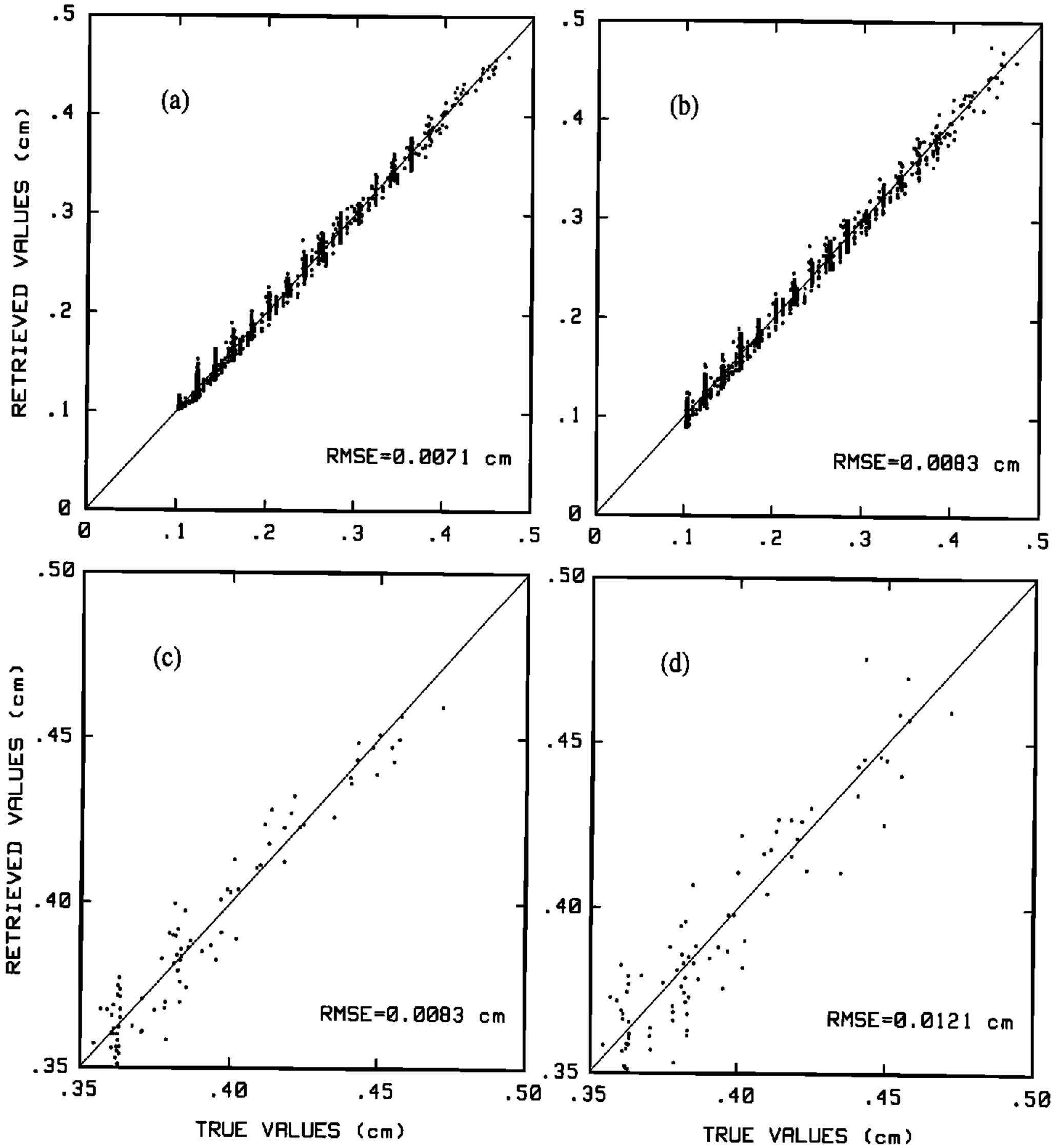

Figure 2. Retrieved versus true values of integrated liquid in the general case (Figures 2a and 2b) and for the case of profiles with an amount of liquid water greater than 0.35 cm (Figures 2c and 2d). Input quantities are the brightness temperatures for the neural networks (Figures 2a and 2c) and the opacities for the linear regression (Figures 2b and 2d). Radiometric noise is of 1.0 K standard deviation.

by radiosondes. These latter results confirm the previous considerations.

In Table 4 the retrieval results, obtained using the two other instruments described in section 3.2, are reported: The ceilometer is supposed to be used in synergy either with the dual-channel radiometer or with the seven-channel radiometer. One more term, $b_{N_f+1}H_c$ or $c_{N_f+1}H_c$, where H_c is the cloud-base height, was added to (12) or (13), respectively, for LR, while one element for H_c was added to the input vector for NN. It can be

Table 4. RMS Errors in the Retrieval of V and L With Four Different Measurement Configurations

	V RMS Error, cm			L RMS Error, cm		
Simulated	Radiometric Noise, K			Radiometric Noise, K		
Instrument	0.5	1.0	2.0	0.5	1.0	2.0
	Neural Network Retrieval Method					
DCR	0.0861	0.1512	0.2730	0.0100	0.0104	0.0117
DCR + C	0.0763	0.1457	0.2705	0.0049	0.0057	0.0079
MFR	0.0636	0.1053	0.1782	0.0056	0.0071	0.0091
MFR + C	0.0505	0.0902	0.1663	0.0026	0.0036	0.0059
	Linear Regression Retrieval Method					
DCR	0.1340	0.1809	0.2909	0.0113	0.0116	0.0128
DCR + C	0.1297	0.1775	0.2880	0.0080	0.0084	0.0099
MFR	0.0740	0.1134	0.1854	0.0066	0.0083	0.0102
MFR + C	0.0956	0.1233	0.1910	0.0047	0.0060	0.0079

Input quantities were the brightness temperatures for the neural network and the opacities for the linear regression. Abbreviations are MFR, multifrequency radiometer; DCR, dual-channel radiometer; and C, ceilometer.

noted that again, the estimation based on NN gives generally better results if compared with LR. This occurs for each of the four simulated sensing configurations. Focusing on the contribution of the ceilometer to the accuracy improvement of the estimates, the NN seems to exploit the information provided by this intrument much better than LR does. In the case of liquid water and for a radiometric noise of 0.5 and 1.0 K of standard deviation, the error is halved using NN, while LR does not show such a good incorporation of the ceilometer measurement. In the case of water vapor, even less accurate results are obtained using LR.

5. Optimization Results and Pruning

We also tried to optimize the NN from the point of view of the number of its adaptive parameters (units and connections). We considered a net which used as inputs the brightness temperature of the seven-channel radiometer, L or V in output, and, desiring to optimize the net in a still more critical range of values of L, we restricted the analysis only to profiles with values of L larger than 0.2 cm.

In Figure 3 the rms error of the retrieval of V and L versus the number of neurons in the single hidden layer is plotted. A noise level of 1.0 K has been assumed. The reported results were obtained by averaging over a number of realizations generated from different initial conditions and randomly varied in time. It can be noted that for both cases an optimum number or at least an optimum range does exist for the number of neurons in the hidden layer. In fact, if the number of neurons is too small, the input-output associative capabilities of the net are too weak. On the other hand, this number should not be too large; otherwise, these capabilities might show a lack of generality (they would be tailored on the training set). It turns out that a fair compromise between these two contrasting requirements has to be found. Hence it is not always true that the more complex the nets are (i.e., the greater the number of neurons is) the better the results are, which is of particular interest for hardware design purposes. For water vapor and for data corrupted with noise of 1.0-K standard deviation, an optimum range for the number of hidden units will be 3–7. For integrated liquid, instead, this study suggests a number of neurons between 2 and 6. We also made some tests with other levels on noise affecting the data, and we generally noted that the lower the level of noise is, the larger the value of the optimum number of neurons is.

As a second step for the optimization of the net we considered a pruning procedure. According to this kind of procedure; a network is examined to assess the rel-

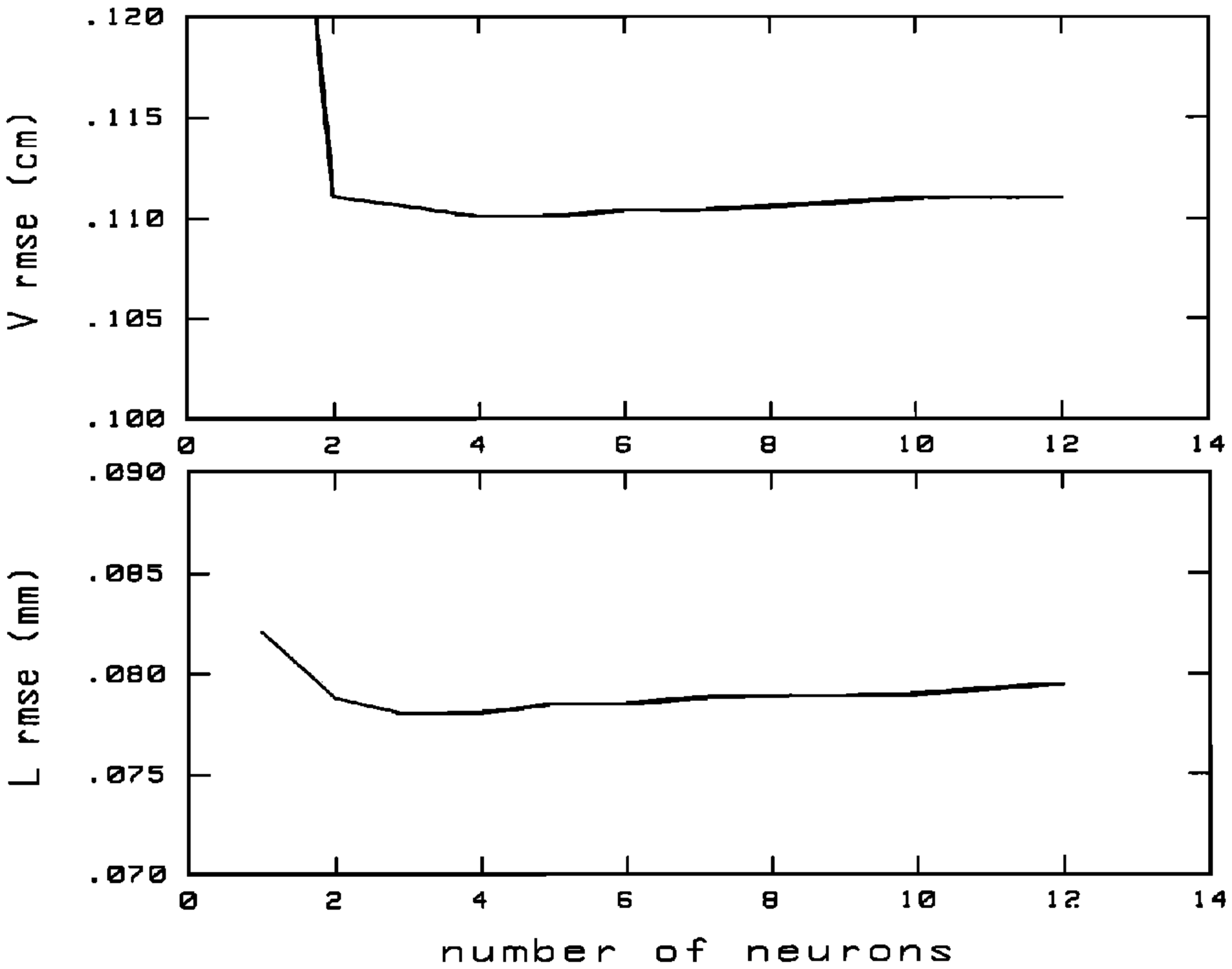

Figure 3. The rms error in the retrieval of (top) vapor and (bottom) liquid with a neural network versus the number of neurons in the hidden layer. Input is seven brightness temperatures. Radiometric noise is of 1.0 K standard deviation.

ative importance of its weights, and the least important ones are deleted. Typically, this is followed by some further training of the pruned network, and the procedure of pruning and training may be repeated for several cycles. Clearly, there are various choices to be made concerning how much training is applied at each stage, what fraction of the weights is pruned, and so on. In our case, every time a weight was removed, we trained the new net as long as was required by the early stopping procedure, and we continued with the pruning procedure until we realized that new removals involved a significant increase of the error in the retrieval operation. The most important consideration, however, is how to decide which weights should be removed. To do this, we need some measure of the relative importance, or saliency, of different weights. Results that will be shown have been obtained applying the simple concept that small weights are less important than large weights and using the magnitude of a weight value as a measure of its importance. We also considered more sophisticated approaches for determining the saliency of a weight as the optimal brain surgeon [*Hassibi and Stork,* 1993], which did not lead to different results. In Figure 4 we show an example of what happens during the pruning procedure. Until a certain point, the removal of the less significant connections is recommended, as it yields a decrease in the retrieval error. Then, after an oscillatory phase, successive removals make the error increase dramatically. The pruning procedure has to be interrupted before this point. In Figure 5 we show how the pruning procedure modified the net configuration for the case where the input vector consisted of the measurements provided by the multifrequency radiometer and the ceilometer and the output was water vapor. Almost 40% of connections were eliminated. Note that the pruning procedure did not sensitively alter the number of connections of units 1–3, but mainly affected links connected

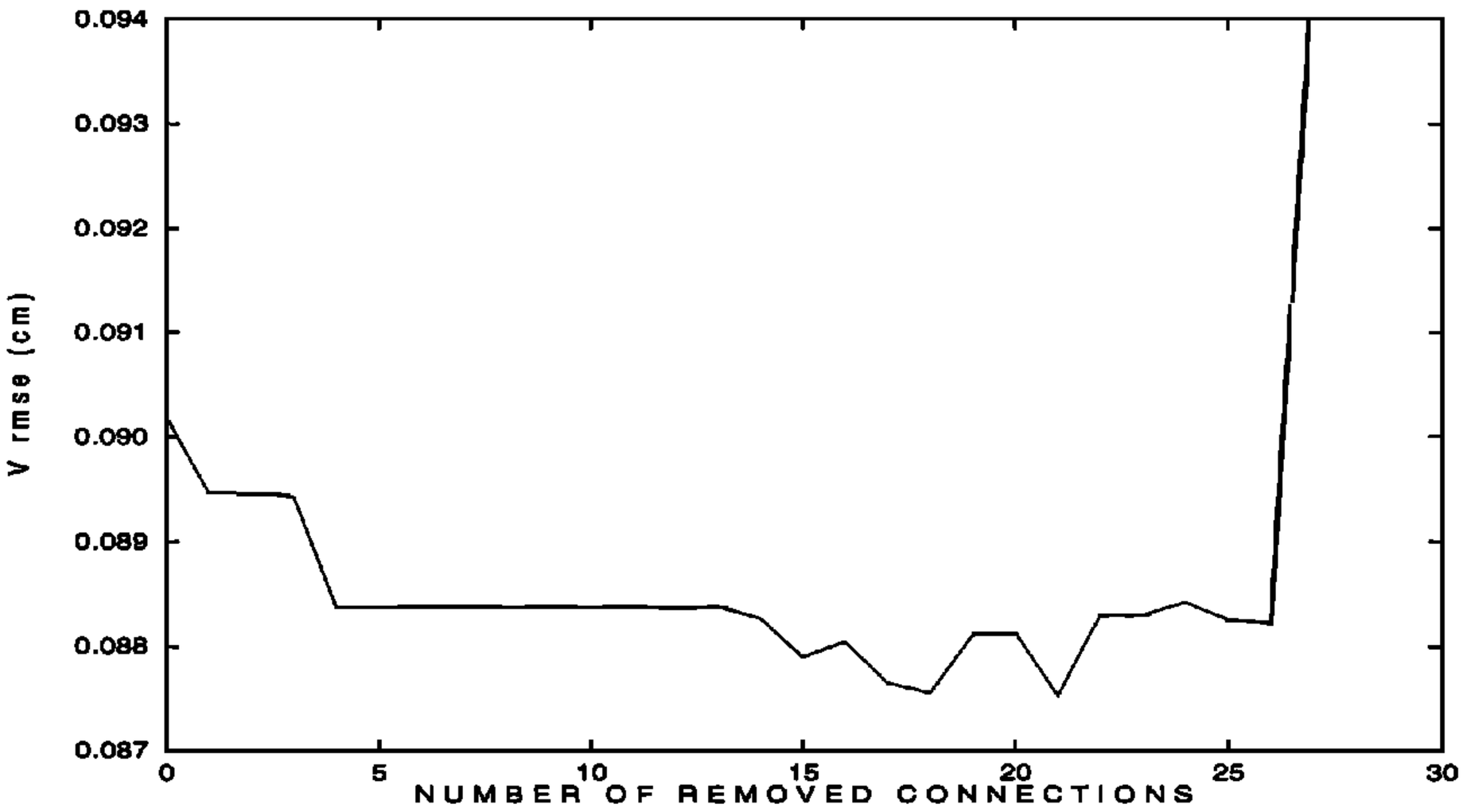

Figure 4. The rms error in the retrieval of vapor with a neural network versus the number of removed connections during the pruning procedure. Input is seven brightness temperatures and height of the cloud base. Radiometric noise is of 1.0 K standard deviation.

to inputs 5–8. In fact, even if the measurements of temperature profile sensitive channels and of the ceilometer can improve the estimation of vapor, most of the information needed for the retrieval is provided by channels 1–3, which include frequencies commonly chosen for dual-channel radiometers. Moreover, if pruning is continued, input units 5–8 would be the first ones to be canceled out, that is, the first ones to lose all their connections. In Table 5 the effects of the pruning procedure on the number of connections and on the rms error are reported for four cases.

Let us now focus on the overall number of coefficients that is needed to perform the inversion with the neural algorithm. In the case of Figure 5, it will be given by the

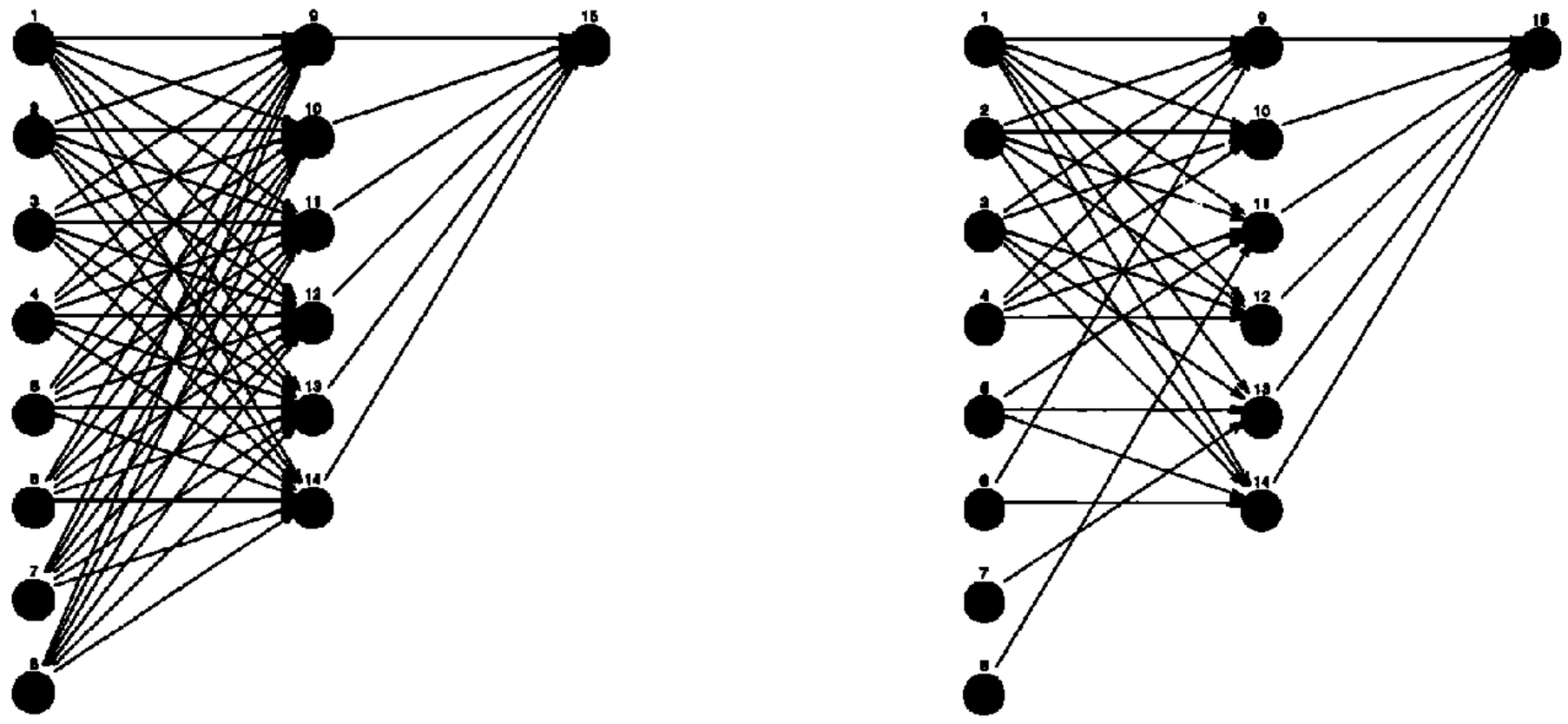

Figure 5. Neural network for the retrieval of vapor (left) before and (right) after the pruning procedure. Input is seven brightness temperatures and height of the cloud base. Radiometric noise is of 1.0 K standard deviation.

Table 5. Effects of the Pruning Procedure on Four Different Nets

Input	Output	ΔC, %	ΔRMS Error, %
MFR	V	−19	−1
MFR + C	V	−39	−3
MFR	L	−19	0
MFR + C	L	−4	0

ΔC and Δrms error are the variations of the number of connections and of the rms error when applying the procedure. Radiometric noise is of 1.0 K standard deviation.

number of the surviving connections plus the number of coefficients that define the bias of all the units. So it will be $33 + 15 = 48$. The number of coefficients that is used instead by LR is given by the coefficients necessary for the estimation of the T_{mr} plus the number of coefficients required for the estimation of vapor from the opacities, as described in section 2. Considering (11) and (12) and also adding in (12) the term for the ceilometer, we have $56 + 9 = 65$ coefficients. It follows that the neural algorithm, besides providing a more accurate estimation, needs a reduced number of adaptive coefficients to perform the retrieval. This remarkable property still holds with the other pruned networks.

6. Fault Tolerance Analysis

Finally, we analyzed the algorithm from the point of view of its fault tolerance. To this aim, we simulated a set of masked fault conditions which would not be detected by a standard automatic quality control system. In fact, some thresholds can be set to define acceptable physical ranges for the brightness temperatures measured by each channel of the radiometer. The detection of values outside these ranges can activate a recovery procedure which, for instance, makes use of an alternative algorithm with fewer channels, perhaps off-line, or discards corrupted data. However, if the fault is such that the output data are within the set ranges, it could not be detected. One at a time the brightness temperatures of all seven channels were permanently set to a value within an acceptable physical range, but very close to either its upper or lower bound, thus simulating a saturation problem.

The rms errors for the fault simulations are reported in Table 6. Three different inversion models are considered: LR, pruned neural network (PNN), and NN. The first seven rows of the table refer to the rms errors obtained for the saturation of each channel. The overall averaged errors for all the considered fault cases are reported in the bottom row. From this line we see that the degree of failure is smaller with NN than with LR, which means that the NN algorithm is more robust. We also note that the pruning procedure does not significantly alter the property of fault tolerance of the net; that is, the effectiveness of the removed connections can also be neglected from the point of view of information redundancy. A more detailed analysis shows that for liquid, and with the exception of the failure of the 51.25-GHz channel, the estimation is always better than that achievable using the a priori information, represented by the standard deviation value (reported in Table 1). The large error for the failure

Table 6. RMS Error in the Retrieval of V and L in Case of Fault Simulation in One Radiometric Channel and Averaged Over All Fault Cases

Saturated Channel Frequency, GHz	V RMS Error, cm			L RMS Error, cm		
	LR	PNN	NN	LR	PNN	NN
22.23	3.010	2.645	2.299	0.0192	0.0225	0.0204
23.87	3.773	2.766	2.739	0.0781	0.0242	0.0224
31.65	5.165	2.427	2.429	0.0687	0.0489	0.0378
51.25	0.381	0.246	0.244	0.2756	0.1489	0.1526
52.85	0.218	0.725	0.716	0.0604	0.0581	0.0626
53.85	0.479	0.521	0.512	0.0199	0.0121	0.0104
54.85	0.250	0.247	0.272	0.0233	0.0226	0.0232
Average	1.896	1.368	1.316	0.0779	0.0482	0.0470

Abbreviations are LR, linear regression; PNN, pruned neural network; and NN, neural network.

of the 51.25-GHz channel is due to the particularly wide dynamic range that the brightness temperature at this frequency spans when liquid varies in the examined interval. For vapor, if the fault occurs in one of the last four channels, the performance degrades slightly, with the rms error again smaller than the standard deviation value, but becomes unacceptable for the failure of the water vapor absorption line channels or of the atmospheric window one. In fact, as already noted for the pruning procedure in the previous section, these latter channels are necessary for a satisfactory retrieval of vapor.

7. Conclusions

The purpose of this study was to investigate the potentiality of NN in estimating the tropospheric integrated water vapor and integrated liquid water from radiometric data. Both dual-channel and multifrequency radiometers have been considered, with the optional addition of a laser ceilometer. Radiometric measurements have been simulated by applying the MPM propagation model to different sets of atmospheric profiles. Simulations have been mainly performed by using statistically generated profiles. Radiosonde-measured profiles have also been used to validate results. The NN retrieval algorithm has been trained and tested on two independent sets, and its performance has been compared with that of LR trained and tested on the same two sets. The obtained results highlight that NN delineates an inversion model capable of performing the retrieval with good accuracy, with generally better performance than LR algorithms, particularly when the presence of nonlinearities, due to high cloud liquid content, is stronger. Beyond that, NN algorithms seem to be more portable and flexible than LR algorithms, since they are capable of taking advantage of the information provided by a ceilometer.

We also collected data for the design of an optimum NN for the proposed inversion problems. It has been found that an optimum range for the choice of the number of the neurons in the hidden layer of the considered network exists. The net must have at least a minimum size to satisfactorily implement input-output associative capabilities, but too many neurons tend to reduce its generalization properties.

Moreover, we examined a pruning procedure for the removal of nonsignificant connections. At the end of this procedure the error is about the same, or even smaller, than the one displayed by the net with all connections. Inspection of the pruned networks shows that the procedure mainly removes connections to inputs with less information content, which, in this case, are temperature profile sensitive channels and the ceilometer. The overall number of the adaptive parameters of the algorithm can even be smaller than the number required by a LR technique.

Finally, we simulated the existence of undetected faults in the radiometric system, saturating, one at a time, the measurements of its channels at the extremes of their physical range of variability. The retrieval performance suffers a stronger degradation when faults are present in the water vapor absorption line channels or in the atmospheric window one. However, the NN is capable of reducing the degradation better than LR, therefore demonstrating itself to be a more robust inversion algorithm.

Acknowledgments. This work has been partially supported by Agenzia Spaziale Italiana (ASI). Radiosonde data have been provided by the Meteorological Service of the Italian Air Force. The authors acknowledge the anonymous reviewers for their helpful criticism and suggestions.

References

Askne, J., G. Skoog, and E. Wimberg, Test of a ground-based microwave radiometer for atmospheric temperature profiling with meteorological applications, *Int. J. Remote Sens.*, *6*, 1241–1256, 1985.

Battistelli, E., C. Capitani, A. Culebras, F. Del Frate, G. Schiavon, B. Arbesser-Rastburg, and J.P. Baptista, A multifrequency mm-wave radiometer (MFR): Instrument description and retrieval algorithm, in *Microwave Radiometry and Remote Sensing of the Environment*, edited by D. Solimini, pp. 529–536, VSP Press, Utrecht, Netherlands, 1995.

Beale, R., and T. Jackson, *Neural Computing: An Introduction*, Adam Hilger, Bristol, England, 1990.

Bishop, C.M., *Neural Networks for Pattern Recognition*, Oxford Univ. Press, New York, 1995.

Cabrera, C.R., and D.H. Staelin, Passive microwave relative humidity retrievals using feedforward neural networks, *IEEE Trans. Geosci. Remote Sens.*, *33*, 1324–1328, 1995.

Churnside, J.H., T.A. Stermitz, and J.A. Schroeder, Temperature profiling with neural network inversion of microwave radiometer data, *J. Atmos. Oceanic Technol.*, *11*, 105–109, 1994.

Damosso, E., L. Stola, and G. Brussaard, Characterization of the 50–70 GHz band for space communications, *ESA J.*, *7*, 25–43, 1983.

Dawson, M.S., Applications of electromagnetic scattering models to parameter retrieval and classification, in *Microwave Scattering and Emission Models and*

Their Applications, edited by A.K. Fung, pp. 527–557, Artech House, Norwood, Mass., 1994.

Decker, M.T., E.R. Westwater, and F.O. Guiraud, Experimental evaluation of ground-based microwave radiometric sensing of atmospheric temperature and water vapor profiles, *J. Appl. Meteorol.*, *17*, 1788–1795, 1978.

Del Frate, F., and G. Schiavon, A combined natural orthogonal functions/neural network technique for the radiometric estimation of atmospheric profiles, *Radio Sci.*, *33*, 405–410, 1998.

Han, Y., and E.R. Westwater, Remote sensing of tropospheric water vapor and cloud liquid water by integrated ground-based sensors, *J. Atmos. Oceanic Technol.*, *12*, 1050–1059, 1995.

Hassibi, B., and D.G. Stork, Second order derivatives for network pruning: Optimal brain surgeon, in *Advances in Neural Information Processing Systems*, vol. 5, edited by S.J. Hanson, J.D. Cowan, and C. L. Giles, pp. 164–171, Morgan Kaufmann, San Francisco, Calif., 1993.

Hogg, D.C., F.O. Guiraud, J.B. Snider, M.T. Decker, and E.R. Westwater, A steerable dual-channel microwave radiometer for measurements of water vapor and liquid in the troposphere, *J. Clim. Appl. Meteorol.*, *22*, 789–806, 1983.

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

Hsu, S.-Y., T. Masters, M. Olson, M. Tenorio, and T. Grogan, Comparative analysis of five neural network models, *Remote Sens. Rev.*, *6*, 319–329, 1992.

Janssen, M.A., An introduction to the passive microwave remote sensing of atmospheres, in *Atmospheric Remote Sensing by Microwave Radiometry*, edited by M. Janssen, pp. 1–35, John Wiley, New York, 1993.

Li, L., J. Vivekanandan, C.H. Han, and L. Tsang, Microwave radiometric technique to retrieve vapor, liquid and ice, 1, Development of a neural network-based inversion method, *IEEE Trans. Geosci. Remote Sens.*, *35*, 224–236, 1997.

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, in *AGARD 52nd Specialists' Meeting of the Electromagnetic Wave Propagation Panel*, pp. 3.1–3.10, Advis. Group for Aerosp. Res. and Dev., Brussels, 1993.

Lippmann, R.P., An introduction to computing with neural nets, *IEEE ASSP Mag.*, *4*, 4–22, 1987.

Peter, R., and N. Kämpfer, Radiometric determination of water vapor and liquid water and its validation with other techniques, *J. Geophys. Res.*, *97*, 18,173–18,183, 1992.

Rumelhart, D.E., G.E. Hinton, and R.J. Williams, Learning internal representations by error propagation, in *Parallel Distributed Processing*, edited by D.E. Rumelhart and J.L. McClelland, pp. 318–362, MIT Press, Cambridge, Mass., 1986.

Schiavon G., D. Solimini, and E.R. Westwater, Performance analysis of a multifrequency radiometer for predicting atmospheric propagation parameters, *Radio Sci.*, *28*, 63–76, 1993.

Soong, T.T., *Probabilistic Modeling and Analysis in Science and Engineering*, pp. 206–209, John Wiley, New York, 1982.

Westwater, E.R., The accuracy of water vapor and cloud liquid determination by dual-frequency ground-based microwave radiometry, *Radio Sci.*, *13*, 677–685, 1978.

Westwater, E.R., Ground based remote sensing of meteorological variables, in *Atmospheric Remote Sensing by Microwave Radiometry*, edited by M. Janssen, pp. 145–213, John Wiley, New York, 1993.

Westwater, E.R., and F.O. Guiraud, Ground-based microwave radiometric retrieval of precipitable water vapor in the presence of clouds with high liquid content, *Radio Sci.*, *15*, 947–957, 1980.

Westwater, E.R., and O.N. Strand, Statistical information content of radiation measurements used in indirect sensing, *J. Atmos. Sci.*, *25*, 750–758, 1968.

Zell, A., et al., *SNNS Stuttgart Neural Network Simulator, User Manual, Version 4.1, Rep. 6/95*, Inst. for Parallel and Distrib. High Performance Syst., Univ. of Stuttgart, Stuttgart, Germany, 1995.

F. Del Frate and G. Schiavon, Università Tor Vergata, Ingegneria, D.I.S.P., Via di Tor Vergata, 110, I-00133, Roma, Italy. (e-mail: delfrate@disp.uniroma2.it; schiavon@disp.uniroma2.it)

(Received October 24, 1997; revised March 11, 1998; accepted June 23, 1998.)