An Evaluation of Neural Networks for Retrieval of Water and Temperature Profiles from SSM-T/T2/I Soundings

Howard E. Motteler

UMBC
NOAA
NASA/GSFC

May 25, 1995

1. Introduction

Neural networks have been used with with some success for retrieval of temperature and water profiles from microwave and infrared soundings [4, 6]. In this report, neural networks are compared with regression for retrieval of water and temperature profiles, both for SSM/T/T-2/I [2] soundings with co-located RAOB data, and for simulated AMSU-A and B brightness temperatures. The neural nets used in these tests have one input for each microwave channel, two hidden layers, and one output for each pressure level of the retrieved profile. The nets used for SSM retrievals also have local and solar zenith angle as inputs. Network training is done with a modified form of back-propagation. For each test, neural networks are compared with a form of regression (used operationally at NOAA for a number of years) in which an orthogonal basis is found for the dependent set, and is then used to provide a compact representation for data in the independent set [8, 6].

In general, neural nets out-performed regression in all tests, though generally by a small margin; the greatest difference was for temperature retrievals performed directly from antenna temperature and beam position. For temperature profile retrievals by both neural nets and regression, adding the SSM-I and T-2 channels gave a slight improvement in accuracy, while working directly with antenna temperatures and beam angle gave a slight decrease in accuracy. For water profile retrievals, adding the SSM-I channels help regression accuracy slightly, and did not significantly effect the neural net accuracy. Distinct and averaged SSM-T2 and SSM-I FOVs gave the same results with regression, while the neural net did slightly better with averaged FOVs.

2. Methods

In this section we briefly describe the neural net and regression methods we use for retrievals. The regression and neural net approaches have a number of features in common, including fast computation of retrievals; also, both methods provide very good linear and non-linear fits, respectively, to noisy, correlated data.

2.1. Neural Networks

The neural networks we use for retrievals are three-layer feed-forward networks, trained with a modified form of back-propagation [5, 7]. Such networks can be represented as a vector valued function $F(\mathbf{x}) = W_3 F_2(W_2 F_1(W_1 \mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2) + \mathbf{b}_3$, where F_1 and F_2 map vectors to vectors by applying a *transfer function* to each vector component. The mapping F_i is often referred to as a *layer*, with the weight matrices representing connections between layers. We use the hyperbolic tangent as a transfer function in the first two layers, and a linear function in the third. The vector $\mathbf{x}$ is the input to the network, the W_i are *weight matrices* and the $\mathbf{b}_i$ are *bias vectors*.

Back-propagation training is a variation of gradient descent, in which a *training set*, a sequence of input/output vector pairs, is presented to the net, and weight and bias vectors

are adjusted incrementally in an attempt to match network output with training set output. A single presentation of all the training data, with a corresponding weight adjustment, is called an *epoch*. Training consists of a sequence of epochs, and typically continues until the difference between training data and net output is acceptable, or alternatively, until the net's performance on a separate set of test data is acceptable. Training is a computationally intensive process for non-trivial networks. In contrast, applying a trained net is very fast, with the runtime being dominated by the time for the three matrix-vector multiplies.

The networks we used for temperature and water retrievals have one input component for each selected instrument channel (including beam angle, when working directly with antenna temperatures) and one output component for each pressure level. For example, the network for experiment 1, which used the T, T2 and I channels, had 21 inputs, 8 elements in each of the two "hidden layers" and 15 outputs. Inputs are brightness or antenna temperatures (depending on the experiment) and the output is a temperature or humidity profile. Inputs and outputs are normalized by subtracting the mean and dividing by the standard deviation of the training set, before presentation to the net.

Training was stopped when the RMS error on the training data stops showing significant improvement; this typically happen at between 2000 and 5000 epochs, for nets of the size used here. As with more traditional methods of interpolation, neural networks can both under- and over-fit. If the hidden layer(s) are too large, and training is continued long enough, the net will over-fit and performance on new data suffers. Over-fitting can also decrease performance on noisy data. On the other hand, if the hidden layer(s) are too small, the net will fail to fit to the training data adequately. For the tests here, the close correspondence between training and testing errors, and the relative smoothness of retrieved profiles, suggest that the size of our hidden layers is not too large, and that we are not over-fitting, while the fact that training errors increase when the hidden layers are made smaller, or when only a single hidden layer is used, suggest that we are not under-fitting.

2.2. Linear Retrievals

We perform linear retrievals using a regression method involving a change of basis for both temperature profiles and brightness temperature spectra, as proposed by Smith and Woolf [8]. The change of basis can give a significant compression of radiance and/or profile data, and makes the regression less sensitive to noise.

The basic idea is to find distinct minimal orthogonal bases for both a representative set of temperature profiles and for the corresponding brightness temperature spectra, for a selected set of channels. Suppose T is an $n \times k$ array of k temperature profiles at n pressure levels, one profile per column, with $k > n$. Let V be the $n \times n$ matrix whose columns are eigenvectors of TT^{T}, ordered by the eigenvalues. (The matrix V can be obtained either by a conventional eigenvector algorithm, or by performing a singular value decomposition of T.) Although V is an n-dimensional basis for T, it may be the case that T has an "effective" basis with a smaller dimension.

The basis B we want is the first $m \leq n$ columns of V, where the value of m is estimated from the eigenvalues, and can be checked by testing the accuracy of mappings

into and back out of the m-dimensional space. If B is orthonormal, then we can use B^{T} as a linear transform from profiles into the m-dimensional basis, and B as a transform from the m-dimensional basis back to profiles. For example, if T is the entire set of 1761 TIGR temperature profiles, then T can be represented quite accurately with $m = 35$; in this case we have $\mathrm{RMS}(T - BB^{\mathrm{T}}T) < 10^{-3}$ K. Because upper levels of the 40-level TIGR set are extrapolations, it is not surprising that there is such an accurate 35-dimensional representation. (Note that the mapping in and out of the new basis is not an interpolation, and that components of profiles in the 35-dimensional space do not correspond to pressure levels.)

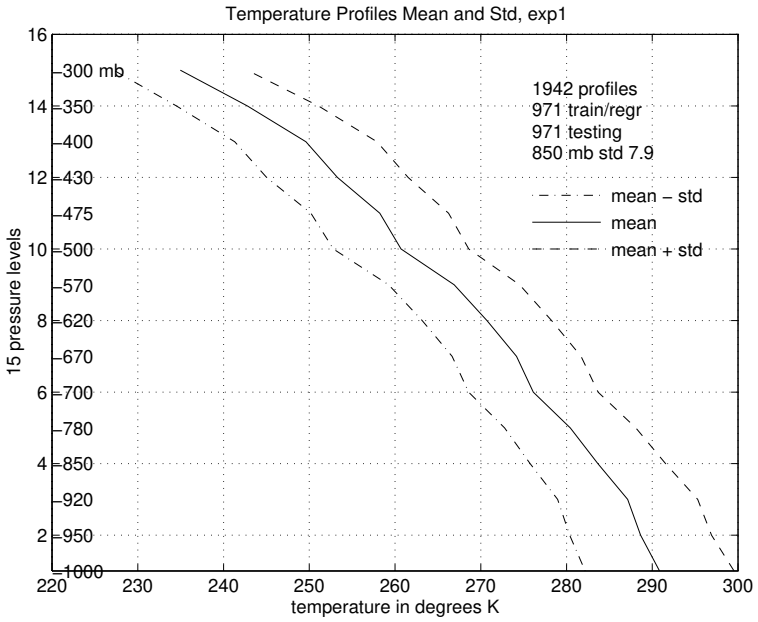

Figure 1: Mean and Variation of Temperature Profiles from RAOB Set 1; 1942 ocean profiles, with both day and night soundings, over a wide range of latitudes. The temperature standard deviation at 850 mb was 7.9K.

A similar change of basis can be performed on the brightness temperature spectra Tb corresponding to T, for various channel sets. Suppose B_1 is a basis for Tb and B_2 a basis for T. Let $Tb' = B_1^{\mathrm{T}}Tb$, $T' = B_2^{\mathrm{T}}T$, and let C be the least squares solution to $CT_b' = T'$. Then $D = B_2CB_1^{\mathrm{T}}$ is the desired linear transform for retrievals. A slight improvement performing retrievals with noisy input (about 0.1K) can be obtained by adding noise after the bases B_1 and B_2 have been obtained, and before the regression. If $\hat{Tb}$ is Tb with added noise, then let $\hat{Tb}' = B_1^{\mathrm{T}}\hat{Tb}$, let C be the least squares solution to $C\hat{T}_b' = T'$, and $D = B_2CB_1^{\mathrm{T}}$.

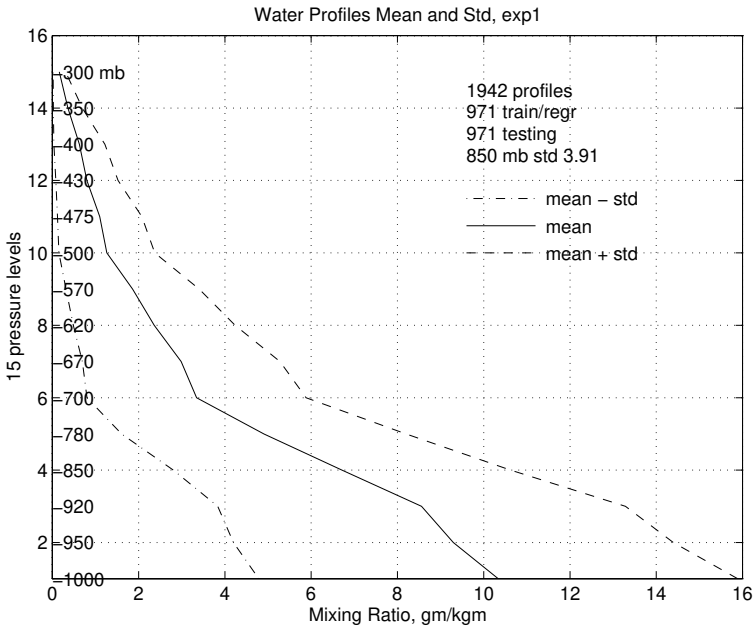

Figure 2: Mean and Variation of Water Profiles from RAOB Set 1. The water standard deviation at 850 mb was 3.91 gm/kg.

---

3. Datasets

Two sets of RAOBS co-located with SSM/T/T2/I soundings were used. The first set (Figure 1, Figure 2, and Figure 3) consisted of 1942 temperature and water profiles, to 300 mb. These profiles were taken over the ocean, both day and night, and at a wide range of latitudes. The standard deviation at 850 mb was 7.9 K for the temperature profiles, and 3.91 gm/kg for the water profiles. The 1942 profiles were split into two sets: 971 for training and regression, and 971 for testing. The second set (Figure 4 and Figure 5) consisted of 1821 temperature profiles, to 50 mb. Profiles were again taken over the ocean, both day and night, and at a wide range of latitudes. The standard deviation at 850 mb was 9.05 K. The 1821 profiles were split into two sets: 910 for training and regression, and 911 for testing.

The SSM/T and T-2 both scan cross-track, with one SSM/T field of view (FOV) containing approximately 16 T-2 FOVs. The SSM/I has a conical scan, with a smaller footprint than the SSM/T and T-2, with the path scanned falling behind the SSM/T and T-2 scan. Co-location data available from [Doug Moore's group], in addition to the RAOB, provides 16 SSM/T-2 FOVs for each SSM/T field, with averaged SSM/I FOVs available for each SSM/T-2 FOV. Limb correction data was available for the SSM/T observations, but only preliminary corrections were available for the SSM/T2 at the time of these tests.

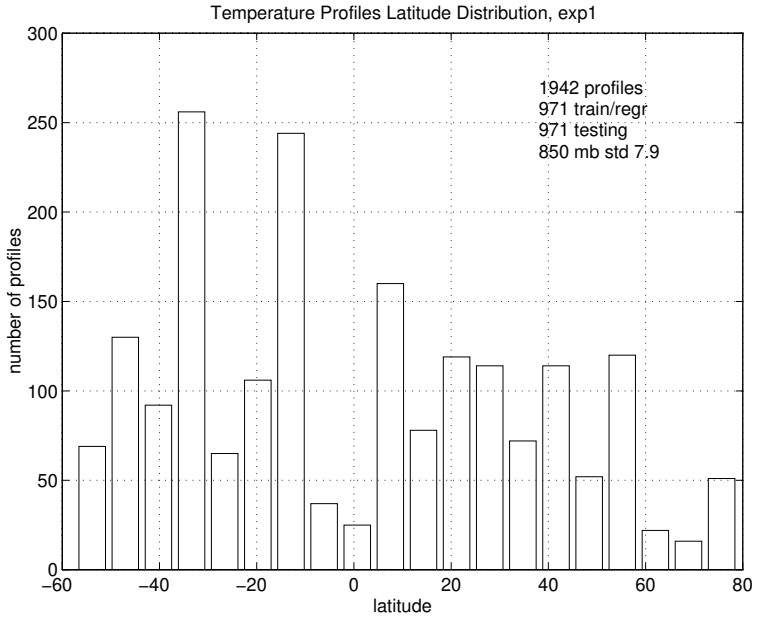

Figure 3: Latitude distribution of RAOB Set 1.

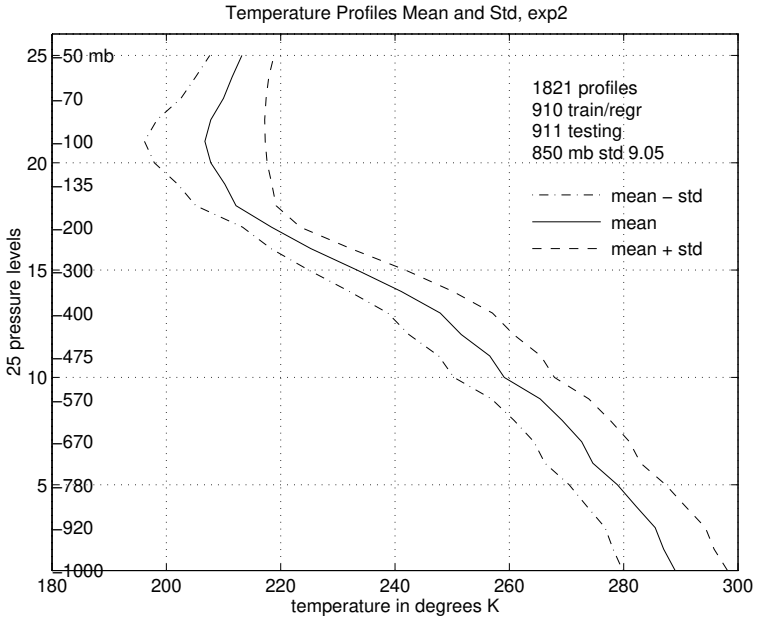

Figure 4: Mean and Variation of Temperature Profiles from RAOB Set 2; This set consists of 1821 ocean profiles, both day and night soundings, over a wide range of latitudes. The temperature standard deviation at 850 mb was 9.05K.

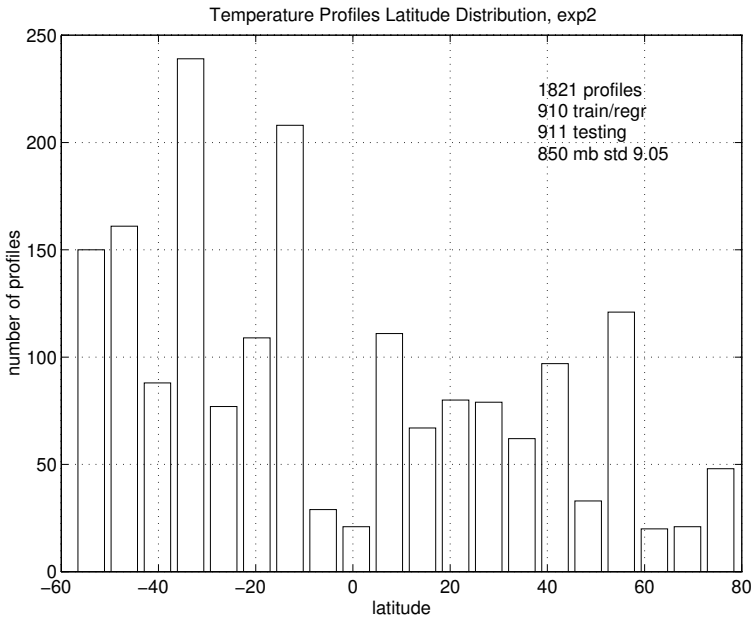

Figure 5: Latitude distribution of RAOB Set 2.

SSM Channels

	SSM/T		SSM/I		SSM/T2	
1	50.5	H	19.35	V	183+-3	V
2	53.2	H	19.35	H	183+-1	V
3	54.35	H	22.235	V	183+-7	V
4	54.9	H	37.0	V	90	V
5	58.4	V	37.0	H	150	V
6	58.825	V	85.5	V		
7	59.4	V	85.5	H		

4. Temperature Retrievals

Four temperature retrieval tests are described; the first with RAOB set 1, providing a set of temperature retrievals to go with the water retrievals described in the next section, and the remaining three with RAOB set 2. Tests two and three compare temperature retrievals with and without the addition of the SSM/T-2 and SSM/I channels. Tests three

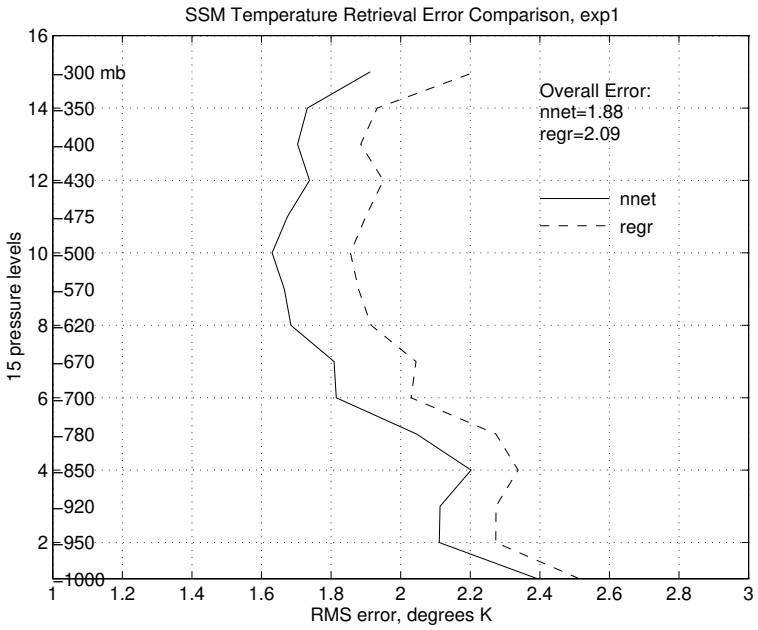

Figure 6: Temperature Retrieval Test 1; SSM/T channels with averaged SSM/T2 and SSM/I FOVs added; RAOB set 1. Input is antenna temperatures and local zenith angle; output is temperature at 15 pressure levels.

---

and four compare retrievals done directly with antenna temperature and beam position with retrievals done using corrected brightness temperatures.

In all of the temperature retrieval tests, in addition to the antenna or brightness temperatures as described below, inputs for retrievals (and for training and regression) included local and solar zenith angles, and a "day/night" flag. All neural networks described in this section used eight hyperbolic-tangent transfer functions in each of two hidden layers; the dimension used for the regression bases was the smaller of the number of inputs and the number of outputs.

4.1. Temperature Retrieval Test 1

The first temperature retrieval test was done with profiles from RAOB set 1 (Figure 1). Input for retrievals (and for training and regression) was antenna temperatures and local zenith angle; all beam angles were used except for beam position 7, which was partly obstructed. All SSM/T, T-2, and I channels were used (except for the the SSM/T 59.4 GHz channel 7 and the SSM/T-2 150 GHz channel 5, which were not available); for each SSM/T FOV, and each T-2 and I channel, all 16 associated T-2 and I FOVS were averaged. A temperature profile at 15 pressure levels, up to 300 mb, was retrieved. Results are summarized in Figure 6; the overall neural net error was 0.21K less than the regression error.

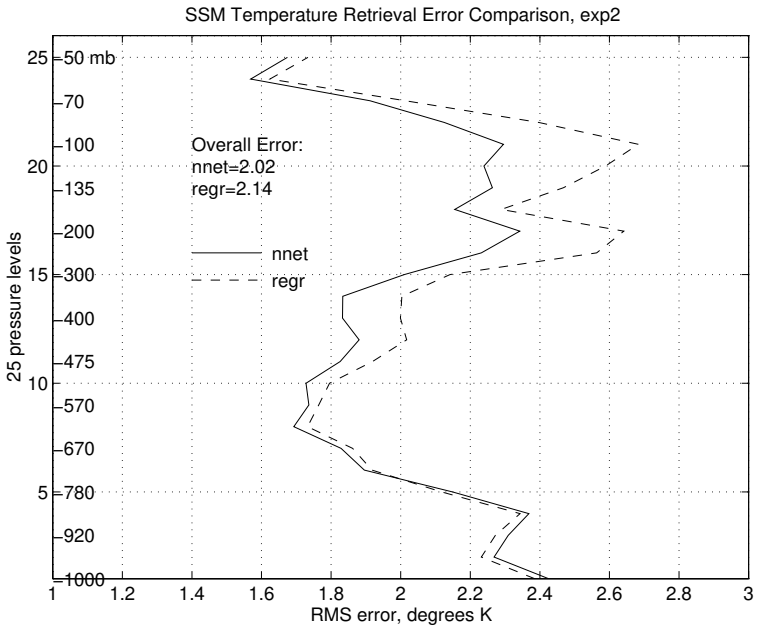

Figure 7: Temperature Retrieval Test 2: SSM/T channels with averaged SSM/T2 and SSM/I FOVs added; RAOB set 2. Input is antenna temperatures and local zenith angle; output is temperature at 25 pressure levels.

4.2. Temperature Retrieval Test 2

The remaining three temperature retrieval tests were done using profiles from RAOB set 2 (Figure 4). For the second test, the same sort of inputs were used as in the first test: antenna temperatures, local zenith, and all channels except SSM/T channel 7 and SSM/T-2 channel 5, with averaged T-2 and I fields. A temperature profile at 25 pressure levels, up to 50 mb, was retrieved. Results are summarized in Figure 7; the overall neural net error is 0.12K less than the regression error.

4.3. Temperature Retrieval Test 3

For the third test, the six available SSM/T channels were used alone. Inputs were antenna temperatures and local zenith angle (and also solar zenith and a day/night flag). A temperature profile at 25 pressure levels, up to 50 mb, was retrieved. Results are summarized in Figure 8; the overall neural net error is 0.32 K less than the regression error. When the T2 and I channels are dropped, the neural net error increased by just under 0.1 K, while the regression error increased by just under 0.2 K; adding the SSM-I and T-2 channels gives a slight improvement in temperature retrieval accuracy.

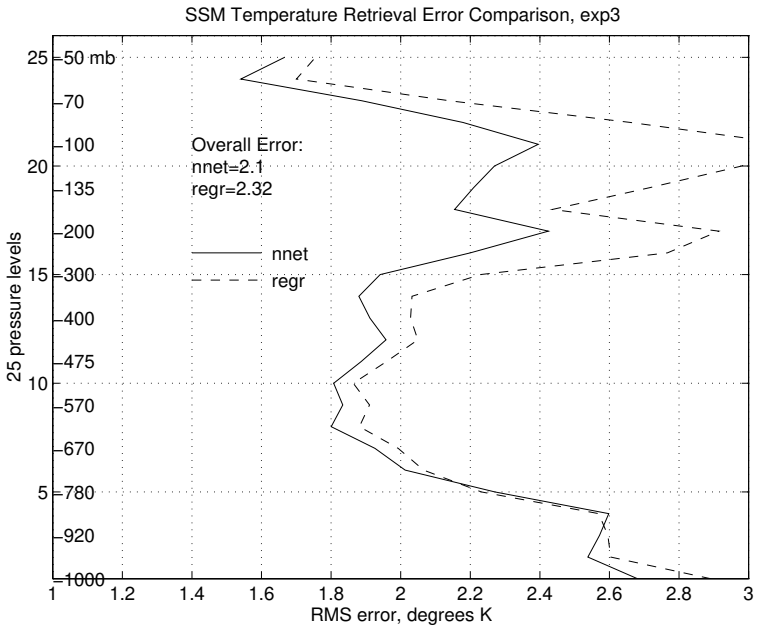

Figure 8: Temperature Retrieval Test 3: SSM/T channels only; RAOB set 2. Input is antenna temperatures and local zenith angle; output is temperature at 25 pressure levels.

4.4. Temperature Retrieval Test 4

For the fourth test, SSM/T channels were again used alone. This time inputs are corrected brightness temperature for the six SSM/T channels. A temperature profile at 25 pressure levels, up to 50 mb, was retrieved. Results are summarized in Figure 9; the overall neural net error is 0.21 K less than the regression error.

In going from antenna temperatures to corrected brightness temperatures, both the regression and neural net error decreased; the regression error by about 0.2 K and the neural net by about 0.1 K. Results with neural nets and regression are very close when using corrected brightness temperatures. In comparison with test three, the neural net does slightly better than regression using beam position and antenna temperatures; this may be particularly useful in those cases where corrected brightness temperatures are not available.

5. Water Retrievals

Three water retrieval tests are described, all using RAOB set 1. The first test used 16 distinct SSM/T-2 FOVs for each SSM/T FOV; the second test adds SSM/I channels, and the third test uses the average of the 16 SSM/T-2 and SSM/I FOVs; the third test thus

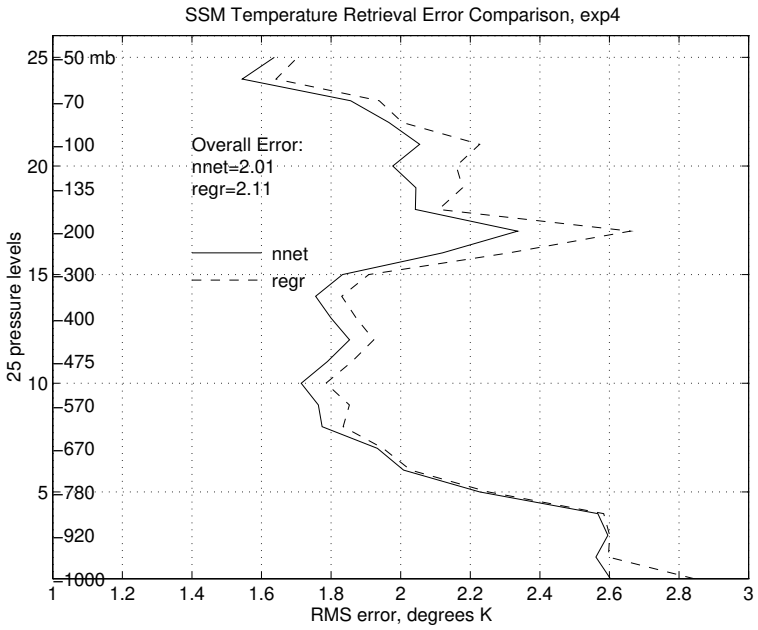

Figure 9: Temperature Retrieval Test 4: SSM/T channels only; RAOB set 2. Input is corrected brightness temperatures; output is temperature at 25 pressure levels.

---

uses the same inputs as temperature tests 1 and 2, and exactly the same input data as was used for temperature test 1.

As with the temperature retrieval tests, in addition to the antenna or brightness temperatures as described below, inputs for retrievals (and for training and regression) included local and solar zenith angles, and a "day/night" flag. The dimension used for the regression bases was the smaller of the number of inputs and the number of outputs.

5.1. Water Retrieval Test 1

For the first water retrieval test, 16 distinct SSM/T-2 FOVs were used for each SSM/T FOV, for a total of 73 inputs: 6 SSM/T channels; local and solar zenith angles; day/night flag; and 16×4 SSM/T-2 antenna temperatures. The neural net had 16 hyperbolic-tangent functions in the first hidden layer, and 12 in the second; a dimension of 10 was used for the regression bases. Mixing ratio in gm/kgm was retrieved at 15 pressure levels, up to 300 mb. Results are summarized in Figure 10; the overall neural net error is 5% less than the regression error.

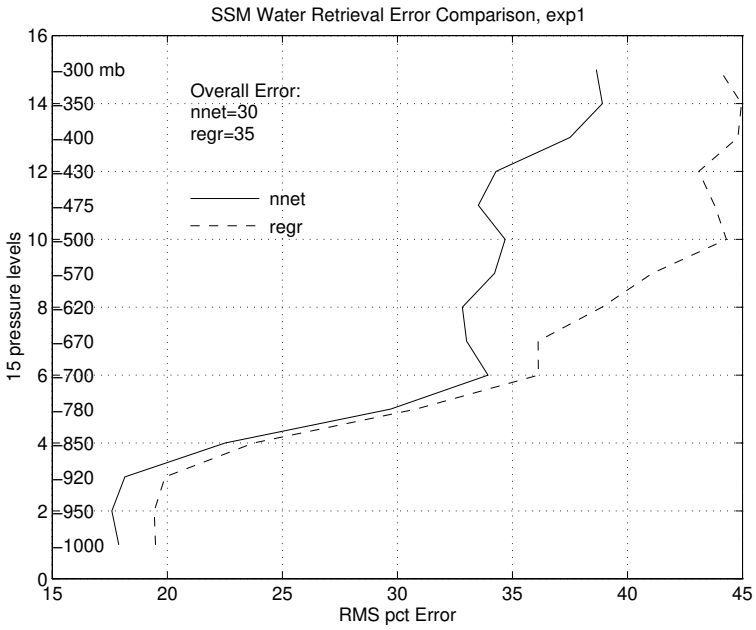

Figure 10: Water Retrieval Test 1: SSM-T with 16 distinct SSM-T2 FOVs for each SSM/T footprint; RAOB Set 1.

5.2. Water Retrieval Test 2

For the second water retrieval test, the 16×7 averaged SSM/I values corresponding to the SSM/T-2 FOVs were added, for a total of 185 inputs. The neural net had 16 hyperbolic-tangent functions in the first hidden layer, and 12 in the second; a dimension of 10 was used for the regression bases. Mixing ratio in gm/kgm was retrieved at 15 pressure levels, up to 300 mb. Results are summarized in Figure 11; the overall neural net error is 2% less than the regression error. In comparison with test 1, we see that the SSM/I channels help regression accuracy slightly.

5.3. Water Retrieval Test 3

For the third water retrieval test, the SSM/T channels and the average of the 16 SSM/T-2 and SSM/I FOVs was used; the inputs for water retrieval test 3 and temperature retrieval test 1 are thus exactly the same. The neural net had 8 hyperbolic-tangent functions in the first hidden layer, and 8 in the second; a dimension of 10 was used for the regression bases. Mixing ratio in gm/kgm was retrieved at 15 pressure levels, up to 300 mb. Results are summarized in Figure 12; the overall neural net error is 4% less than the regression error, and results overall were slightly better than for the first two tests. Distinct and averaged SSM/T2 and SSM/I FOVs gave the same results with regression, while the net did slightly better with averaged FOVs.

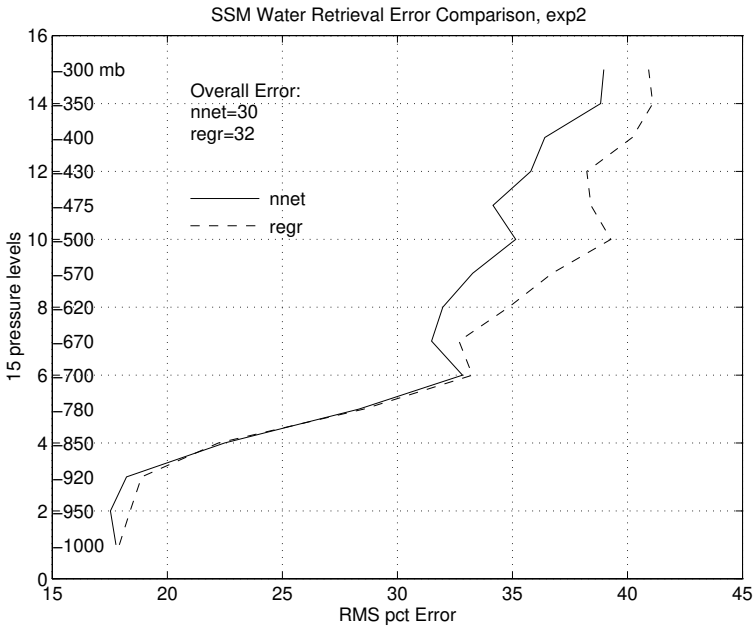

Figure 11: Water Retrieval Test 2: SSM-T with 16 distinct SSM-T2 FOVs for each SSM/T footprint and closest SSM-I FOVs averaged for each SSM-T2 FOV; RAOB Set 1.

---

6. Simulation Studies

Before the co-located SSM data became available, a simulation study was done using the 1761 TIGR temperature and water profiles [3] (Figure 13) and simulated brightness temperatures at the 20 AMSU (AIRS microwave sounder) channels [1]. The AMSU and SSM/T/T-2/I channel sets are close enough for a rough comparison, and the AMSU-B channels include the 183 GHz water line.

In these tests, all 20 AMSU channels were used as inputs for both temperature and water retrievals; the view was assumed to be at nadir. Temperature was retrieved at 64 levels, to 4 mb, and water at 30 levels, to 300 mb. The neural nets had 20 hyperbolic-tangent functions in both hidden layers and dimension of 20 was used for the regression bases. Emissivity was fixed at 1 for the temperature tests and 0.5 for the water tests, and noise with a standard deviation of 0.5K was added to the simulated brightness temperatures.

Figure 14 shows RMS error retrieving 880 TIGR temperature profiles at 64 TOVS pressure levels, and RMS percentage error retrieving 880 TIGR water profiles at the lowest 30 TOVS pressure levels. The overall retrieval errors were as follows:

Method	temperature	water
20 channel neural net	2.16 K	25.1 %
20 channel regression	2.22 K	34.2 %

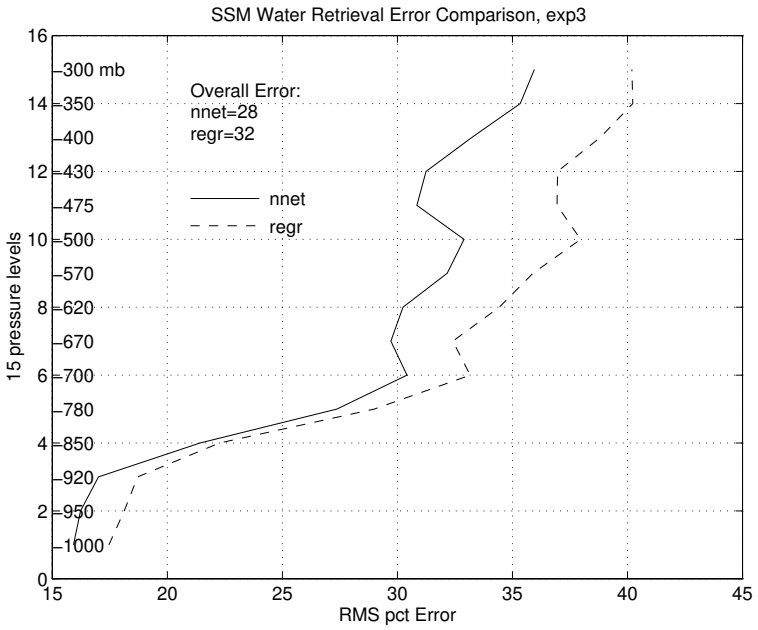

Figure 12: Water Retrieval Test 3: SSM/T with averaged SSM/T2 and SSM/I FOVs; RAOB Set 1.

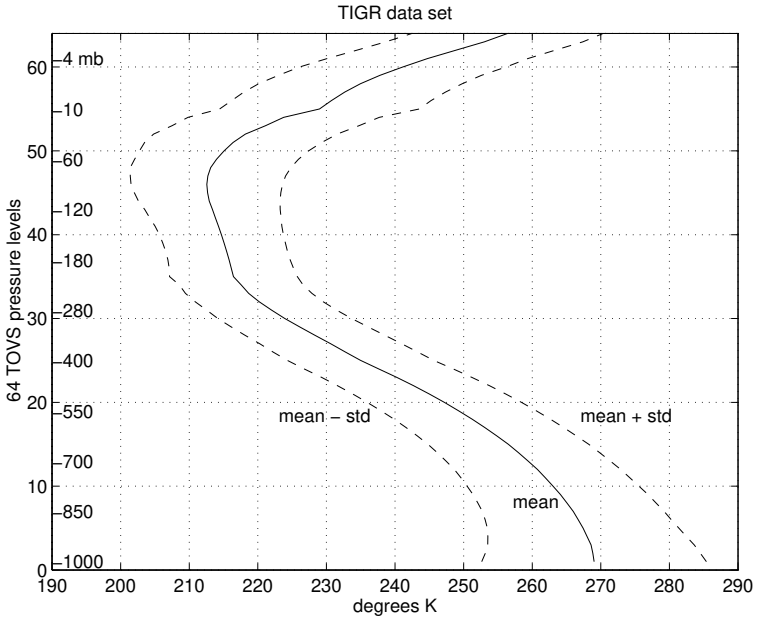

Figure 13: Mean and standard deviation of 1761 TIGR temperature profiles, used for AMSU simulation studies.

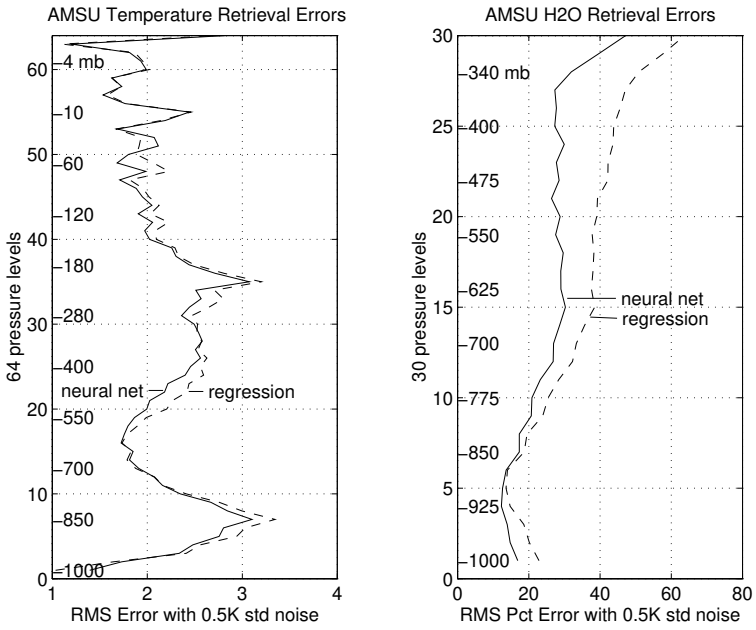

Figure 14: Temperature and water retrieval error using TIGR profiles and simulated AMSU A and B brightness temperatures.

These results are in rather close agreement with our results using actual data, however this should be viewed with a degree of caution: in some ways the AMSU tests may have been harder, because of the somewhat greater variability of the TIGR profile set and the greater number of levels retrieved, while in other ways the AMSU tests were easier, due to the single (nadir) beam position, and other simplifications inevitable in a simulation study.

7. Conclusions

Neural networks have many of the advantages of regression, are almost as easy to use, for moderate-sized channel sets, and are potentially more accurate. A neural net might also be useful in providing a good "first guess" for an iterative physical retrieval system. Also, the good results we had using neural nets directly with antenna temperatures suggests they might provide a way to get a reasonably accurate retrieval system on-line very quickly, for a new instrument.

In general, neural nets out-performed regression in all tests, though generally by a small margin; the greatest difference was for temperature retrievals performed directly from antenna temperature and beam position. For temperature profile retrievals by both neural nets and regression, adding the SSM-I and T-2 channels gave a slight improvement in accuracy, while working directly with antenna temperatures and beam angle gave a slight decrease in accuracy. For water profile retrievals, adding the SSM-I channels help regression

accuracy slightly, and did not significantly effect the neural net accuracy. Distinct and averaged SSM-T2 and SSM-I FOVs gave the same results with regression, while the neural net did slightly better with averaged FOVs.

Possible future work includes delivering a production retrieval system; refining results with larger datasets; trying other neural net techniques; trying cloud detection and clearing with SSM/I channels; investigating microwave retrievals over land; and using neural nets or a simple forward model for fast radiative transfer calculations, for a quality check on retrievals.

8. References

[1] Atmospheric infrared sounder: Science and measurement requirements. Technical Report D6665 Rev. 1, Jet Propulsion Laboratory, 1991.

[2] Feasibility study of 183 ghz capability. Technical Report Report 6404, Areojet Electrosystems Company, 1100 West Hollyvale St., Azusa, CA 91702, 1983.

[3] A. Chedin, N. A. Scott, C. Wahiche, and P. Moulinier. The improved initialization inversion method: A high resolution physical method for temperature retrievals from satellites of the tiros-n series. *Journal of Climate and Applied Meteorology*, 24:128–143, 1985.

[4] J. Escobar-Munoz, A. Chedin, F. Cheruy, and N. Scott. Réseaux de neurones multicouches pour la restitution de variables thermodynamiques atmosphériques à l'aide de sonduers verticaux satellitaires. Technical report, Laboratoire de Météorologie Dynamique du CNRS, Ecole Polytechnique, 91128 Palaiseau Cedex, France, January 1993.

[5] Robert Hecht-Nielsen. *Neurocomputing.* Addison-Wesley Publishing Company, New York, NY, 1990.

[6] H. E. Motteler, L. L. Strow, L. McMillin, and J. A. Gualtieri. A comparison of neural networks and regression-based methods for temperature retrievals. *Applied Optics*, to appear 1995.

[7] Patrick K. Simpson. *Artificial Neural Systems.* Pergamon Press, Inc., Elmsford, New York, 1990.

[8] W. L. Smith and H. M. Woolf. The use of eigenvectors of statistical covariance matrices for interpreting satellite sounding radiometer observations. *Journal of the Atmospheric Sciences*, 33(7):1127–1140, July 1976.