Neural Networks for Atmospheric Temperature Retrieval

from AQUA AIRS/AMSU/HSB Measurements

Deming Jiang[1] [3], Chaohua Dong[2], Peng Zhang[2], Xuebao Wu[2]

[1] *Nanjing Meteorological Institute, Nanjing 210044, P.R.China*
[2] *National Satellite Meteorological Center, Beijing, 100081, P.R.China*
[1] *Hunan Meteorological Institute, 163 Yuhua Load, Changsha, Hunan Province 410007, P.R.China*

ABSTRACT

The high spectral resolution sounding suite instruments on EOS Aqua are expected to provide estimated atmospheric profiles with higher accuracy and better vertical resolution. To validate and test it's performance with it's orbital measurements, temperature profiles at 101 pressure levels, up to 0.005 hPa, with 1km vertical resolution at troposphere, were retrieved on different types of terrain with different spectral bands in the middle latitude area by using a three-layered feed-forward neural networks with back-propagation algorithm. Results show that the Qinghai-Tibet Plateau has a measurable impact on the retrieval accuracy which is corresponding to the spectral bands used in performing retrievals. A promising approach to the elimination of this effect is to apply additional predictors which are non-satellite observed (e.g. surface altitude).

Key words: AIRS, Temperature Profile, Retrieval, Neural Networks

1 Introduction

With the development of numerical weather prediction and four-dimensional data assimilation, the accuracy of model-derived atmospheric temperature fields is comparable to or better than those obtained from existing operational satellite soundings (e.g., HIRS). The retrieval vertical resolution is limited mainly by the broadness of the contribution functions (i.e., the weighting function multiplied by the Planck function) of current instruments. When the contribution functions are expanded, emitted energy reaching the satellite in each channel will have components originating from a thick layer of the atmosphere, thereby making the discrimination of fine-scale vertical details practically impossible. This problem is exacerbated by the limited number of HIRS channels. Furthermore, because of the broad width of the contribution functions and the difficulties in eliminating cloud contamination effects, as well as effects from the surface emissivity, O_3, H_2O, and other minor constituents, the RMS errors in the retrieved temperature profiles remain high. Until the launch of Aqua in

2002, the advanced sounding suite system, especially the advanced hyperspectral IR instrument AIRS made this deficiency eliminable (Jun Li 1994).

AIRS (Atmospheric Infrared Sounder) on the EOS Aqua spacecraft that launched on 4 May 2002 from Vandenburg Air Force Base, California, into a 705km altitude sun-synchronous orbit, is an infrared spectrometer/radiometer which employs two multi-aperture slit and is pupil-imaging, with spectral coverage from 3.74 to 4.61μm, from 6.20 to 8.22μm, and from 8.8 to 15.4μm, with nominal spectral resolution $\lambda / \Delta\lambda = 1200$. The spectrum is sampled twice per spectral resolution element for a total of 2378 spectral samples. These large numbers of channels provide sharpen weighting functions, and also permit selection of sounding channels not contaminated by water vapor lines or by emission from other active gases. The AIRS as a key facility instrument on the EOS Aqua satellite, working together with AMSU (Advanced Microwave Sounding Unit) and HSB (Humidity Sounder for Brazil), forms a complementary sounding system for NASA's Earth Observing System (Aumann et al. 2003).

The AMSU-A is primarily a temperature sounder co-aligned with the AIRS instrument, which measures radiant energy with 15 channels between 23 and 89 GHz provides means to independently account for clouds in the AIRS Field of View and corrects the infrared measurements for the effects of clouds so as to resulting in more accurate temperature and humidity retrievals. The HSB is 4 moist sounding channel version of AMSU-B but has footprints of the same spatial resolution as AIRS IR footprints. It detects radiances in the range of 150 to 183.31 GHz to provide the atmospheric information needed to correct infrared measurements (e.g., from AIRS) for the effects of clouds and to determine global humidity profiles in 2-km thick layers in the lower atmosphere (troposphere) (Susskind et al. 1998 and Susskind et al. 2003).

Before and after the AIRS had been launched into its orbit, algorithms and monographs for atmospheric profiles retrieval have been published by various researchers, such as Motteler et al. 1995, Susskind et al. 1998, Kuligowski and Barros, 2001, Cabrera-Mercader and Staelin 1995, Goldberg et al. 2003, Susskind et al. 2003, Chaohua et al. 2004,and etc.

Spectral bands were fine-scaled for AIRS instrument to achieve higher accuracy and better vertical resolution in estimated profiles. This means larger amount of channels data have to be dealt with. In the physical retrieval algorithms, Forward Models (FMs) are the physical kernel to invert atmospheric parameters numerically. In the data assimilation systems, they are used to connect satellite measurements with numerical models. Both numerical inversion and direct assimilation iteratively require FMs and their Jacobeans to calculate many times for each satellite measurement. Thus, the retrieval process becomes very time consuming, sometimes prohibitively expensive for operational applications (Strow et al. 2003). For such applications it is essential to have fast and accurate versions of FMs. Neural networks enable us to construct such fast and accurate FMs

Neural networks technique is a very promising mathematical tool for modeling complicated nonlinear relationships (Attali et al. 1997, Chen and Chen 1995, Hornik 1991, Deming et al. 2003) and is also an appropriate and efficient tool for solving

forward and inverse problems in remote sensing and for developing fast and accurate forward models and accurate and robust retrieval algorithms. (Kuligowski and Barros 2001, Aires et al. 2002, Cronford et al. 2001, Gottsche and Olesen 2002). Various neural inversion techniques have been developed, such as the "iterative inversion" (Kindermann and Linden 1990), the "distal learning" (Jordan and Rumelhart 1992), and etc.

In this paper, a three-layered feed-forward neural network similar to that used by H. E. Motteler et al. in 1995, is applied to estimate atmospheric temperature profiles by using AQUA/AIRS, AMSU and HSB real data.

2 The neural networks inversion approach

The neural networks used in the tests as a atmospheric parameters retrieval system consists of a three-layered Feed-forward networks with the back-propagation algorithm, as is shown in Fig. 1. It can be represented as a vector valued function:

$$\mathbf{F}(\mathbf{p}) = \mathbf{w_3}\mathbf{F_2}(\mathbf{w_2}\mathbf{F_1}(\mathbf{w_1}\mathbf{p}+\mathbf{b_1})+\mathbf{b_2})+\mathbf{b_3},$$

where $\mathbf{F_1}$ and $\mathbf{F_2}$ map vectors to vectors by applying a transfer function to each vector component. The mapping $\mathbf{F_i}$ (i=1, 2, 3) is referred to as a layer, with the weight matrices representing connections between layers. The hyperbolic tangent as a transfer functions in the two hidden layers, and a linear function in the output layer. The vector $\mathbf{p}$ is the input to the network, the $\mathbf{w_i}$ (i=1, 2, 3) are weight matrices and $\mathbf{b_i}$ (i=1, 2, 3) are bias vectors.
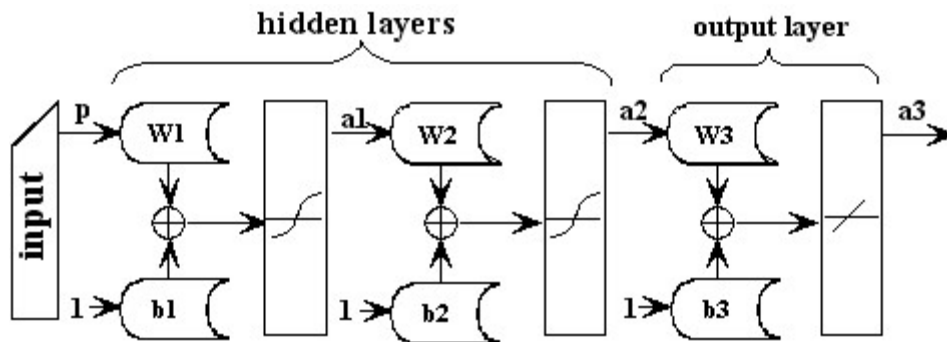

Fig.1 The three-layered feed-forward neural networks employed in the experiments.

The term *back-propagation* refers to the manner in which the gradient is computed for nonlinear multilayer networks. Standard back-propagation is a gradient descent algorithm, in which the network weights are moved along the negative of the gradient of the performance function. There are a number of variations on the basic algorithm that are based on other standard optimization techniques, such as conjugate gradient and Newton methods. The neurons in the two hidden layers are nonlinear hyperbolic tangent activation functions (the sigmoid function) and the output neurons are linear. The input to the network is a vector of PCs (the Principal Components) of the observed radiances obtained from the AIRS, AMSU and HSB measurements (Huang 2001). For each input vector the system produce estimates of atmospheric temperature at pressure levels ranging from 0.005 hPa to 1010 hPa.

The procedure of retrieval can be divided into several steps. The first step is to

divide the data up into training, validation and test subsets. One fourth of the data were taken as the validation set, one fourth as the test set and one half as the training set. The sets were picked as equally spaced points throughout the original data. Then the inputs and targets of the training set were normalized so that they will have zero mean and unity standard deviation. The next step is to perform a principal component analysis - PCA. PCA technique has three effects: it orthogonalizes the components of the input vectors (so that they are uncorrelated with each other); it orders the resulting orthogonal components (principal components) so that those with the largest variation come first; and it eliminates those components that contribute the least to the variation in the data set.
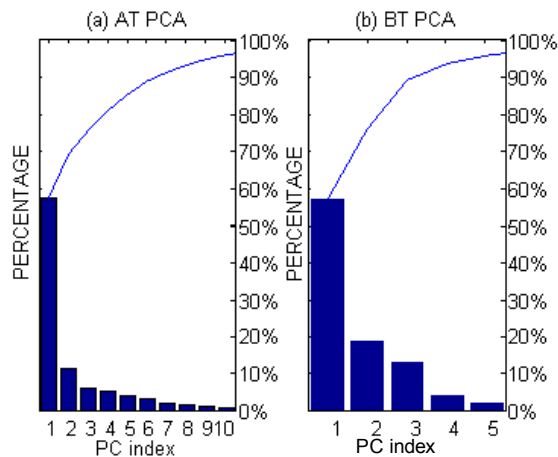
Fig. 2 PCs on complicated terrain. (a) The percentage of the total variance in the atmospheric temperature profile explained by each principal component in explained. (b) The same as (a) but for satellite observed brightness temperature.
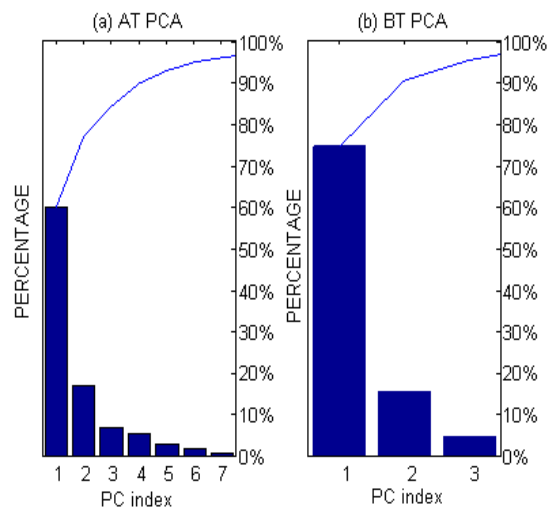
Fig.3 PCs on flat terrain. (a) and (b) The same as in figure 2, but for flat terrain.

Figure 2 and Figure 3 indicated the percentage of the total variance in the observations explained by each principal component in explained. Those principal components have been conservatively retained which account for 99.9998 of the variation in the data set. This means that those principal components that contribute less than 0.0002% to the total variation in the data set will be eliminated. This procedure reduces the size of the input vectors from 2127 to 230 (There are 270 bad

channels among 2378 AIRS channels. These bad channels are eliminated. And 15 AMSU channels and 4 HSB channels are involved in. So we use totally 2127 channels). The third step is to create a network and train it. The last step is to perform some analysis of the network response. The entire data set were input through the network (training, validation and test) and a linear regression was performed between the network outputs and the corresponding targets. With all test data set, the root mean square errors (RMSE) were also computed between the retrieved profiles and the real-time ECMWF analyzing profiles.

3. Training and Retrieving

Dataset we used in the study include totally 11,658 cloud-free samples were derived from 10 granules which numbered as 043, 044, 045, 060, 061, 076, 077, 078, 093, 094, at time 04:17.26s, 04:23.26s, 04:29.26s, 05:59.26s, 06:5.26s, 07:35.26s, 07:41.26s, 07:47.26s, 09:17.26s, 09:23.26s (UTC), on 6 September 2002. Each sample was co-located with the real-time ECMWF analysis profile at 06 (UTC). These samples and profiles were arbitrarily clustered in two classes: (1) on complicated terrain (with Qinghai-Tibet and sea surface involved) located at latitude between $20°N$ and $40°N$; (2) on plain terrain located at latitude between $40°N$ and $60°N$. And each class presented to the networks was divided up into two subsets for training and test, which were picked as equally spaced points throughout the original dataset.

Figure 4 shows the spatial distribution of the observations belonging to each class with respect to the entire ensemble. And figure 5 shows the surface altitude distribution in each sample subset, (a) for the complicated terrain, (b) for the flat terrain.
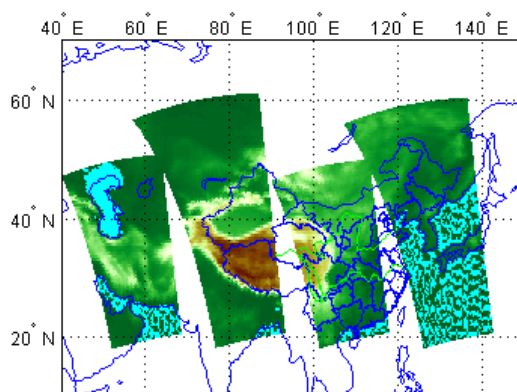

Fig. 4 Sampling measurements coverage (Ascending orbit)

As it was known, the training for a feed-forward network with back-propagation method is a computationally intensive process. Especially for a non-trivial one, it has hundreds of neurons inside so as to adapt the input vectors to the target vectors and the target vectors also have hundreds of components in each. A conventional process for neural networks retrieval system usually has one input component for each selected instrument channel and one output component for each pressure level. Inputs are brightness temperatures and output is a temperature or humidity profile.
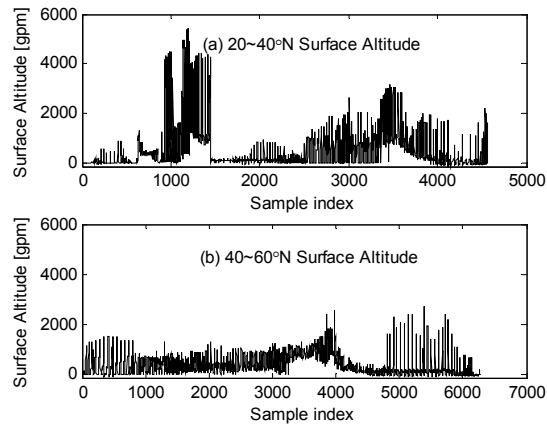
Fig. 5 Surface altitude distribution. (a) the complicated terrain, (b) on the flat terrain.

This design is very time-consuming to train in case for high spectral resolution satellite data and high vertical resolution profile. A typical training process for temperature retrieval is as shown in figure 3. The training process stopped at 5,000 epochs and the training RMS error goal was not met. Motteler et al. trained for 20,000 to 100,000 epochs for smaller RMS error (Motteler et al. 1995). Hence, an improved design is made in our retrieval system.
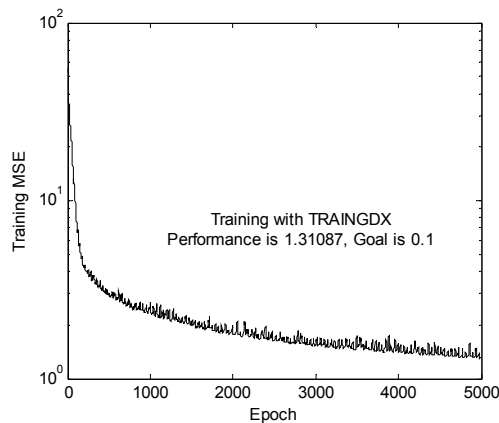
Fig. 6   A typical conventional training process of neural networks for

temperature profile retrieval.

In the training procedures, brightness temperatures (training subset, pretreated by using PCA) and one single layer of temperature as target are presented to the net at a time. Due to only one component in the target vector, the size of networks could be very small. Consequently, the net should cost less training epochs to fit and less time for each single epoch as well. In retrieving phase following the training phase, one single layer of temperature would be output. When temperature at the first pressure level is retrieved, the next layer adjacent to the first layer then begins to perform. At this phase, we use the prior weight matrix to initialize the network. Because the temperatures at adjacent pressure layers should have less difference on value, the weight and bias matrix should need less adjustment and the net should converge to the target more quickly. This process somewhat looks like in the physical retrieval algorithm with a "good first guess". In our tests, this method performed very well and

saved a great quantity of time. And the retrieved profiles look more smoothly.

4. Results

Results are summarized in table 1 and in figure 7 ~ 9.

Spectral coverage sensitivity test: table 1 shows the overall retrieval RMS errors in different terrain by using channels within long wave spectral range (L: 8.8 - 15.4μm), within middle wave range (M: 6.20 - 8.22μm) and within short wave range (S: 3.74 - 4.61μm) and their combinations.

Table 1   Overall RMS errors of spectral coverage

Spectral coverage	20~40°N	40~60°N
L	1.0220	0.8122
M	0.9524	0.7635
S	1.0406	0.8519
L+M	0.9650	0.8191
L+S	0.9864	0.8299
M+S	0.9829	0.8107
L+M+S	0.9687	0.8080
AIRS+AMSU+HSB	0.9795	0.8074
Adding additional predictors	0.8990	0.7510

The middle wave channels perform best, while the short wave channels do the worst, both on the complicated and flat terrain. Using all channels does not yield better result. Microwave channels help improvement in temperature retrieval accuracy within the flat terrain, but just the reverse on the complicated terrain. By adding the surface altitude as an additional predictor, the RMS error decrease. The overall error in the region $20°N \sim 40°N$ is 0.8990 less than 0.9795, and in $40°N \sim 60°N$ is 0.7510 less than 0.8074.

Figure 7 and figure 8 demonstrate RMS errors of temperature retrieval at latitude between $20°N$ and $40°N$, and between $40°N$ and $60°N$ respectively. The dotted line represents the retrieved RMS temperature errors by using IR channels only (AIRS) and the solid is similar but with microwave channels (AMSU & HSB) merged in.

In comparison with figure 7 and figure 8, we see that retrievals performance much better at flat terrain than at complicated surface and microwave channels help to improve accuracy in temperature retrievals at the near surface pressure levels at complicated terrain more obviously than at plain surface.

To confirm if these differences really arose from the Qinghai-Tibet plateau, we eliminated the sample profiles of Qinghai-Tibet from the dataset. The results are summarized in figure 9, in which the solid line indicated the RMS error with the Plateau eliminated. In the comparison of two RMS lines shows the improvement in retrieval accuracy overall non-trivial. In other words, results have posed a major challenge to achieve the nominal accuracy of 1K/1km for temperature retrieval in high spectral resolution soundings: the Qinghai-Tibet will be a major obstacle.
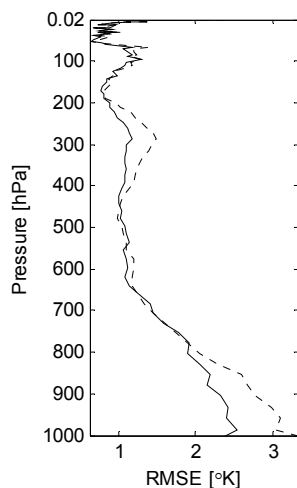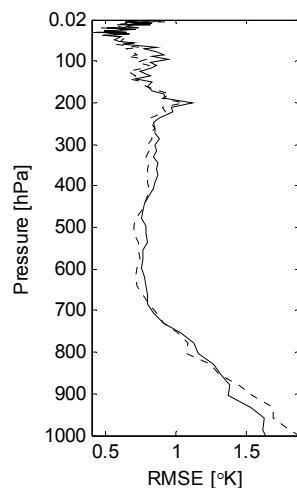
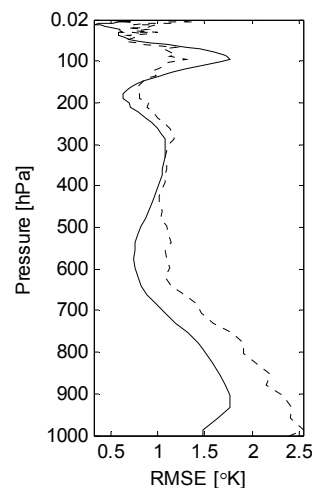**Fig.7 complicated terrain**   **Fig.8 flat terrain**   **Fig.9 Qinghai-Tibet eliminated**

5. Conclusions

1) Different spectral bands have a variety of sensitivity to the topography. The short wave range spectral band is of the most sensitive, while the medium wave channels are of less sensitive. To reduce the impact of terrain, it is necessary to select proper channels to be used in retrieval.

2) Using microwave channels help the improvement of retrieval accuracy at the lower troposphere, but do worse at the upper troposphere, as a result, the overall RMS errors increase.

3) Adding additional predictors, such as surface altitude, can largely help in the improvement of the retrieval accuracy at the complicated terrain.

4) Retrieval errors near the tropopause are still too large and no method was found to get the situation improved. Using GPS data may be a prospective approach (Borbas et al. 2003).

5) High spectral resolution instruments are capable of providing a way prospectively to obtain high vertical resolution retrieved profiles. The greatest difference between actual retrievals and simulated retrievals was the RMS errors at the tropopause, and at pressure levels near the surface, where large errors stick out obviously for actual retrievals.

6) Training for neural networks is a computationally intensive process, but applying a trained net is very fast. Results showed in this paper suggest neural networks are capable of providing a way to get a reasonably accurate retrieval system on-line very quickly, and might also be useful in providing a good "first guess" for an iterative physical retrieval system to obtain better accuracy. The "layer-by-layer" method for profile retrieval proposed in this paper provide an efficient way to training a net and to retrieving profiles, especially for the retrieval of a high vertical resolution profile with fine-scale more than one hundred pressure levels. But some people consider it is not an acceptable retrieval strategy. Because retrieving one temperature at a time does not use the fact that the profiles have regularities. This means that it has irregular, noisy retrieved profiles since no constrains of regularity is used (in the form of covariance

100

matrices generally). So, it is not optimal at all. But this method performed very well and saved a great quantity of time in our test. I can not tell the reason.

ACKNOWLEDGEMENTS

REFERENCES

Aires F, Chedin A, Scott N A, et al., 2002, A regularized neural net approach for retrieval of atmospheric and surface temperature with the IASI instrument. J. App. Met., **41**, 144-159

Attali J, Pagés G., 1997, Approximation of functions by a multilayer perceptron: a new approach. Neural Networks, **10**, 1069-1081.

Aumann H H, Chahine M T, Gautier C, et al., 2003, AIRS/AMSU/HSB on the Aqua Mission: Design, Science Objectives, Data Products, and Processing Systems. IEEE Transactions on Geoscience and Remote Sensing. **41**, 253-264

Borbas E, Menzel W P, Jun Li, et al., 2003, Effects of GPS/RO refractivities on IR/MW retrievals. Proceedings of Thirteenth International TOVS Study Conference, 29 October - 4 November 2003.

Cabrera-Mercader C R, Staelin D H., 1995, Passive microwave relative humidity retrievals using feed-forward neural networks. IEEE Trans. Geosci. Remote Sens., **33**, 1324-1328

Chaohua Dong, Deming Jiang, Chengli Qi, Peng Zhang, 2004, Experimental study on atmospheric parameters using new atmospheric sounding instrument data. Multispectral and Hyperspectral Remote Sensing Instruments and Applications II, Proc. of SPIE, **5655**, 24-32.

Chen T, Chen H., 1995, Approximation capability to functions of several variables, nonlinear functionals, and operators by radial basis function neural networks. Neural Networks, **6**, 904-910

Cronford D., 2001, Nabney I T, Ramage G. Improved neural network scatterometer forward models. J. Geophys. Res., **106**, 22331-22338

Deming, J., Weimin, C., Bingshan, F., Jiankai, W. 2003. A Neural Network Approach to the Automated Cloud Classification of GMS Imagery over South-East China Maritime Regions. Journal of Nanjing Institute of Meteorology, **26** (1), 89-95.

Goldberg M D, Qu Y, McMillin L M, et al., 2003, AIRS near-real-time products and algorithms in support of operational numerical weather prediction. IEEE Trans. Geosci. Remote Sensing, **41**, 379-389

Gottsche F M, Olesen F S., 2002, Evolution of neural networks for radiative transfer calculations in the terrestrial infrared. Remote Sensing of Environment, **80**, 157–164

Hornik K., 1991, Approximation capabilities of multilayer feedforward network. Neural Networks, **4**, 251-257

Huang H. L., Antonelli P., 2001, Application of Principal Component Analysis to High-Resolution Infrared Measurement Compression and Retrieval. J. App. Met., **40**, 365-388

Jordan, M., and Rumelhart, D., 1992, Forward models: Supervised learning with a distal teacher. Cognitive Science, **16**, 307-354.

Jun Li, 1994, The Capability of Atmospheric Profile Retrieval from Satellite High Resolution Infrared   Sounder Radiances. Advances in Atmospheric Science, **12**(2), 255-258

Kindermann, J., and Linden, A., 1990, Inversion of neural networks by gradient descent, Parallel computing, **14**, 277-286.

Kuligowski R J, Barros A P., 2001, Combined IR-Microwave Satellite Retrieval of Temperature and Dewpoint Profiles Using Artificial Neural Networks. J. App. Met., **40**, 2051-2067.

Motteler H E, Larrabee S L, McMillin L, et al., 1995, Comparison of neural networks and regression-based methods for temperature retrievals.   Applied Optics, **34**(24), 5390-5397

Strow L L, Hannon S, Machado S, et al., 2003, An Overview of the AIRS Radiative Transfer Model. IEEE Transactions on Geoscience and Remote Sensing, **41**(2), 303-313.

Susskind J, Atlasa R, Barnet C D, et al., 2003, Current results from AIRS/AMSU/HSB. Proceedings of Thirteenth International TOVS Study Conference, 29 October - 4 November 2003.

Susskind J, Barnet C D, Blaisdell J M., 1998, Determination of atmospheric and surface parameters form simulated AIRS/AMSU/HSB sonding data:retrieval and cloud clearing methodology. Adv.Space Res., **21**(3), 369-384.

Susskind J, Barnet C D, Blaisdell J M., 2003, Retrieval of atmospheric and surface parameters from AIRS/AMSU/HSB data in the presence of clouds. IEEE Trans. Geosci. Remote Sensing, **41**, 390-409

INTERNATIONAL
ATOVS
WORKING GROUP

Proceedings of the Fourteenth International TOVS Study Conference

Beijing, China
25-31 May 2005