

## Determination of Cloud Liquid Water Path over the Oceans from Special Sensor Microwave/Imager (SSM/I) Data Using Neural Networks

THOMAS JUNG AND EBERHARD RUPRECHT

*Institut für Meereskunde, Kiel, Germany*

FRIEDRICH WAGNER

*Institut für Theoretische Physik, Kiel, Germany*

(Manuscript received 23 June 1997, in final form 10 December 1997)

### ABSTRACT

A neural network (NN) has been developed in order to retrieve the cloud liquid water path (LWP) over the oceans from Special Sensor Microwave/Imager (SSM/I) data. The retrieval with NNs depends crucially on the SSM/I channels used as input and the number of hidden neurons—that is, the NN architecture. Three different combinations of the seven SSM/I channels have been tested. For all three methods an NN with five hidden neurons yields the best results. The NN-based LWP algorithms for SSM/I observations are intercompared with a standard regression algorithm. The calibration and validation of the retrieval algorithms are based on 2060 radiosonde observations over the global ocean. For each radiosonde profile the LWP is parameterized and the brightness temperatures (Tb's) are simulated using a radiative transfer model.

The best LWP algorithm (all SSM/I channels except T85V) shows a theoretical error of  $0.009 \text{ kg m}^{-2}$  for LWPs up to  $2.8 \text{ kg m}^{-2}$  and theoretical “clear-sky noise” ( $0.002 \text{ kg m}^{-2}$ ), which has been reduced relative to the regression algorithm ( $0.031 \text{ kg m}^{-2}$ ). Additionally, this new algorithm avoids the estimate of negative LWPs.

An indirect validation and intercomparison is presented that is based upon SSM/I measurements (*F-10*) under clear-sky conditions, classified with independent IR-Meteosat data. The NN-based algorithms outperform the regression algorithm. The best LWP algorithm shows a clear-sky standard deviation of  $0.006 \text{ kg m}^{-2}$ , a bias of  $0.001 \text{ kg m}^{-2}$ , nonnegative LWPs, and no correlation with total precipitable water. The estimated accuracy for SSM/I observations and two of the proposed new LWP algorithms is  $0.023 \text{ kg m}^{-2}$  for  $\text{LWP} \leq 0.5 \text{ kg m}^{-2}$ .

### 1. Introduction

Understanding the hydrological cycle is one of the major tasks in atmospheric research. This cycle is responsible for the redistribution of water and (latent) energy in the climate system. In addition clouds—one component of this cycle—significantly affect the earth radiation budget (Harrison et al. 1990). Small variations in cloud amount can alter the radiative forcing at the top of the atmosphere and consequently change the atmospheric temperature. Modeling results further suggest that a radiative perturbation due to a  $\text{CO}_2$  doubling could be balanced by an increase in liquid water path (LWP) from 0.05 to  $0.06\text{--}0.068 \text{ kg m}^{-2}$  (Slingo 1990). Thus, a prerequisite for a better understanding of climate is the knowledge of the temporal and spatial behavior of the cloud fields.

The aim of this study is to contribute to the obser-

vational part, that is, to apply neural networks (NNs) to the retrieval of LWP from spaceborne microwave measurements over the oceans where ground-based measurements are limited.

The microwave radiometer SSM/I (Special Sensor Microwave/Imager) on board the satellites of the Defense Meteorological Satellite Program (DMSP) measures thermal-emitted radiation from the earth's surface and the atmosphere at 19.35, 37.0, and 85.5 GHz, vertically and horizontally polarized, and at 22.235 GHz, vertically polarized only. These observational data can be used to retrieve surface and atmospheric parameters simultaneously, in particular those that are included in the hydrological cycle. (For further details, see Hollinger et al. 1987.) The first DMSP satellite was launched in June 1987. The long time series provided by the SSM/I makes this radiometer attractive for many research purposes: for example, investigation of seasonal and interannual variability of the hydrological parameters (Ferraro et al. 1996).

Several retrieval algorithms for the LWP can be found in the literature. On the one hand, there exist (semi-) statistical algorithms based on regression techniques (Alishouse et al. 1990b; Karstens et al. 1994; Weng and

---

*Corresponding author address:* Thomas Jung, Institut für Meereskunde an der Universität Kiel, Düsterbrookweg 20, D-24105 Kiel, Germany.  
E-mail: tjung@ifm.uni-kiel.de

Grody 1994). On the other hand, there are several algorithms of physical origin; that is, they are based on radiative transfer calculations (Greenwald et al. 1993; Liu and Curry 1993; Prigent et al. 1994). The former have advantages due to their simplicity in operational use, and the latter are assumed to perform better, as validation results for six cases with simultaneous aircraft and SSM/I measurements indicate (Cober et al. 1996). A more detailed description of most of the cited algorithms is given by Cober et al. (1996).

As noted by Petty at the SSM/I Algorithm Symposium (see Colton and Poe 1994), retrieval problems arise due to correlations between the retrieved LWP and high humidity values (“cross talk”) and due to “clear-sky noise,” which is accompanied by negative LWPs. However, setting negative LWPs to zero, as is sometimes proposed, biases a retrieval mean proportional to the level of “noise.” Thus, it is desirable to develop LWP algorithms that combine simplicity in applications with accuracy and that show neither cross talk nor clear-sky noise.

During the past five years the application of NNs has become a valuable tool in meteorology. Among these applications are predictions, for example, on time series data (Elsner and Tsonis 1992; Tang et al. 1994; Hastenrath et al. 1995) and nonlinear principal component analysis of climate data (Sengupta and Boyle 1995). However, most applications can be found in remote sensing, either to classify properties within an image (for an overview, see Atkinson and Tatnall 1997) or to solve inverse problems: for example, to retrieve vertical profiles of temperature (Churnside et al. 1994; Butler and Meredith 1996) and of relative humidity (Cabrera-Mercader and Staelin 1995) from microwave data. With NNs, improvements were found over regression algorithms (Cabrera-Mercader and Staelin 1995; Butler and Meredith 1996). Retrieval algorithms for SSM/I observations on the basis of NNs are also available to infer wind speed (Stogryn et al. 1994; Krasnopolsky et al. 1995) and snow parameters (Tsang et al. 1992). Krasnopolsky et al. (1995) report significant improvements compared to regression algorithms, in particular under cloudy and very cloudy conditions. To our knowledge, however, NNs have not been applied to the retrieval of LWP from SSM/I measurements.

Neural networks have several properties that make them attractive for solving inverse problems. First, they are able to detect and approximate multiple, nonlinear correlations without a priori information about the problem under consideration. Second, they are able to approximate relationships using only the *data* presented. That is, they do not depend on *model equations* like regression techniques. Once training has been finished, they are easy to implement, and the computer cost is relatively low.

In this study, the potential of NNs for the development of LWP retrieval algorithms for SSM/I measurements is investigated. In section 2 the data basis used for the

TABLE 1. Summary of radiosonde observations used to parameterize the liquid water paths and simulate the SSM/I brightness temperatures.

Source	<i>n</i>	Location
FGGE <sup>a</sup>	1311	North Atlantic Ocean
ICE <sup>b</sup>	132	North Sea
SONNE <sup>c</sup>	141	Subtropical Pacific Ocean
MEPO <sup>d</sup>	476	Atlantic Ocean

<sup>a</sup> FGGE: First Global GARP (Global Atmospheric Research Programme) Experiment (April–October 1979).

<sup>b</sup> ICE: International Cirrus Experiment (September–October 1989).

<sup>c</sup> MEPO: R/Vs *Meteor* and *Polarstern* (different cruises between 1983 and 1986).

<sup>d</sup> SONNE: R/V *Sonne* (cruise between February and March 1995).

calibration and validation is described. The NN methodology is introduced in section 3. The algorithm development and the results are described in sections 4 and 5, and the summary and conclusions are given in section 6.

## 2. Data and radiative transfer model

Three types of data are used in this study. The calibration and validation of the LWP algorithms are carried out with pairs of simulated brightness temperatures (Tb’s) for the SSM/I and parameterized LWPs, both calculated from radiosonde observations (RAOBs). An indirect validation is performed with SSM/I observations under clear-sky conditions from the *F-10* instrument. The clear-sky cases were defined with collocated IR-Meteosat data.

About 2060 RAOBs over most of the global oceans with corresponding synoptic observations (SYNOPs) are the main data basis in this study. The sources, number of ascents, locations, and time periods for the different subsets are given in Table 1. RAOBs from two regions are missing: no data are available over the Indian Ocean, and except for one cruise, the subtropical South Pacific is not represented at all in our dataset.

For every RAOB, liquid water content is parameterized by the method of Karstens et al. (1994). The parameterization is carried out in three steps. First, if the relative humidity exceeds 95% in a given layer, it is assumed that a cloud is present. Second, for each layer, the adiabatic liquid water content at height *h* ( $LWC_{adi}$ ) is calculated by

$$LWC_{adi}(h) = \int_{z_0}^h \rho(z) \frac{C_p}{L} (\Gamma_D - \Gamma_S) dz, \quad (1)$$

where  $z_0$  is the height of the cloud base,  $\rho(z)$  denotes air density,  $L$  is the latent heat of vaporization, and  $\Gamma_D$  and  $\Gamma_S$  are the dry and moist adiabatic lapse rates, respectively. Third, to account for entrainment and adiabatic processes (e.g., condensation and precipitation),  $LWC_{adi}$  is modified empirically according to

$$LWC_{mod} = LWC_{adi}(1.239 - 0.145 \ln \Delta h), \quad (2)$$

where  $\Delta h$  is the height above cloud base in meters. Note that (2) reduces the  $LWC_{\text{adi}}$ , which is consistent with aircraft measurements (Warner 1955). Finally, the LWP is calculated as a vertical integral of  $LWC_{\text{mod}}$  of all cloudy layers:

$$LWP = \int_{z_0}^H LWC_{\text{mod}}(h) dz, \quad (3)$$

where  $H$  denotes the top of the uppermost cloud. For temperatures below  $-20^\circ\text{C}$  in the radiosonde profile, it is assumed that clouds exist totally of ice particles. Using a relative humidity threshold method (first step) it is possible to obtain multiple cloud layers, in contrast, for example, to the one-layer artificial clouds of Greenwald et al. (1993).

The mean and standard deviation of the estimated LWP for all 2060 RAOBs is given by 0.13 and 0.35  $\text{kg m}^{-2}$ , respectively. For 1408 RAOBs (approximately 70%) LWP is equal to 0.0, for 476 cases LWP is between 0.0 and 0.5  $\text{kg m}^{-2}$ , and for 176 cases LWP is greater than 0.5  $\text{kg m}^{-2}$ .

Both the RAOBs and the calculated profiles of the  $LWC_{\text{mod}}$  serve as input for the radiative transfer model to compute the  $Tb$ 's. Missing wind speed data in the SYNOPs were filled using equally distributed random values between 0 and 14  $\text{m s}^{-1}$ . Missing SSTs (sea surface temperatures) were substituted by the surface air temperature. For calculating the microwave radiances at the top of the atmosphere the radiative transfer model developed by Simmer (1994) was used. This one-dimensional monochromatic model solves the polarized radiative transfer equation by successive order of scattering (e.g., Goody and Yung 1989). The absorption and scattering coefficients were calculated using Mie theory. Rain clouds are not included. For the ocean surface, specular reflection is assumed modified by wind-speed-dependent surface roughness by a formulation of Wisler and Hollinger (1977). The influence of foam (reflection) on the oceanic surface emissivity was allowed for according to Stogryn (1972). The simulated  $Tb$ 's refer to an incident angle of  $53^\circ$ . Normally distributed noise has been added to the  $Tb$ 's with standard deviations equal to the noise-equivalent temperature differentials given by Hollinger et al. (1987).

The clear-sky dataset of microwave observation has been compiled with independent IR data (10.5–12.5  $\mu\text{m}$ ) obtained from Meteosat. This dataset comprises 70 000 SSM/I observations (*F-10* instrument) taken over the Atlantic Ocean for June 1992.

### 3. Neural networks

The task of an NN is to determine for each set of  $\mathbf{Tb} = (Tb_1, \dots, Tb_K)$  a value for the geophysical parameters  $\mathbf{P} = (P_1, \dots, P_M)$ . Even though we will use only one parameter ( $P = LWP$ ) in this study,  $\mathbf{P}$  is retained as a vector for the theoretical discussion. A standard

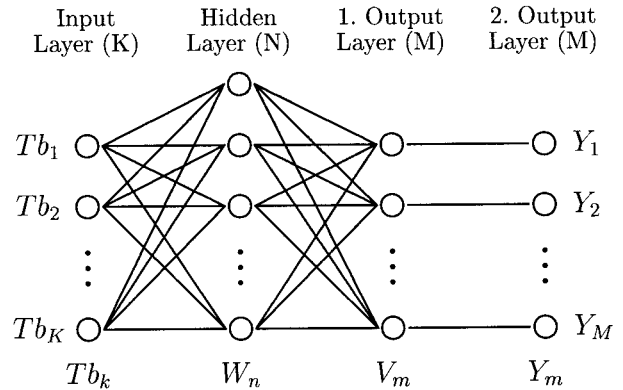


FIG. 1. Schematic of a layered feed-forward NN architecture. The components of the input vector  $\mathbf{Tb}$  are the brightness temperatures, and those of the output vector  $\mathbf{Y}$  are the geophysical parameters. The number of neurons in each layer is given in parentheses.

feed-forward NN with one hidden layer augmented by two additional output layers will be applied. The NN architecture is depicted in Fig. 1. Note that only the hidden layer is *fully* connected to the input and the first output layer. The calculation of the NN output  $\mathbf{Y}$  proceeds as follows. From the input value  $\mathbf{Tb}$  the activations  $W_n$  for the  $N$  neurons of the hidden layer are obtained by

$$W_n = \tanh\left(\omega_{n0} + \sum_{k=1}^K \omega_{nk} Tb_k\right) \quad n = 1, \dots, N, \quad (4)$$

where  $\omega_{nk}$  ( $\omega_{n0}$ ) denote the weights (offsets) of the  $n$ th hidden neuron. The activations  $W_n$  serve as input for the first output layer with  $M$  neurons. Its activations  $V_m$  are obtained in a similar way as in (4): that is,

$$V_m = \tanh\left(\Omega_{m0} + \sum_{n=1}^N \Omega_{mn} W_n\right) \quad m = 1, \dots, M. \quad (5)$$

Again the components of  $\mathbf{\Omega}$  are free parameters to be determined. The values of  $V_m$  are restricted to the interval  $-1 < V_m < 1$ , which, in general, does not coincide with the range of  $\mathbf{P}$ . In principle, one could use the minimum and maximum values of  $\mathbf{P}$  to map  $\mathbf{P}$  in the interval  $(-1, 1)$  by a linear transformation (e.g., Churnside et al. 1994). However, the extreme values of  $\mathbf{P}$  may depend on a few nonrepresentative cases. To overcome this problem we included a second-output layer into the NN, the results of which are calculated by

$$Y_m = a_{0m} + a_{1m} V_m \quad m = 1, \dots, M. \quad (6)$$

According to (4)–(6) the NN output  $\mathbf{Y}$  depends on  $\mathbf{Tb}$  and on the  $(K + 1 + M)N + 3M$  free parameters  $\mathbf{X} = (\boldsymbol{\omega}, \mathbf{\Omega}, \mathbf{a})$ . They have to be determined by a learning algorithm.

Using a training set  $(\mathbf{Tb}^\mu, \mathbf{P}^\mu)$  of  $\mu = 1, \dots, N_E$  data pairs, where the answer  $\mathbf{Y}^\mu = \mathbf{P}^\mu$  is known, the best set of  $\mathbf{X}$  minimizes a so-called cost function  $E$ ,

which is obtained from both the NN output  $Y_m^\mu(\mathbf{Tb}^\mu, \mathbf{X})$  and the nominal values  $P_m^\mu$  for each  $\mu$ th event by

$$E_\mu = \sum_{m=1}^M [Y_m^\mu(\mathbf{Tb}^\mu, \mathbf{X}) - P_m^\mu]^2$$

and

$$E = \sum_{\mu=1}^{N_E} E_\mu. \quad (7)$$

Other cost functions are possible. We prefer the quadratic form since its value at the minimum is essentially the rms difference between the data  $\mathbf{P}^\mu$  and the calculated NN output  $\mathbf{Y}^\mu$ . That is,

$$\text{rms} = (EN_E^{-1})^{1/2}.$$

Therefore,  $E$  is a quantitative measure for the quality of the NN.

To find the minimum of  $E$  with respect to  $\mathbf{X}$ , the so-called back-propagation algorithm (Rumelhart et al. 1986) is frequently used (e.g., Churnside et al. 1994; Krasnopolsky et al. 1995) since the formulas (4)–(6) allow an easy calculation of the gradient  $\nabla E_\mu$ . For each event the weight vector  $\mathbf{X}$  is updated according to

$$\Delta \mathbf{X} = -\gamma \cdot \nabla E_\mu. \quad (8)$$

The update (8) for all  $N_E$  events is repeated until the changes in  $\Delta \mathbf{X}$  become small or  $E$  is less than some prescribed value corresponding to an acceptable rms error. (See below for how this method is applied in this study.) Various improvements such as the adjustable learning rate  $\gamma$  or the inclusion of a momentum term (see, e.g., Rojas 1995) do not remove the disadvantage that back propagation is a simple gradient method and therefore computationally slow. The advantage of (8) is of course that the update is done for each event instead of all events, as in conventional optimization methods. The noise introduced by this method should prevent the algorithm from getting stuck in one of the many local minima of  $E$ . In our case, however, the training set contains  $\mathcal{O}(1000)$  events and the dimension of  $\mathbf{X}$  is  $\mathcal{O}(50)$ , which means computing time is of vital importance. Since  $\nabla E$  can be calculated, one should choose derivative methods. In particular, the so-called variable metric methods (Fletcher and Powell 1963) that we adopt are considered as very efficient (e.g., Press et al. 1992; van der Smagt 1994). With such a method, the cost function  $f(t) = E(\mathbf{X} + t \cdot \mathbf{s})$  is minimized at each step along the search direction  $\mathbf{s}$  given by

$$\mathbf{s} = -\mathbf{H}^{-1} \cdot \nabla E(\mathbf{X}). \quad (9)$$

Here  $\mathbf{H}$  denotes the Hessian matrix  $(\nabla \nabla^T)E$ . With known  $\mathbf{H}$ , cost functions quadratic in  $\mathbf{X}$  are minimized in one step. A general problem in applying (9) is that the calculation of the Hessian and its inversion take large amounts of computer time. This problem is avoided using the Fletcher–Powell algorithm (Fletcher and Powell 1963). Instead of the exact  $\mathbf{H}^{-1}$  an approximation  $\tilde{\mathbf{H}}^{-1}$

is used. In each step,  $\tilde{\mathbf{H}}_i^{-1}$  (the lower index refers to the iteration step) is corrected in such a way that it will converge after a sufficiently large number of steps to the true  $\mathbf{H}^{-1}$ . The success of this method depends somewhat on the choice of the initial matrix  $\mathbf{H}_0$ . It can be chosen as unit matrix; in this case,  $\mathbf{s}$  in (9) is just the direction of the steepest descent. In general, however, the large-scale differences of the parameters  $\mathbf{X}$  and the correlations between these parameters, which are described by the nondiagonal elements of  $\mathbf{H}^{-1}$ , have to be taken into account. For this purpose, we preprocessed the data. Instead of  $\mathbf{Tb}_k^\mu$ , we used decorrelated data  $d_k^\mu$  given by

$$d_k^\mu = \sum_{k'=1}^K C_{kk'}^{-1/2} (\mathbf{Tb}_k^\mu - \langle \mathbf{Tb}_k \rangle), \quad (10)$$

where  $\langle \cdot \rangle$  denotes the expectation operator and  $\mathbf{C}$  the covariance matrix taken over all  $N_E$  training events. Due to this linear transformation,  $\langle d_k^\mu \rangle = 0$  and  $\langle d_k^\mu d_{k'}^\mu \rangle = \delta_{kk'}$ ; that is, all input data have the same scale (e.g., Campbell 1996). The dataset  $\mathbf{d}^\mu$  is needed only during the training phase. Once the optimum  $\mathbf{X}$  is found, the NN output can be obtained by calculating the input-hidden weights  $\boldsymbol{\omega}$  referring to  $\mathbf{Tb}$  from  $\bar{\boldsymbol{\omega}}$  referring to  $\mathbf{d}$  by

$$\omega_{nk} = \sum_{k'=1}^K \bar{\omega}_{nk'} C_{k'k}^{-1/2}$$

and

$$\omega_{n0} = \bar{\omega}_{n0} - \sum_{k=1}^K \omega_{nk} \langle \mathbf{Tb}_k \rangle. \quad (11)$$

The actual implementation is highly nontrivial. Protections against instabilities and ill-conditioned  $\mathbf{H}^{-1}$  have to be provided. We use the minimizing part of the CERN–Munich phase shift analysis program, which has been used and tested extensively during the past 30 years (Wagner and Lovelace 1971). Since the minimization concerns the inner part, we could not use any customary NN program but had to design and write a completely new one (Wagner 1996). Using this new program, Jung (1996) showed that computer cost is reduced by two orders of magnitude in comparison to classic back-propagation methods.

Two important differences in the use of NN programs, as compared to other optimization problems, need to be mentioned. Conventionally, minimization is said to have converged if the changes  $\Delta \mathbf{X}$  and/or  $\Delta E$  had become sufficiently small. This criterion cannot be applied to our problem for the following reason. Any method minimizing  $E$  will in the beginning fit the gross features of the data. This is reflected by a large gain in  $E$  in each step. Once a reasonable  $E$  is obtained, it will start to optimize particular properties of the training events (which may also be a few badly measured events). To prevent this “overfitting,” the usual method is to divide the available data into two parts: a training set that is

used for the minimization and a test set for the validation. In each iteration step the cost function for the latter is also calculated. The overfitting phase is signaled if the cost function for the training dataset decreases, but the cost function for the test set increases over several subsequent iterations. Independent of the minimization method, any further minimization is meaningless. Therefore, we restart the minimization using the same initial weights and stop the learning process before the overfitting phase starts.

The second remark concerns the problem of several minima. The minimization has to be repeated many times with different randomly chosen starting values for  $\mathbf{X}$ . This is of particular importance for the variable metric method. From the distribution of the achieved cost functions one can estimate whether the best value is purely accidental or representative of the dataset.

#### 4. Algorithm development—Different neural network architectures

To investigate the architecture of NNs for the development of an LWP algorithm, the 2060 available data pairs have been randomly divided into two parts. The first 1030 data pairs were used as a training set, and the remaining as a test set. We found for all of our results that interchanging the role of the training and the test sets did not produce any significant change in the results.

The following questions have been addressed to investigate the best possible architecture.

- 1) NN input: Which channels of the SSM/I should be used at best?
- 2) Number of hidden neurons: How many do we need?
- 3) Application: Are the channels always available?

The first and third criteria are common to other retrieval techniques as well. The number of hidden neurons is proportional to the flexibility of the NN: the more neurons that are used, the better the training data that are reproduced. Nevertheless, too many neurons may result in greater errors for the independent test set. The last criterion, applicability, is motivated by the fact that, for example, the SSM/I measurements at T85V on board the *F-8* satellite are not reliable after 1989 due to high instrumental noise (which hampers a long homogeneous time series) and that corrections for systematic deviations between measured and simulated Tb's (Simmer 1994) are not available for T85V.

Three different input combinations are investigated: all SSM/I channels except the T85V one (ALL/85V), the low-frequency SSM/I channels (19, 22, 37 GHz) for both polarizations (LOW), and a two-channel algorithm—T22V and T37V (TWO). To estimate the influence of different local minima, 40 random searches from different starting points in weight space were carried out for zero to seven hidden neurons and for each of the three input combinations. In each run a maximum of 2000 iterations was performed and the best solution

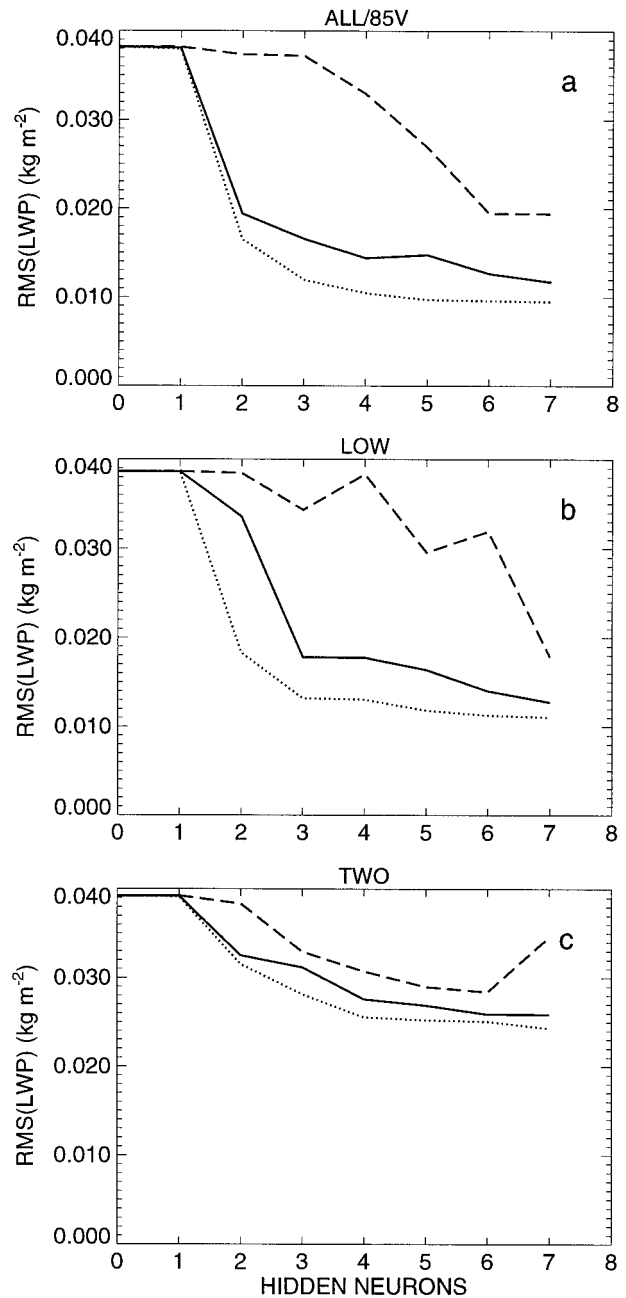


FIG. 2. Dependency of the rms error in the LWP retrieval on the number of hidden neurons for three different inputs: (a) all SSM/I channels except T85V (ALL/85V); (b) low-frequency SSM/I channels (LOW); and (c) T22V, T37V (TWO). Median (solid), 10th (dotted), and 90th (long dashed) percentiles. The results refer to 40 realizations and the independent test set.

$\mathbf{X}$  for the test set was saved for further investigation. Note that, typically, fewer than 2000 iterations were sufficient to reach convergence.

The rms errors for ALL/85V, LOW, and TWO depend on the number of hidden neurons (Fig. 2). Differences are obvious except between zero and one neuron. With more than five hidden neurons, the improvement shows

TABLE 2. The best results in terms of the rms error of all 40 random searches for the independent test sets and different NN architectures. The number of weights (complexity) are given in parentheses. Three different input combinations are considered: all SSM/I channels except T85V (ALL/85V), the low-frequency SSM/I channels (LOW), and T22V, T37V (TWO).

Hidden neurons	rms error ( $10^{-3}$ kg m $^{-2}$ )		
	ALL/85V	LOW	TWO
0	38.2 (9)	38.6 (8)	30.2 (5)
1	37.9 (11)	38.5 (10)	39.1 (7)
2	16.1 (19)	18.2 (17)	31.3 (11)
3	10.6 (27)	13.2 (24)	27.4 (15)
4	9.8 (35)	11.7 (31)	25.5 (19)
5	8.9 (43)	11.2 (38)	24.1 (23)
6	8.7 (51)	10.6 (45)	24.0 (27)
7	8.9 (59)	10.6 (52)	23.8 (31)

some kind of saturation, which is obvious from the 10th percentile (dotted in Fig. 2). Note that the 10th percentile separates the four best solutions with respect to the rms error. The spread of the curves in Fig. 2 indicates that the range of the rms errors is wide for ALL/85V and LOW and very narrow for TWO for more than one hidden neuron. This indicates that few of the runs were stuck in local minima of the cost function, which is true especially for the former two algorithms. This result confirms the importance of multiple random searches. Despite the narrow range of the rms errors in TWO, the accuracies are better for ALL/85V and LOW, which can be seen from the 10th percentile.

The best results of all 40 realizations are summarized in Table 2. In addition, the number of weights is given to show the complexity of the method. The differences between ALL/85V and LOW are relatively small. Note that the best results agree well with the 10th percentile in Fig. 2, which enhances confidence that the best values are not purely accidental.

The rms errors for TWO with more than four hidden neurons are considerably higher than those for ALL/85V and LOW, indicating that there is additional valuable information in the other channels for the LWP retrieval. It also shows the ability of NNs to make use of small correlations in the data. Note that this result is in contrast with that reported by Karstens et al. (1994) with their regression method. Karstens et al. obtained no further improvement when all SSM/I channels were used as predictors. This might be the consequence of collinearities between the Tb's, which are problematic for ordinary least squares techniques (Crone et al. 1996).

We conclude that five hidden neurons seem to be a good compromise with respect to accuracy and simplicity. We have also applied Akaike's Information Criterion (AIC), which is a more objective criterion for model selection in regression and NN applications (e.g., von Storch and Zwiers 1997). AIC takes into account the two parameters' model error and complexity (number of weights), for which a minimum is required. AIC as function of hidden neurons (not shown) leads essen-

tially to the same conclusions described above but in a more objective manner; that is, five hidden neurons seem to be sufficient. Consequently, further investigations in the remainder of this paper refer to NN architecture with five neurons in the hidden layer.

## 5. Results

### a. Algorithm intercomparison for simulated data

In the following, the new algorithms will be compared to the regression algorithm of Karstens et al. (1994) (denoted as REG hereafter). Karstens et al. (1994) propose a two-channel approach of the form

$$\begin{aligned} \text{LWP} = & a_0 + a_1 \ln(280 - \text{T22V}) \\ & + a_2 \ln(280 - \text{T37V}), \end{aligned} \quad (12)$$

where  $a_0$ ,  $a_1$ , and  $a_2$  denote the regression coefficients. The transformation  $\ln(280 - \text{Tb})$  linearizes the LWP–Tb relationship. Nevertheless, this kind of linearization is valid only for optically thin atmospheres (e.g., Simmer 1994) since scattering is neglected. The regression coefficients of REG were recalculated with our training dataset. No significant differences were obtained to the original one, so that the original values were further used. This result could be expected since Karstens et al. (1994) used the same radiative transfer model, LWP parameterization scheme, and subset of RAOBs as used in this study.

The results for ALL/85V and REG for low and high LWPs are shown in Fig. 3. Obviously, the scatter is reduced for the NN-retrieved results. This is true in particular for cloud-free cases and LWP larger than 1.0 kg m $^{-2}$ . For the latter, REG shows a similar behavior if  $a_0$ ,  $a_1$ , and  $a_2$  were recalibrated for the whole LWP range. Note that Karstens et al. (1994) calibrated their algorithm only for LWPs less than 1.0 kg m $^{-2}$ . The low accuracy at high LWPs might be due to the fact that the linearization is not valid for optically thick clouds. The results of the NN algorithm ALL/85V agree very well with the parameterized LWP over the whole range, but there is a tendency to underestimate LWPs less than 0.02 kg m $^{-2}$  (Fig. 3a). This behavior is more strongly pronounced for the two-channel algorithm (not shown). The results for LOW resemble those for ALL/85V (not shown).

The performances of ALL/85V, LOW, TWO, and REG are summarized in Table 3. The bias is not included because it is close to zero for all algorithms. TWO is twice as accurate as REG for all LWP ranges and explains significantly more variance. Using more than two channels (LOW and ALL/85V) provides a significant gain in accuracy and explained variance (Table 3). The differences between LOW and ALL/85V are negligible.

Since, in general, measurements of LWP are not available to validate the algorithm results, a test of the algorithm behavior at LWP = 0.0 is often the only solution to this problem (see, e.g., Karstens et al. 1994).

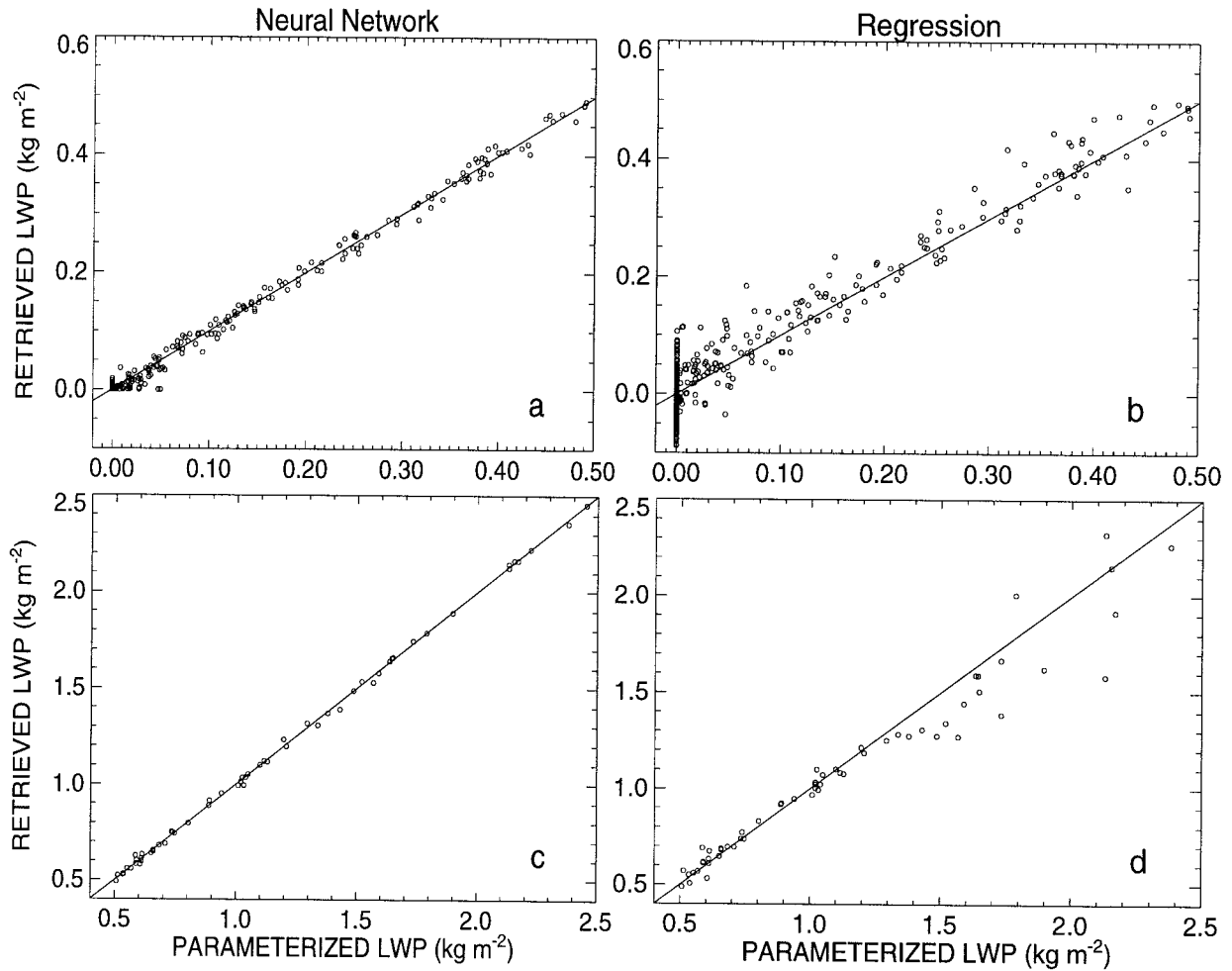


FIG. 3. Retrieved vs parameterized LWP for the independent test set. (a) and (c) Neural network (ALL/85V) for low ( $LWP < 0.5 \text{ kg m}^{-2}$ ) and high ( $LWP \geq 0.5 \text{ kg m}^{-2}$ ) LWP, respectively. (b) and (d) Regression algorithm (REG) for low and high LWP, respectively. Line of perfect agreement (solid).

However, algorithms can have problems at the low end of the LWP range. Therefore, we investigate the NN algorithm behavior for  $LWP = 0.0$ . The normalized frequency distributions of the results of REG, TWO, LOW, and ALL/85V in clear-sky conditions are shown

in Fig. 4. The three NN-based algorithms identify clear-sky cases much more accurately than the regression algorithm that shows a relatively flat distribution. For LOW and ALL/85V, approximately 80% and 90% of all clear-sky cases ( $n = 691$ ), respectively, were cor-

TABLE 3. Summary of retrieval results from four different algorithms for the independent test set. The results are subdivided into three different LWP ranges. Here  $B$  denotes the explained variance, and  $\Delta B$  is the corresponding 99% confidence interval for the lower/upper boundary. The rms errors are given in ( $\text{kg m}^{-2}$ ), and  $B$  as well as  $\Delta B$  in (%).

Algorithm	All LWPs			$LWP < 0.5 \text{ kg m}^{-2}$			$LWP \geq 0.5 \text{ kg m}^{-2}$		
	rms	$B$	$\Delta B^*$	rms	$B$	$\Delta B^*$	rms	$B$	$\Delta B^*$
REG	0.044	98.81	-0.21/+0.17	0.032	94.51	-0.96/+0.82	0.125	96.86	-2.79/+1.49
TWO	0.023	99.76	-0.04/+0.04	0.018	98.87	-0.21/+0.17	0.061	99.24	-0.69/+0.36
LOW	0.011	99.96	-0.01/+0.00	0.009	99.81	-0.04/+0.03	0.027	99.85	-0.14/+0.07
ALL/85V	0.009	99.97	-0.01/+0.00	0.008	99.87	-0.02/+0.02	0.022	99.89	-0.09/+0.05
	$n = 1030$			$n = 964$			$n = 66$		

\* To account for deviations from the normal distribution,  $\Delta B$  values were estimated using a Fisher transformation.

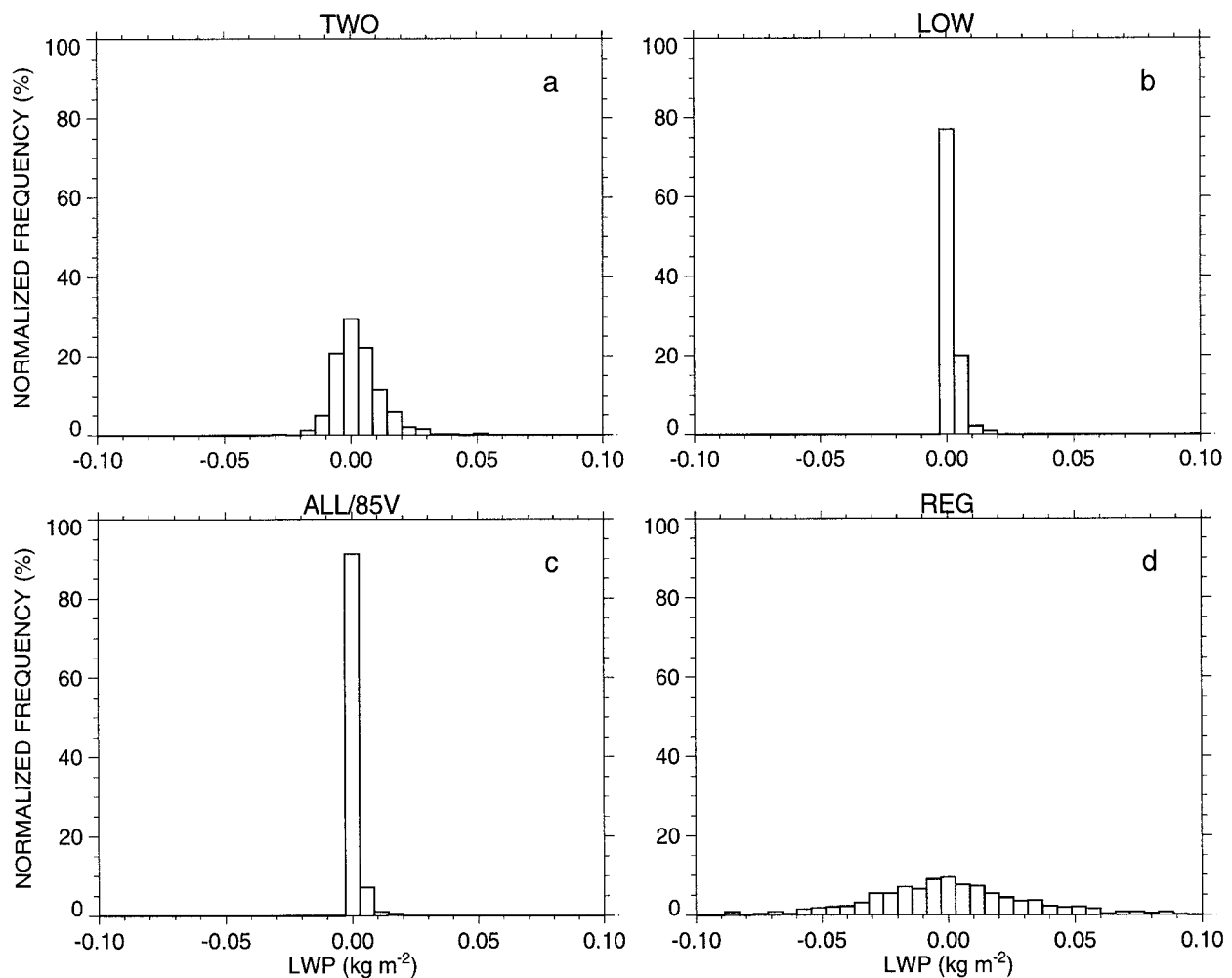


FIG. 4. Normalized (by sample size) frequency distribution of the retrieved LWP for clear-sky cases and different algorithms (interval width is  $0.006 \text{ kg m}^{-2}$ ): (a) TWO, (b) LOW, (c) ALL/85V, and (d) REG.

rectly estimated within the range of  $\pm 0.006 \text{ kg m}^{-2}$ . The corresponding clear-sky statistic is summarized in Table 4. Again, the bias is negligible for all algorithms. The standard deviation for REG is significantly higher than those for TWO, LOW, and ALL/85V. ALL/85V shows the lowest standard deviation of  $2.0 \times 10^{-3} \text{ kg m}^{-2}$ . A remarkable feature of ALL/85V is that no negative values were retrieved (Table 4). The weights  $a_{10}$

and  $a_{11}$  in (6) were adjusted in such a manner (during the training process) that ALL/85V yields no negative values.

Statistical results show that the large clear-sky standard deviation of REG is due partly to the influence of other geophysical parameters such as surface wind speed ( $V$ ), total precipitable water (TPW), and SST. The linear correlations and their error estimates between the retrieved LWP under clear-sky conditions and these three geophysical parameters are shown in Table 5. The results of REG are significantly correlated with all three parameters. The explained variances, which contribute to the “clear-sky noise,” range from 10% for TPW to 25% for  $V$  and SST. The results of TWO are also significantly correlated with  $V$ ; however, the correlation is much smaller with SST and TPW. The results of LOW are influenced only by the SST, and those of ALL/85V show no significant correlations with all three parameters within the range of uncertainty.

Note, however, that only linear correlations were con-

TABLE 4. Deviation of the retrieved values for the test set ( $n = 691$ ) from the expected clear-sky value  $\text{LWP} = 0.0$ . Given are the minimum, maximum, and average values, as well as the width of the distribution (std dev) of the LWP retrieved by different algorithms.

Algorithm	Retrieval statistics ( $\text{kg m}^{-2}$ )			
	Min	Max	Avg	Std dev
REG	$-8.8 \times 10^{-2}$	$1.1 \times 10^{-1}$	$0.4 \times 10^{-3}$	0.031
TWO	$-2.8 \times 10^{-2}$	$5.1 \times 10^{-2}$	$2.7 \times 10^{-3}$	0.009
LOW	$-2.7 \times 10^{-3}$	$1.9 \times 10^{-2}$	$1.1 \times 10^{-3}$	0.003
ALL/85V	$0.2 \times 10^{-3}$	$1.8 \times 10^{-2}$	$0.9 \times 10^{-3}$	0.002



TABLE 5. Dependencies of the (parameterized) clear-sky retrievals of four different algorithms on total precipitable water, surface wind speed, and SST;  $r$  is the correlation coefficient between the algorithm results and the three parameters, and  $\Delta r$  is the corresponding 99% confidence interval for the lower/upper boundary. The presented statistics are based on 691 cases.

Algorithm	Total precipitable water		Wind speed		SST	
	$r$	$\Delta r$	$r$	$\Delta r$	$r$	$\Delta r$
REG	-0.31	-0.40/-0.21	0.52	0.44/0.59	-0.51	-0.58/-0.43
TWO	0.10	0.00/0.20	0.48	0.40/0.55	-0.18	-0.27/-0.08
LOW	0.03	-0.08/0.13	-0.03	-0.13/0.07	-0.25	-0.34/-0.15
ALL/85V	0.08	-0.02/0.18	-0.08	-0.18/0.02	-0.04	-0.14/0.06

sidered, so that the given dependencies may have to be regarded as lower boundaries. For example, REG systematically overestimates LWP up to  $0.1 \text{ kg m}^{-2}$  mainly for dry atmospheres with  $\text{TPW} \leq 15 \text{ kg m}^{-2}$ , which contributes to the negative sign of the correlation coefficient.

The validation with the independent test set described above is based on data that were generated (parameterized) in the same manner as the training set. To test the algorithms for "other clouds," which are more independent, we have changed the LWP parameterization; that is, the relative humidity threshold was set to 90% instead of 95% (section 2). For the 1030 independent RAOBs, the LWP and  $T_b$ 's were recalculated using the lower threshold. As a result, the relationship between the temperature-humidity profiles and the cloud properties (height, water content) has changed.

As expected, the behavior of the different algorithms has not changed significantly for clear-sky conditions. To estimate the sensitivity of the algorithms for cloudy conditions, two intervals were investigated:  $0 < \text{LWP} \leq 0.5 \text{ kg m}^{-2}$  and  $\text{LWP} \geq 0.5 \text{ kg m}^{-2}$ . The latter range is more or less of theoretical interest since in the real

atmosphere clouds with LWPs higher than  $0.5 \text{ kg m}^{-2}$  are commonly believed to rain (e.g., Karstens et al. 1994), and rain is not included in the training set.

The correlation coefficients as well as the biases are not affected by the change in the relative humidity threshold. For  $0 < \text{LWP} \leq 0.5 \text{ kg m}^{-2}$ , the increase in rms errors range from 0.001 (ALL/85V and REG) to  $0.004 \text{ kg m}^{-2}$  (TWO).

The increase in rms error for  $\text{LWP} \geq 0.5 \text{ kg m}^{-2}$  is more pronounced:  $\Delta \text{rms} \approx 0.01 \text{ kg m}^{-2}$  (ALL/85V, LOW, and REG). The increase for TWO is three times larger ( $=0.036 \text{ kg m}^{-2}$ ). The larger rms errors are mainly caused by a shift in the LWP distribution toward higher values for which the retrieval errors are also larger (not shown). Therefore, the main conclusions given above still hold in the case when the algorithms are validated with clouds that due to a change in the parameterization method are different from those used during the training procedure.

#### b. Indirect validation for clear-sky cases

After the test of the algorithms with simulated data, direct measurements of the SSM/I are used in this section. Again no actual observations of cloud liquid water content are available; thus, the test is carried out for the clear-sky cases. The clear-sky noise and cross-talk problems can be investigated with observed data, too.

Before the algorithms could be applied, the  $T_b$ 's had to be adjusted for systematic differences between the results of the model used for the algorithm development and the SSM/I observations. Corrections were estimated from comparisons under clear-sky conditions between simulated and measured  $F-8$   $T_b$ 's (Fuhrhop and Simmer 1996). The adjustments are quite similar for the  $F-10$  and  $F-11$  instruments (H. Gäng 1997, personnel communication).

The standard deviation for the retrieved LWPs for clear-sky cases (classified with IR-Meteosat data) from measured SSM/I  $T_b$ 's for REG, TWO, and ALL/85V as a function of latitude are depicted in Fig. 5. The results of LOW are similar to those of ALL/85V and therefore are not shown. REG shows the highest scatter with a dependency on the geographical latitude. The standard deviation for TWO are lower, but latitudinal dependency is still evident. The clear-sky noise of ALL/85V is considerably reduced. All algorithms estimate

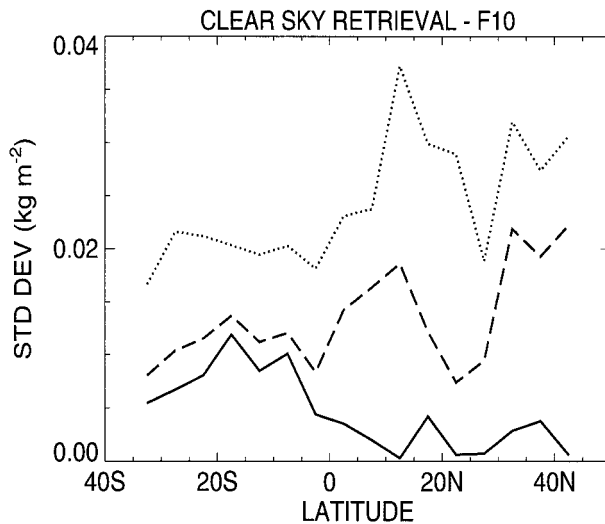


FIG. 5. Meridional cross section over the Atlantic Ocean of the std dev of retrieved LWP for clear-sky conditions and different algorithms: (solid) ALL/85V, (dashed) TWO, and (dotted) REG. The retrievals are based upon measured  $T_b$ 's ( $F-10$ ) and were classified as cloud free with independent IR-Meteosat data.

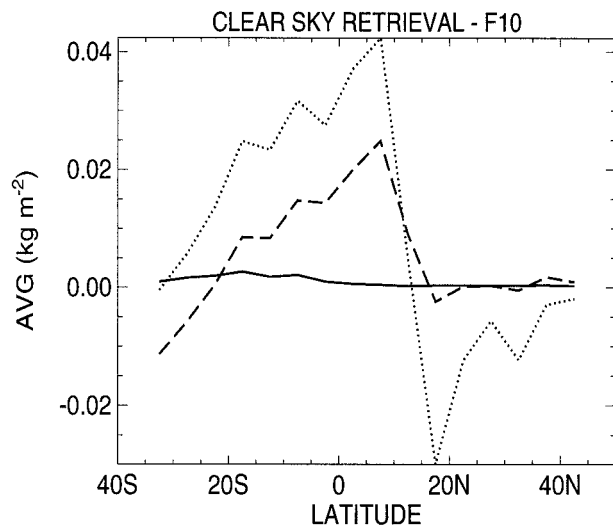


FIG. 6. Meridional cross section over the Atlantic Ocean of the averaged retrieved LWP for clear-sky conditions and different algorithms: (solid) ALL/85V, (dashed) TWO, and (dotted) REG. The retrievals are based upon measured Tb's ( $F-10$ ) and were classified as cloud free with independent IR-Meteosat data.

LWPs exceeding  $0.1 \text{ kg m}^{-2}$  in few cases (less than 1%), especially over the Southern Hemisphere and near  $40^\circ\text{N}$ . This might be due in part to errors in the clear-sky classification procedure.

The biases of the retrieved LWPs under the same clear-sky conditions (which is equal to the average retrieval) are shown in Fig. 6. The biases for LOW (not shown) and ALL/85V are independent of latitude and therefore independent of the total TPW. In contrast, there are significant latitudinal dependencies for TWO and REG. This behavior is somewhat masked by a negative peak near  $15^\circ\text{N}$  for REG. Overall, however, the correlations with TPW are obviously positive for both algorithms. The corresponding normalized frequency distributions are presented in Fig. 7. They are like those for the parameterized clear-sky cases (Fig. 4) except for positive biases for TWO and REG.

Clear-sky statistics and correlations between retrieved LWPs and TPWs (Alishouse et al. 1990a), as well as surface wind speeds  $V$  calculated with the retrieval algorithms (Goodberlet et al. 1989) are summarized in Table 6. As noted before, there are significant positive biases for REG and TWO, and negligible positive biases for LOW and ALL/85V. The standard deviation for REG is twice as high as for TWO. ALL/85V shows the highest accuracy with a standard deviation of  $0.006 \text{ kg m}^{-2}$ . The standard deviation of REG agrees well with that for the simulated data, but for the NN-based algorithms the standard deviation is higher, which is mainly caused by outliers that were not present in the simulated data. Note that only the two-channel NN (TWO) shows a significantly positive bias. This might be due in part to the positive correlations ( $r = 0.38$ ) with TPW (Table 6) that are not present for LOW and ALL/85V ( $0.07$

and  $-0.03$ ). The mean TPW for the clear dataset (SSM/I estimate) is  $30 \text{ kg m}^{-2}$ , which is approximately  $4 \text{ kg m}^{-2}$  higher than the mean TPW in the training and test set. Consequently, moist conditions are somewhat overrepresented, so that the positive biases may be due in part to the TPW influence. Note that the sign of the correlation with TPW has reversed for REG in comparison to the parameterized clear-sky cases (Table 5). The negative sign for the latter dataset is due mainly to an overestimation for dry atmospheres, which are underrepresented in the former set. Additionally, the overestimation of REG at high humidity values is much more pronounced for the measured clear-sky cases (not shown). Surface wind speed  $V$  explains 1% of the variance of the results of REG and TWO (Table 6). In contrast, LOW and ALL/85V show no significant correlation with  $V$ . The positive sign of the correlation coefficients for the two-channel approaches is due mainly to an overestimation of LWP for  $V$  higher than  $12 \text{ m s}^{-1}$  (not shown).

## 6. Summary and conclusions

The potential of NNs for the development of retrieval algorithms for LWP from SSM/I data are studied. The effects of NN architecture and different predictors are investigated, and an intercomparison with a standard regression algorithm (Karstens et al. 1994) is presented.

The NN is trained with a dataset comprising 1030 pairs of simulated Tb's and parameterized LWPs that are based on global distributed RAOBs over the oceans.

The choice of the NN architecture has a significant influence on the retrieval accuracy. Important architecture parameters are the number of neurons in the hidden layer (only one layer was employed). Multiple random searches in weight space were necessary in order to get a *good* minimum of the cost function and to ensure that the solution found during the training procedure was not purely accidental. This might be mainly a problem of variable metric and conjugate gradient methods because standard back propagation is less likely to get stuck in local minima (van der Smagt 1994). Nevertheless, second-order methods are far superior with respect to learning time.

A validation of the algorithm is presented that relies on 1030 independent pairs of Tb's and LWPs, simulated and parameterized in the same manner as for the training set. The NN algorithms show significantly better retrieval results in comparison to the regression algorithm, especially for clear-sky cases and high LWPs ( $\text{LWP} > 0.5 \text{ kg m}^{-2}$ ). For the latter, however, the results are of more theoretical interest since clouds contain rainwater for such LWP (Karstens et al. 1994) and rain was not included in the training and test set. A direct intercomparison between regression and NNs was possible since both algorithms use the same Tb's as predictors (T22V and T37V). Using these two channels reduces but does not fully cancel the problems of clear-sky noise and

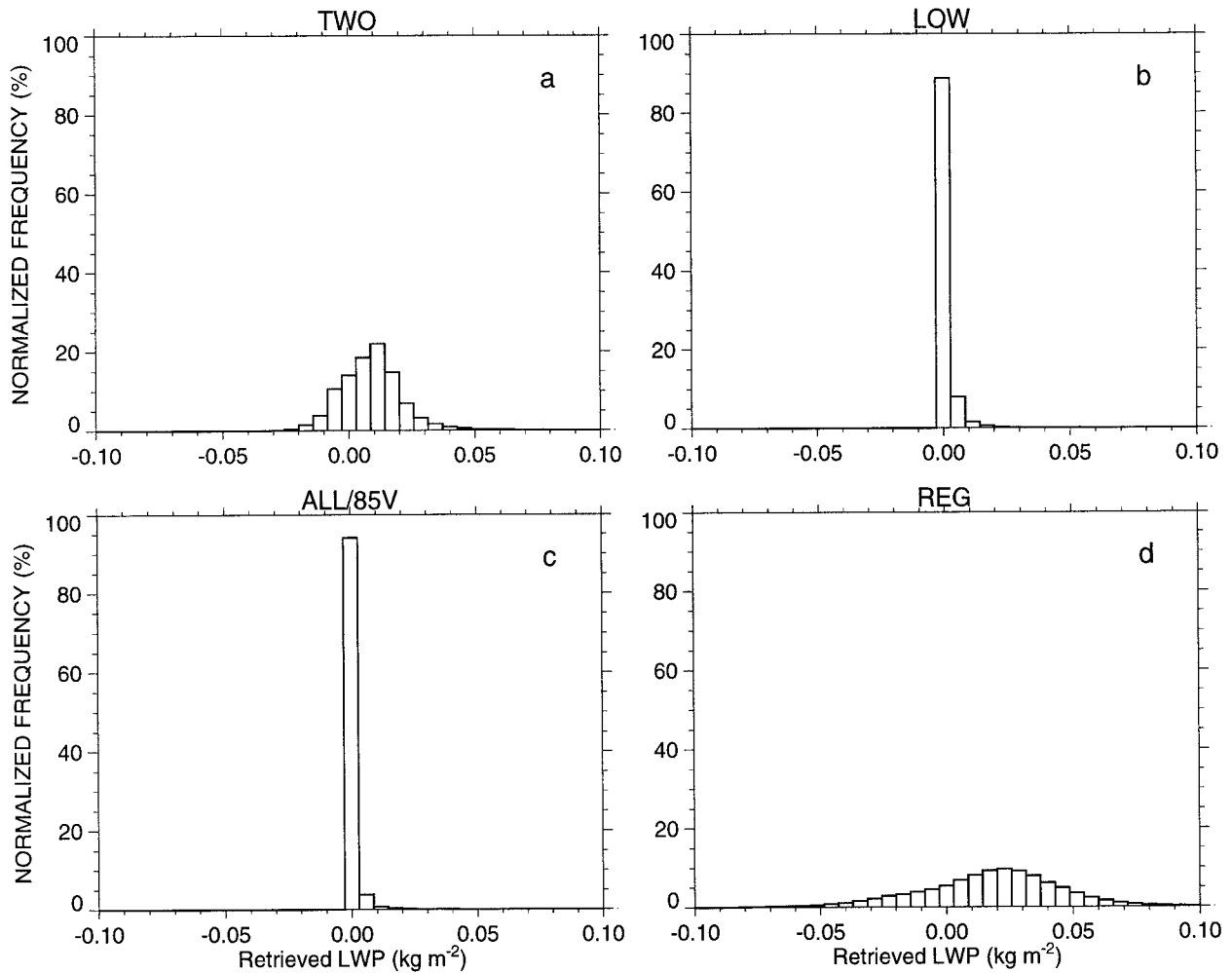


FIG. 7. Normalized (by sample size) frequency distribution of the retrieved LWP from SSM/I observations under clear-sky conditions for different algorithms: (a) TWO, (b) LOW, (c) ALL/85V, and (d) REG (interval width is 0.006 kg m<sup>-2</sup>).

TABLE 6. The summary of retrieval results for clear cases (identified with Meteosat data) of four different algorithms applied to SSM/I measurements (*F-10*). The correlations between the LWP retrievals and the total precipitable water (Alishouse et al. 1990a) as well as the surface wind speed (Goodberlet et al. 1989) calculated from the same Tb's are also presented. The presented statistics are based on 70 000 cases.

Algorithm	Avg (kg m <sup>-2</sup> )	Std dev (kg m <sup>-2</sup> )	Retrieval correlation with	
			Total precipitable water	Surface wind speed
REG	$17.5 \times 10^{-3}$	$29.6 \times 10^{-3}$	0.29	0.10
TWO	$9.4 \times 10^{-3}$	$15.5 \times 10^{-3}$	0.38	0.08
LOW	$-1.0 \times 10^{-5}$	$7.6 \times 10^{-3}$	0.07	0.02
ALL/85V	$1.1 \times 10^{-3}$	$6.4 \times 10^{-3}$	-0.03	0.02

cross talk with TPW. These effects can be reduced significantly if NNs with more SSM/I channels as predictors (LOW and ALL/85V) are used in order to correct for the influence of other geophysical parameters. The advantage of NNs lies in the fact that they can extract information even if the channels are redundant and the *linear* correlations with the geophysical parameters are small. This is apparent from clear-sky investigations (clear-sky noise and cross talk) and high LWP cases. For the latter, the linearization of the regression model (12), which is valid only under nonscattering conditions, is not appropriate. With respect to LWP, NNs can approximate the (unknown) Tb-LWP relationship, taking into account nonlinear effects such as scattering. These results agree quite well with those from Stogryn et al. (1994) and Krasnopolsky et al. (1995), who found that the wind speed retrieval for the SSM/I radiometer is also possible under more adverse conditions (high LWPs) when NNs are used instead of regression algorithms.

A direct validation of LWP algorithms is difficult to perform since measurements of LWP over the whole horizontal scale ( $\approx 50$  km) and vertical column do not exist. To test the algorithms for more *independent* cases, two additional validations are presented. Changing the relative humidity threshold for the LWP parameterization changes the vertical distribution and total amount of cloud water, and with this the relationship between the vertical structure of temperature as well as humidity and the parameterized clouds. Using this more independent test set of Tb's and LWPs does not change the conclusions described above. The second indirect validation is based on measured Tb's (*F-10* instrument) for clear-sky cases, classified with independent IR-Meteosat data. Applying this set of noisier Tb's shows that clear-sky noise and cross talk is also removed for LOW and ALL/85V applications to SSM/I observations.

To estimate the influence of noisier Tb's under cloudy conditions we performed Monte Carlo experiments. Normally distributed, uncorrelated, and zero mean noise with standard deviations between 0 and 5 K and  $\Delta \text{std dev} = 0.5$  was added to the simulated Tb's (clean apart from the radiometer's noise) of the test set. For each noise level this procedure was repeated 50 times (i.e., 50 realizations of 11 different noise levels and 1030 pairs). The rms errors as estimates of the scatter as well as the biases were calculated for all algorithms. The rms ratio between the NN (TWO) and the regression algorithm (REG) using the same predictors increases from 0.6 for zero noise to 0.8 for an added noise higher than 3 K. The rms difference, however, remains constant. The increase in bias for higher noise levels is greater for TWO than for REG. These results still hold when noise is added only to one channel. The sensitivity to noise in channels T22V and T37V (the other held fixed) is significantly reduced for LOW and ALL/85V in comparison to the two-channel algorithms because the other noise-free channels stabilize the retrieval. These results enhance our confidence that the algorithms give better results than the regression algorithms, even when noisier Tb's are used as input, that is, applications of observed data.

To estimate the retrieval error for the new algorithms under cloudy conditions ( $\text{LWP} \leq 0.5 \text{ kg m}^{-2}$ ), we assume that the rms ratio between clear-sky and cloudy conditions remains equal when *measured* instead of *simulated* data are used (first-order error estimate). Taking the estimates from Tables 3 and 4 (simulated) and Table 6 (measured), we can solve for the unknown rms error. Using this procedure, the rms errors for  $\text{LWP} \leq 0.5 \text{ kg m}^{-2}$  become equal for REG and TWO ( $\approx 0.031 \text{ kg m}^{-2}$ ). However, the error estimates for LOW and ALL/85V are much smaller ( $\approx 0.023 \text{ kg m}^{-2}$ ). Note that possible classification errors (IR clear-sky detection) propagate into these error estimates. Another indirect validation can be performed with ground-based observations. Such measurements are currently being carried out by our group, but more data are needed for a sufficient test.

As noted above, the new algorithms are not valid when rain clouds are present. To exclude such cases, we propose a polarization threshold method that relies on the Tb's at 37 and 85 GHz. If we assume that clouds rain if  $\text{LWP} > 0.5 \text{ kg m}^{-2}$  (Karstens et al. 1994), the new algorithms should not be used if the polarization (vertical minus horizontal polarization) is lower than 40 or 7 K for 37 or 85 GHz, respectively. The LWP algorithms can be applied, however, even when ice clouds are present because these clouds are included in the training set. Problems might arise due to inhomogeneously distributed clouds in the radiometer's field of view—the “beam-filling problem,” the influence of which is difficult to estimate since no quantitative numbers can be given.

New NN algorithms that are also applicable under rainy conditions are currently under development. The theoretical results, that is, the retrievals for  $\text{LWP} > 0.5 \text{ kg m}^{-2}$ , presented in this study indicate that NNs are able to map high nonlinearities. An NN-based LWP algorithm, which is still applicable when rain is present, will be a considerable improvement to the algorithms presented in this study.<sup>1</sup>

We propose to use NNs instead of regression techniques when nonlinearities in the radiative transfer equation become important during the retrieval process because the form of the inverse radiative transfer operator is a priori unknown in such cases. If the linearization of the radiative transfer equation is valid over the whole range, for the retrieval of total precipitable water, for example, the advantage of NNs is comparably small (Jung 1996). We would like to note that the computational cost of the new algorithms is comparable to the regression method, once training has been finished, because only the feed-forward phase has to be processed [(4)–(6)]. This aspect may be important, for example, when long times series of geophysical fields have to be calculated from satellite observations.

*Acknowledgments.* The authors would like to acknowledge Dr. Rolf Fuhrhop for valuable discussions and for providing the radiative transfer model. Dr. Holger Gäng is acknowledged for providing the clear-sky dataset. We also wish to thank three reviewers for constructive comments on this manuscript. This work was supported by the German Research Foundation (DFG) under Contract Ru 375/5-1.

#### REFERENCES

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

<sup>1</sup> The new LWP algorithms, described in this paper, are available in the form of FORTRAN functions from the World Wide Web at <http://www.ifm.uni-kiel.de/me/research/Projekte/RemSens/hypam.html>, or directly from the corresponding author.

- , J. B. Snider, E. R. Westwater, C. T. Swift, C. S. Ruf, S. A. Snyder, J. Vongsathorn, and R. R. Ferraro, 1990b: Determination of cloud liquid water content using the SSM/I. *IEEE Trans. Geosci. Remote Sens.*, **28**, 817–822.
- Atkinson, P. M., and A. R. L. Tatnall, 1997: Neural networks in remote sensing. *Int. J. Remote Sens.*, **18**, 699–709.
- Butler, C. T., and R. v. Z. Meredith, 1996: Retrieving atmospheric temperature parameters from DMSP SSM/T-1 data with a neural network. *J. Geophys. Res.*, **101**, 7075–7083.
- Cabrera-Mercader, C. R., and D. H. Staelin, 1995: Passive microwave relative humidity retrievals using feedforward neural networks. *IEEE Trans. Geosci. Remote Sens.*, **33**, 1324–1328.
- Campbell, N. A., 1996: The decorrelation stretch transformation. *Int. J. Remote Sens.*, **17**, 1939–1949.
- Churnside, J. H., T. A. Stermitz, and J. A. Schroeder, 1994: Temperature profiling with neural network inversion of microwave radiometer data. *J. Atmos. Oceanic Technol.*, **11**, 105–109.
- Cober, G. C., A. Tremblay, and G. A. Isaac, 1996: Comparison of SSM/I liquid water paths with aircraft measurements. *J. Appl. Meteor.*, **35**, 503–519.
- Colton, M. C., and G. A. Poe, 1994: Shared Processing Program, Defense Meteorological Satellite Program, Special Sensor Microwave/Imager Algorithm Symposium, 8–10 June 1993. *Bull. Amer. Meteor. Soc.*, **75**, 1663–1669.
- Crone, L. J., L. M. McMillin, and D. S. Crosby, 1996: Constrained regression in satellite meteorology. *J. Appl. Meteor.*, **35**, 2023–2035.
- Elsner, J. B., and A. A. Tsonis, 1992: Nonlinear prediction, chaos, and noise. *Bull. Amer. Meteor. Soc.*, **73**, 49–60.
- Ferraro, R. R., F. Weng, N. C. Grody, and A. Basist, 1996: An eight-year (1987–1994) time series of rainfall, clouds, water vapor, snow cover, and sea ice derived from SSM/I measurements. *Bull. Amer. Meteor. Soc.*, **77**, 891–905.
- Fletcher, R., and M. J. D. Powell, 1963: A rapidly convergent descent method for minimization. *Comput. J.*, **6**, 163–168.
- Fuhrhop, R., and C. Simmer, 1996: SSM/I brightness temperature corrections for incident angle variations. *J. Atmos. Oceanic Technol.*, **13**, 246–254.
- Goodberlet, M. A., C. T. Swift, and J. C. Wilkerson, 1989: Remote sensing of ocean surface winds with the Special Sensor Microwave/Imager. *J. Geophys. Res.*, **94**, 14 547–14 555.
- Goody, R. M., and Y. L. Yung, 1989: *Atmospheric Radiation—Theoretical Basis*. Oxford University Press, 519 pp.
- Greenwald, T. J., G. L. Stephens, T. H. Vonder Haar, and D. L. Jackson, 1993: A physical retrieval of cloud liquid water over the global oceans using Special Sensor Microwave/Imager (SSM/I) observations. *J. Geophys. Res.*, **98**, 18 471–18 488.
- Harrison, E. F., P. Minnis, B. R. Barkstrom, V. Ramanathan, R. D. Cess, and G. G. Gibson, 1990: Seasonal variation of cloud radiative forcing derived from the Earth Radiation Budget Experiment. *J. Geophys. Res.*, **95**, 18 687–18 703.
- Hastenrath, S., L. Greischar, and J. van Heerden, 1995: Prediction of the summer rainfall over South Africa. *J. Climate*, **8**, 1511–1518.
- Hollinger, J. P., R. Lo, G. Poe, R. Savage, and J. Peirce, 1987: Special Sensor Microwave/Imager user's guide. Tech. Rep., Naval Research Laboratory, Washington, DC, 120 pp.
- Jung, T., 1996: Bestimmung des Wasserdampf- und Flüssigwassergehaltes über den Ozeanen aus simulierten Special Sensor Microwave/Imager (SSM/I)-Daten mit neuronalen Netzen. M.S. thesis, Diplomarbeit am Institut für Meereskunde an der Christian-Albrechts-Universität Kiel, Kiel, Germany, 87 pp. [Available online at <http://www.ifm.uni-kiel.de/me/research/Projekte/RemSens/hypam.html>.]
- Karstens, U., C. Simmer, and E. Ruprecht, 1994: Remote sensing of cloud liquid water. *Meteor. Atmos. Phys.*, **54**, 157–171.
- Krasnopolsky, V. M., L. C. Breaker, and W. H. Gemmill, 1995: 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.
- Liu, G., and J. A. Curry, 1993: Determination of characteristic features of cloud liquid water from satellite microwave measurements. *J. Geophys. Res.*, **98**, 5069–5092.
- Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, 1992: *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, 963 pp.
- Prigent, C., A. Sand, C. Klapisz, and Y. Lemaître, 1994: Physical retrieval of liquid water contents in a North Atlantic cyclone using SSM/I data. *Quart. J. Roy. Meteor. Soc.*, **120**, 1179–1207.
- Rojas, R., 1995: *Neural Networks—A Systematic Introduction*. Springer-Verlag, 467 pp.
- Rumelhart, D. E., G. E. Hinton, and R. J. Williams, 1986: Learning representations by back-propagating errors. *Nature*, **323**, 533–536.
- Sengupta, S. K., and J. S. Boyle, 1995: Nonlinear principal component analysis of climate data: Program for climate model diagnosis and intercomparison. PCMDI Rep. 29, Lawrence Livermore National Laboratory, Livermore, CA, 21 pp.
- Simmer, C., 1994: *Satellitenfernerkundung hydrologischer Parameter der Atmosphäre mit Mikrowellen*. Verlag, 313 pp.
- Slingo, A., 1990: Sensitivity of the earth's radiation budget to changes in low clouds. *Nature*, **343**, 49–51.
- Stogryn, A. P., 1972: The emissivity of sea foam at microwave frequencies. *J. Geophys. Res.*, **77**, 1658–1666.
- , C. T. Butler, and T. J. Bartolac, 1994: Ocean surface wind retrievals from Special Sensor Microwave/Imager data with neural networks. *J. Geophys. Res.*, **99**, 981–984.
- Tang, B., G. M. Flato, and G. Holloway, 1994: A study of arctic sea ice and sea-level pressure using POP and neural network methods. *Atmos.–Ocean*, **32**, 507–529.
- Tsang, L., Z. Chen, S. Oh, R. J. Marks, and A. T. C. Chang, 1992: Inversion of snow parameters from passive microwave remote sensing measurements by a neural network trained with a multiple scattering model. *IEEE Trans. Geosci. Remote Sens.*, **30**, 1015–1024.
- van der Smagt, P. P., 1994: Minimization methods for training feedforward neural networks. *Neural Networks*, **7**, 1–11.
- von Storch, H., and F. W. Zwiers, 1997: *Statistical Analysis in Climate Research*. Cambridge University Press, in press.
- Wagner, F., 1996: Manual for Netfit (Version 1.2). Institut für Theoretische Physik an der Christian-Albrechts-Universität Kiel, Kiel, Germany, 11 pp.
- , and C. Lovelace, 1971: Phase shift analysis of  $\Pi-p \rightarrow K1^b$ . *Nucl. Phys.*, **25B**, 411–427.
- Warner, J., 1955: The water content of cumuli form cloud. *Tellus*, **7**, 449–457.
- Weng, F., and N. C. Grody, 1994: Retrieval of cloud liquid water using the Special Sensor Microwave/Imager (SSM/I). *J. Geophys. Res.*, **99**, 25 535–25 551.
- Wisler, M. M., and J. P. Hollinger, 1977: Estimation of marine environmental parameters using microwave radiometric remote sensing systems. NRL Tech. Rep. 3661, Naval Research Laboratory, Washington, DC, 27 pp.