

RESEARCH

Open Access

# Neural network Jacobian analysis for high-resolution profiling of the atmosphere

William J Blackwell

## Abstract

Neural networks have been widely used to provide retrievals of geophysical parameters from spectral radiance measurements made remotely by air-, ground-, and space-based sensors. The advantages of retrievals based on neural networks include speed of execution, simplicity of the trained algorithm, and ease of error analysis, and the proliferation of high quality training data sets derived from models and/or operational measurements has further facilitated their use. In this article, we provide examples of geophysical retrieval algorithms based on neural networks with a focus on Jacobian analysis. We examine a hypothetical 80-channel hyperspectral microwave atmospheric sounder (HyMAS) and construct examples comparing neural network water vapor retrieval performance with simple regressions. Jacobians (derivatives of the outputs with respect to the network weights and with respect to the inputs) are also presented and discussed. Finally, a discussion of the Jacobian operating points is provided.

## 1. Introduction

Three-dimensional (3D) measurements of the Earth's surface and atmospheric thermodynamic state (temperature, moisture, pressure, precipitation, and so forth) have been made indirectly from satellite measurements for many years [1,2]. These measurements are inferred from direct observations of upwelling thermal emission and scattered radiance in microwave and infrared spectral regions, typically near the peaks and troughs of atmospheric absorption lines due largely to molecular oxygen, water vapor, and carbon dioxide. Physical considerations involving the use of these spectral regions include the relatively high cloud-penetrating capability at microwave wavelengths and the relatively sharp weighting functions at infrared wavelengths, particularly in the shortwave region near four micron where Planck nonlinearity further increases temperature sensitivity. A 3D characterization of the atmosphere comprises a 2D array of vertical profile measurements, sometimes referred to as "soundings", and the sensors that acquire such measurements are referred to as "sounders".

Modern spaceborne atmospheric sounders consist of passive spectrometers that measure spectral radiance intensity in a number of frequency bands. The vertical

resolution of a single sounding is a function of the number of frequency bands that are simultaneously used—bands near the peak of an absorption line measure atmospheric features near the top of the atmosphere, as lower levels are obscured due to high atmospheric absorption, and bands near the troughs of absorption lines are sensitive to the lower layers of the atmosphere and the surface. The Atmospheric InfraRed Sounder (AIRS) launched on the NASA Aqua satellite in 2002 was the first spaceborne infrared "hyperspectral" sounder, simultaneously measuring spectral radiance intensity in 2378 channels in the thermal infrared wavelength region from approximately 3.7 to 15.4 micron using a grating spectrometer [3], and the Infrared Atmospheric Sounding Interferometer (IASI), launched in 2006, measures 8461 channels from 3.6 to 15.5 micron using a Fourier transform interferometric spectrometer [4]. Infrared hyperspectral observations provide vertical resolution approaching 1 km in the lower troposphere for nadir observations and have improved Numerical Weather Prediction (NWP) forecast accuracy [5,6]. Infrared measurements, however, are significantly perturbed by the presence of clouds. Microwave measurements, at lower vertical resolution but lower sensitivity to clouds, are therefore used synergistically with infrared measurements in "cloud-clearing" algorithms to provide global, all-weather atmospheric sounding capability. The

Correspondence: wjb@ll.mit.edu  
Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA  
02420, USA

Advanced Microwave Sounding Unit (AMSU), first launched in 1998, provides measurements in 20 noncontiguous spectral bands spanning approximately 23 to 190 GHz. The spatial resolution of AIRS and AMSU varies from approximately 15 km to 150 km, depending on frequency and sensor scan angle. Additional advanced sounders operating in microwave and infrared spectral regions have recently been developed, and in 2011, the United States will launch the NPOESS preparatory project (NPP) the first satellite in its next-generation civilian operational satellite system in cooperation with the European meteorological satellite (EUMETSAT) system, a collaborative venture now termed the Joint Polar Satellite System (JPSS). The NPP satellite will host five sensors, including the advanced technology microwave sounder (ATMS) and the cross-track infrared sounder (CrIS), together referred to as the cross-track infrared and microwave sounding suite (CrIMSS). A new generation of “hyperspectral microwave” systems has recently been proposed [7], and we examine the performance of these sensors in this paper. The analyzes presented here build upon those presented in [8] by considering Jacobian analysis in the context of water vapor retrieval. Furthermore, the Jacobian operating point is also a focus of the present work.

For the following analyzes, we assume that a noisy observation of a random radiance vector  $\tilde{R}$  is related to some atmospheric state vector  $S$  through a forward model  $f(\cdot)$  as follows

$$\tilde{R} = f(S) + \Psi = R + \Psi \quad (1)$$

where  $\Psi$  is a random noise vector (that may depend on  $S$ ), and  $R$  is the “noise-free” radiance observation. The retrieval seeks to estimate the state vector  $S$  given an observation of  $\tilde{R}$ , where we use  $\hat{S}(\tilde{R})$  to denote the estimate of  $S$  given an observation of  $\tilde{R}$ .

## 2. Neural network estimation of geophysical parameters

Artificial neural networks, or neural nets, are computational structures inspired by biological networks of densely connected neurons, each of which is capable only of simple computations. Just as biological neural networks are capable of learning from their environment, neural nets are able to learn from the presentation of training data, as the free parameters (weights and biases) are adaptively tuned to fit the training data. Neural nets can be used to learn and compute functions for which the analytical relationships between inputs and outputs are unknown and/or computationally complex and are therefore useful for pattern recognition, classification, and function approximation. Neural nets are particularly appealing for the inversion of atmospheric remote

sensing data, where relationships are commonly nonlinear and non-Gaussian, and the physical processes may not be well understood. Neural networks were perhaps first applied in the atmospheric remote sensing context by Escobar-Munoz et al. [9], and many other investigators have recently reported on the use of neural networks for inversion of microwave sounding observations for the retrieval of temperature and water vapor [8,7][10-13] and hydrologic parameters [14-22], as well as inversion of infrared sounding observations for retrieval of temperature and water vapor [23-27] and trace gases [28]. Neural networks have also been used in the geophysical context for nonlinear data representation [29].

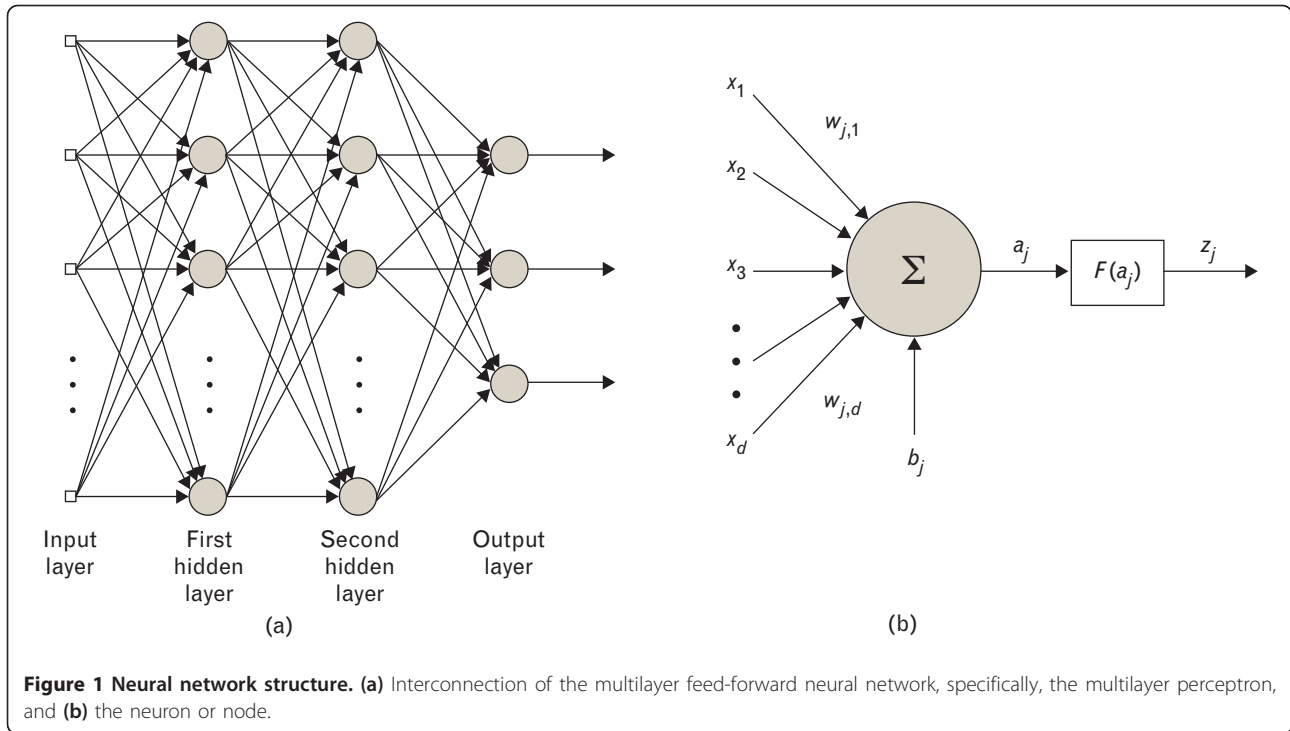
In this article, we focus on feedforward multilayer perceptron (FFMLP) neural networks due to their simplicity, flexibility, and ease of use. Multilayer neural networks most often consist of an input layer, one or more nonlinear hidden layers, and a linear output layer. The mathematical functions implemented by FFMLP nets are continuous and differentiable, which greatly simplifies training and error analysis. But perhaps the most useful attribute of neural nets is their scalability; a network with a sufficient number of weights and biases is capable of approximating a bounded, continuous function to an arbitrary level of precision over a finite domain [30]. Therefore, neural networks can be used as universal function approximators.

### 2.1. Feedforward neural networks

Feedforward neural networks propagate the inputs (the *input layer*) through a set of computational nodes arranged in layers to calculate the network outputs. The *output layer* is the final layer of the neural network and usually contains linear elements. The layers between the input layer and the output layer are called *hidden layers* and usually contain nonlinear elements. This network topology is depicted graphically in Figure 1.

The various types of feedforward neural networks differ primarily in the nonlinear functions (the so-called *activation functions*) that are used in the hidden layer nodes and the training algorithms that are used to optimize the free parameters of the network. In general, the connections shown in Figure 1 need not be fully populated: some optimization strategies start with a large number of hidden nodes and “prune” the network by eliminating connections, and possibly nodes, as training progresses.

The neuron is the basic structural element of feedforward multilayer perceptron networks. The inputs to a neuron are weighted, summed over the  $n$  inputs, translated, and passed through an activation function. The neuron is shown graphically in Figure 1, and the transfer function can be written as follows:



$$y = f \left( \sum_{i=1}^n w_i x_i + b \right) \quad (2)$$

where  $x_i$  is the  $i$ th input,  $w_i$  is the weight associated with the  $i$ th input,  $b$  is the bias,  $f(\cdot)$  is the activation function of the neuron, and  $y$  is the output. The activation functions are generally chosen to be strictly increasing, smooth (continuous first derivative), and asymptotic. Neurons with sigmoidal (soft limit) activation functions are commonly used in the hidden layer (s), and the identity function is used in the output layer. The logistic function

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

with first derivative  $f'(x) = f(x) - f^2(x)$  can be used as a sigmoidal activation function. However, a multilayer perceptron trained with the backpropagation algorithm may, in general, learn faster when the activation function is antisymmetric, that is,  $f(-x) = -f(x)$  [31]. The logistic function is not antisymmetric, but can be made antisymmetric by a simple scaling and shifting, resulting in the hyperbolic tangent function

$$f(x) = \tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (4)$$

with first derivative  $f'(x) = 1 - f^2(x)$ . The simple form of sigmoidal function and its derivative allows fast and accurate calculation of the gradients needed to optimize

selection of the weights and biases and carry out second-order error analysis.

## 2.2. Feedforward multilayer perceptron neural networks

Neurons can be combined to form a multilayer network. In this type of network, individual neurons are arranged in layers, and the neurons in each layer all use the same transfer function. The inputs to the network are fed to every node of the first layer, and the outputs of each layer (except the output layer) are fed to every node of the next layer.

An example of a two-layer network (that is, one hidden layer and one output layer) is shown in Figure 1. In Figure 1,  $x_i$  is the  $i$ th input,  $n$  is the number of inputs,  $w_{ij}$  is the weight associated with the connection from the  $i$ th input to the  $j$ th node in the hidden layer,  $b_i$  is the bias of the  $i$ th node,  $m$  is the number of nodes in the hidden layer,  $f(\cdot)$  is the transfer function of the neurons in the hidden layer,  $v_i$  is the weight between the  $i$ th node and the output node,  $c$  is the bias of the output node,  $g(\cdot)$  is the transfer function of the output node, and  $y$  is the output. We can then relate the network output to the inputs as follows:

$$y = g \left( \sum_{j=1}^m v_j f \left( \sum_{i=1}^n w_{ij} x_i + b_j \right) + c \right) \quad (5)$$

The weights ( $w_{ji}$ ) and biases ( $b_j$ ) for the  $j$ th neuron are chosen to minimize a cost function over a set of  $P$

training patterns. A common choice for the cost function is the sum-squared error, defined as

$$E(\mathbf{w}) = \frac{1}{2} \sum_p \sum_k \left( t_k^{(p)} - y_k^{(p)} \right)^2, \quad (6)$$

where  $y_k^{(p)}$  and  $t_k^{(p)}$  denote the network outputs and target responses, respectively, of each output node  $k$  given a pattern  $p$ , and  $\mathbf{w}$  is a vector containing all the weights and biases of the network. The “training” process involves iteratively finding the weights and biases that minimize the cost function through some numerical optimization procedure. Second-order methods are commonly used, where the local approximation of the cost function by a quadratic form is given by

$$E(\mathbf{w} + d\mathbf{w}) \approx E(\mathbf{w}) + \nabla E(\mathbf{w})^T d\mathbf{w} + \frac{1}{2} d\mathbf{w}^T \nabla^2 E(\mathbf{w}) d\mathbf{w}, \quad (7)$$

where  $\nabla E(\mathbf{w})$  and  $\nabla^2 E(\mathbf{w})$  are the gradient vector and the Hessian matrix of the cost function, respectively. Setting the derivative of (7) to zero and solving for the weight update vector  $d\mathbf{w}$  yields

$$d\mathbf{w} = -[\nabla^2 E(\mathbf{w})]^{-1} \nabla E(\mathbf{w}). \quad (8)$$

Direct application of (8) is difficult in practice, because computation of the Hessian matrix (and its inverse) is nontrivial, and usually needs to be repeated at each iteration. For sum-squared error cost functions, it can be shown that

$$\nabla E(\mathbf{w}) = \mathbf{J}^T \mathbf{e} \quad (9)$$

$$\nabla^2 E(\mathbf{w}) = \mathbf{J}^T \mathbf{J} + \mathbf{S}, \quad (10)$$

where  $\mathbf{J}$  is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases,  $\mathbf{e}$  is a vector of network errors, and  $\mathbf{S} = \sum_{p=1}^P \mathbf{e}_p \nabla^2 \mathbf{e}_p$  [31]. The Jacobian matrix can be computed using a standard backpropagation technique [32] that is significantly more computationally efficient than direct calculation of the Hessian matrix [33]. However, an inversion of a square matrix with dimensions equal to the total number of weights and biases in the network is required. For the Gauss-Newton method, it is assumed that  $\mathbf{S}$  is zero (a reasonable assumption only near the solution), and the update Equation (8) becomes

$$d\mathbf{w} = -[\mathbf{J}^T \mathbf{J}]^{-1} \mathbf{J} \mathbf{e}. \quad (11)$$

The Levenberg-Marquardt modification [34] to the Gauss-Newton method is

$$d\mathbf{w} = -[\mathbf{J}^T \mathbf{J} + \mu \mathbf{I}]^{-1} \mathbf{J} \mathbf{e}. \quad (12)$$

As  $\mu$  varies between zero and  $\infty$ ,  $d\mathbf{w}$  varies continuously between the Gauss-Newton step and steepest descent. The Levenberg-Marquardt method is thus an example of a model trust region approach in which the model (in this case the linearized approximation of the error function) is trusted only within some region around the current search point [35]. The size of this region is governed by the value  $\mu$ .

### 3. Neural network retrieval examples

Consider a simulated hyperspectral microwave atmospheric (HyMAS) sounding system operating near the opaque 118.75 GHz oxygen line and the 183.31GHz water vapor line [7]. The system consists of 64 temperature channels, each of approximately 500 MHz bandwidth equally spaced over a 5 GHz intermediate frequency bandwidth and 16 water vapor channels, each of approximately 1 GHz bandwidth over a 10 GHz intermediate frequency bandwidth. Channel 1 is the most transparent temperature channel (farthest from the center of the oxygen line), and channel 64 is the most opaque temperature channel (closest to the center of the oxygen line). Channel 65 is the most transparent water vapor channel and channel 80 the most opaque water vapor channel. The temperature weighting functions for this 64-channel microwave sounder are shown in Figure 2. Vertical coverage of the atmosphere is most dense in the lower atmosphere where variability is the largest and becomes sparser with increasing altitude.

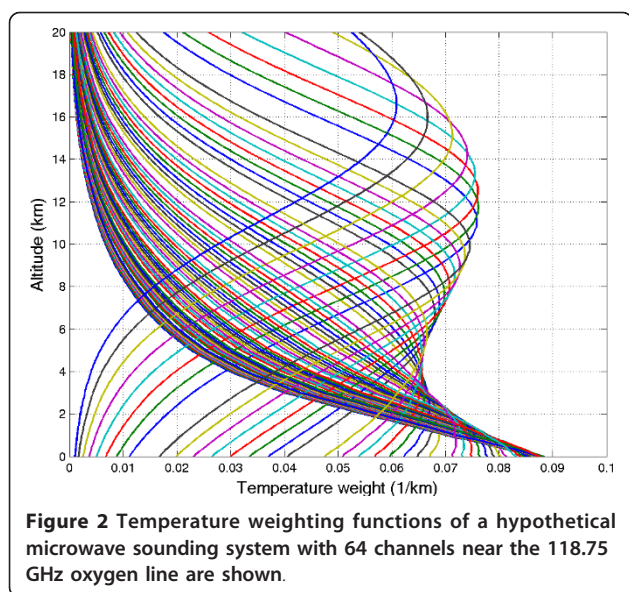
The channel sensitivities ( $\Delta T_{\text{rms}}$ ) are each approximately 0.2 K. The measurements are simulated using a radiative transfer algorithm [36] and an ocean surface model [37].

#### 3.1. Atmospheric state model

The NOAA88b global profile ensemble [38] was used to produce the training, validation, and testing sets, with an 80-10-10 split of the data. The NOAA88b data set contains 7,547 radiosonde/rocketsonde profiles, globally distributed seasonally and geographically. Atmospheric temperature, moisture, and ozone are given at 100 discrete levels from the surface to altitudes exceeding 50 km. Skin surface temperature is also recorded. NOAA88b water vapor measurements above approximately 12 km are of questionable quality and are not considered in this paper.

#### 3.2. Radiative transfer model

Simulated brightness temperature observations for atmospheric profiles in the NOAA88b data set were calculated using the TBARRAY software package of



Rosenkranz [39]. TBARRAY is a line-by-line routine based on the Liebe millimeter-wave propagation model (MPM) [40,41]. Scattering was not modeled because cloud liquid water content was not recorded in the NOAA88b data set. All radiative transfer calculations for the temperature and water vapor retrieval simulations were performed at a single angle at nadir incidence.

### 3.3. Ocean surface emissivity model: FASTEM and FASTEM2

English and Hewison developed the FASTEM model [37], which parameterizes an “effective” ocean surface emissivity for frequencies between 10 and 220 GHz for earth incidence angles less than 60° and for oceanic surface wind speeds less than 20 m/sec. FASTEM2, an updated version of FASTEM, uses an approach similar to that of Petty and Katsaros [42] to compute the surface emissivity. FASTEM and FASTEM2 both incorporate geometric optics, Bragg scattering, and foam effects. FASTEM2 (with the optical depth option set to zero) was used for the simulations of the cases presented in this paper. The oceanic surface wind speed is not recorded in the NOAA88b data set, and the FASTEM2 wind speed input for these cases was therefore randomized using a uniform distribution between 0.5 and 10 m/sec.

### 3.4. Land surface emissivity model

Land surface emissivity values were assigned randomly using a uniform distribution between 0.8-1.0. The same emissivity value was used for all frequencies. Recent study has shown that this simple model is fairly representative of most naturally occurring land emissivities

[43], although improvements are planned in future study.

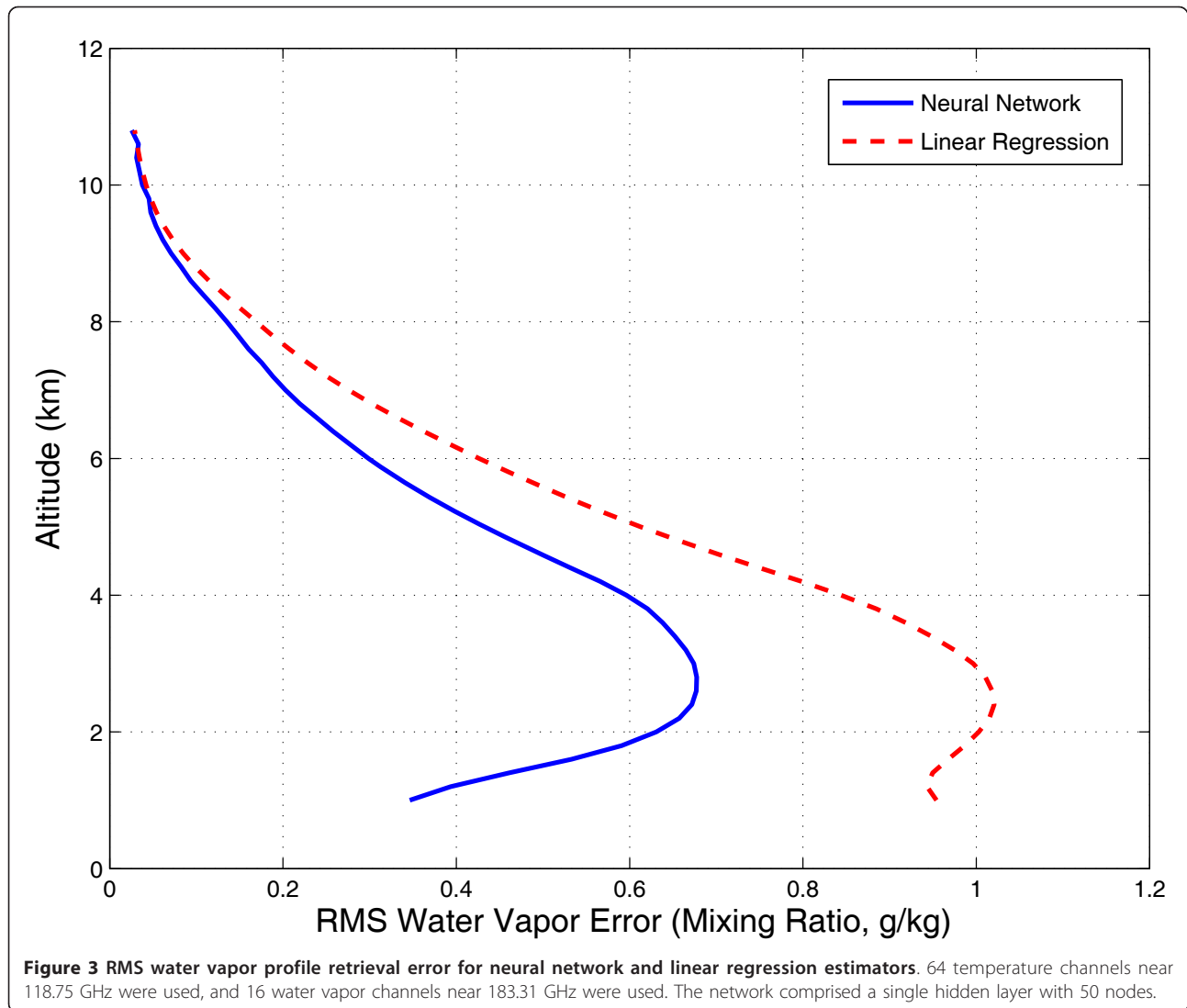
First, a linear regression water vapor (mixing ratio) profile retrieval operator was derived. The linear regression RMS error on the testing set is shown in Figure 3. A neural network with a single hidden layer of 50 nodes was initialized using the Nguyen-Widrow procedure and trained with the Levenberg-Marquardt learning algorithm. Random noise was added to the training set at each iteration, and early stopping (after approximately 50 epochs) was used to prevent overfitting by terminating the learning algorithm if the validation error failed to decrease for any of five successive epochs. The water vapor profile was estimated at 50 levels from the surface to approximately 12 km. The neural networks were each trained three times, and the validation data set was used to select the network with the best performance.

Neural network retrieval performance is shown in Figure 3. The 80-50-50 network (6,600 weights) trained at a rate of 280 seconds per epoch on a desktop Intel Xeon PC operating at a clock speed of 3 GHz. The retrieval performance of the neural network is superior to that of linear regression throughout the atmosphere. Note that the bias of the retrieval error in this example is small relative to the error variance, as the mean of the training set is very close to the mean of the validation and test sets.

## 4. Neural network Jacobians

We now further evaluate performance by examining the neural network Jacobian, that is, the sensitivity of the output values to changes in either the input values or the network weights. Jacobian analysis can be used to assess neural network performance in a variety of ways [23][44]. For example, the effect of sensor noise (or other interfering signals) on retrieval accuracy can be easily evaluated. Jacobians provide information on the relevance of network inputs and can therefore be used (usually in concert with other techniques) to select only the most significant inputs. Finally, Jacobians facilitate system optimization, where various parameters of the observing system can be optimized jointly.

A complete performance characterization of a given atmospheric retrieval algorithm requires an analysis of the retrieval errors. The retrieval error depends on several components, including sensor noise, the vertical resolution of the sensor, and many others. The assessment of these components in isolation is difficult for complex retrieval schemes, and finite-differencing methods are often used to approximate the effects of these contributions over a limited set of atmospheric cases. A significant advantage of atmospheric retrievals based on neural networks is that Jacobians can be calculated analytically, and the calculations can be carried out using



relatively simple methods, such as the backpropagation algorithm. We now present basic methods for calculating neural network Jacobians and simple examples illustrating system characterization.

A retrieval averaging kernel can be defined as:

$$\frac{\partial \hat{S}}{\partial S} = \frac{\partial R}{\partial S} \frac{\partial \hat{S}}{\partial R} \quad (13)$$

We now focus on the second term in (13), the retrieval Jacobian, which consists of the partial derivatives of the outputs (estimates of the atmospheric state vector,  $S$ ) with respect to the inputs (the radiance vector,  $R$ ). Note that the first term in (13) can readily be calculated with modern radiative transfer packages, see [45], for example. We reconcile neural network and atmospheric retrieval terminologies by equating the neural network inputs to the observed radiances,  $X = R$ , and equating

the neural network outputs to the estimates of the atmospheric states,  $Y = \hat{S}$ . Returning to the equation relating the inputs and outputs of a simple feedforward multilayer perceptron with one hidden layer of  $m$  nodes and a single output,  $y_k$ :

$$y_k = g \left( \sum_{j=1}^m v_{jk} f \left( \sum_{i=1}^n w_{ij} x_i + b_j \right) + c_k \right) \quad (14)$$

we can express the derivative  $dy_k/dx_i$  using the chain rule, as follows:

$$\frac{dy_k}{dx_i} = g'(a_k) \sum_{j=1}^m v_{jk} f'(a_j) w_{ij} \quad (15)$$

where  $a_k$  and  $a_j$  are the weighted sum of the inputs to the output and hidden layers, respectively. We assume

that all inputs other than  $x_i$  are fixed. The derivative of the hyperbolic tangent function is related to the function itself as

$$f' = 1 - f^2 \quad (16)$$

If a linear output layer is used,  $g'(ak) = 1$ , and (15) becomes

$$\frac{dy_k}{dx_i} = \sum_{j=1}^m v_{jk}(1 - f^2(a_j))w_{ij} \quad (17)$$

and we see that the Jacobian is easily calculated from the network outputs and weights. This result for a network with a single hidden layer is readily generalized to networks with multiple hidden layers. Note that the mapping function implemented by the neural network could be highly nonlinear, and therefore care must be taken to ensure that the Jacobian is evaluated near an appropriate operating point. The network Jacobian is generated as a simple “byproduct” of the forward propagation of the inputs through the network. Because of this, neural networks are well suited to complicated function approximation problems, as the computation required for error analysis is greatly reduced in comparison to other methods requiring numerical finite-difference techniques.

## 5. Neural network error analysis using the Jacobian

We now present a typical case study to illustrate several facets of retrieval system analysis using the Jacobian. We return to the neural network example presented earlier, which is based on a simulated spaceborne microwave sounding system with 80 channels operating near the 118.75 GHz oxygen line and the 183.31GHz water vapor line. The neural network is used to retrieve the water vapor (mixing ratio) profile at 50 levels. The NOAA88b global ensemble [38] of over 7,000 profiles was used to produce the training, validation, and testing sets, with an 80-10-10 split of the data. An FFMLP network with a single hidden layer of 30 nodes was initialized using the Nguyen-Widrow procedure and trained with the Levenberg-Marquardt learning algorithm. Random noise ( $\sigma = 0.2$  K) was added to the training set at each iteration, and early stopping was used to prevent overfitting. Both the simulated radiances and the temperature profile elements were normalized to unit standard deviation and zero mean to simplify the interpretation of the resulting Jacobians.

### 5.1. The network weight Jacobian

The Jacobians (both with respect to the network weights and inputs) can now be calculated using the trained

neural network as discussed in Section 4. First, we will examine the “network weight Jacobian,” the derivative of the outputs with respect to the weights. This derivative is the primary component of network learning algorithms based on gradient descent. Large values of the Jacobian indicate that a given weight has high influence on the output parameter (that is, the output is highly sensitive to the value of that weight) and small values indicate a small influence. The network weight Jacobian for the microwave sounder example is shown in Figure 4. Only the 4,000 (80 inputs  $\times$  50 nodes) hidden layer weights are shown in the figure. The weights have been arranged in increasing order of the derivative at 10 km. There are several interesting features evident in this figure. First, a band of approximately 1,000 weights (near the vertical center in the figure) have very small influence on any output. This suggests that the neural network could be simplified by removing the connections corresponding to the least influential weights. Second, outputs corresponding to altitudes in the mid-troposphere are highly influenced (both in the positive and negative directions) by some of the weights. This could indicate that the relationships between the radiometric observations and the temperature at these altitudes are substantially nonlinear.

### 5.2. The network input Jacobian

We now consider the “network input Jacobian,” the derivative of the outputs with respect to the inputs. This Jacobian reveals the inputs that most significantly influence the outputs. The influence of input perturbations on the outputs can be calculated as follows [35]:

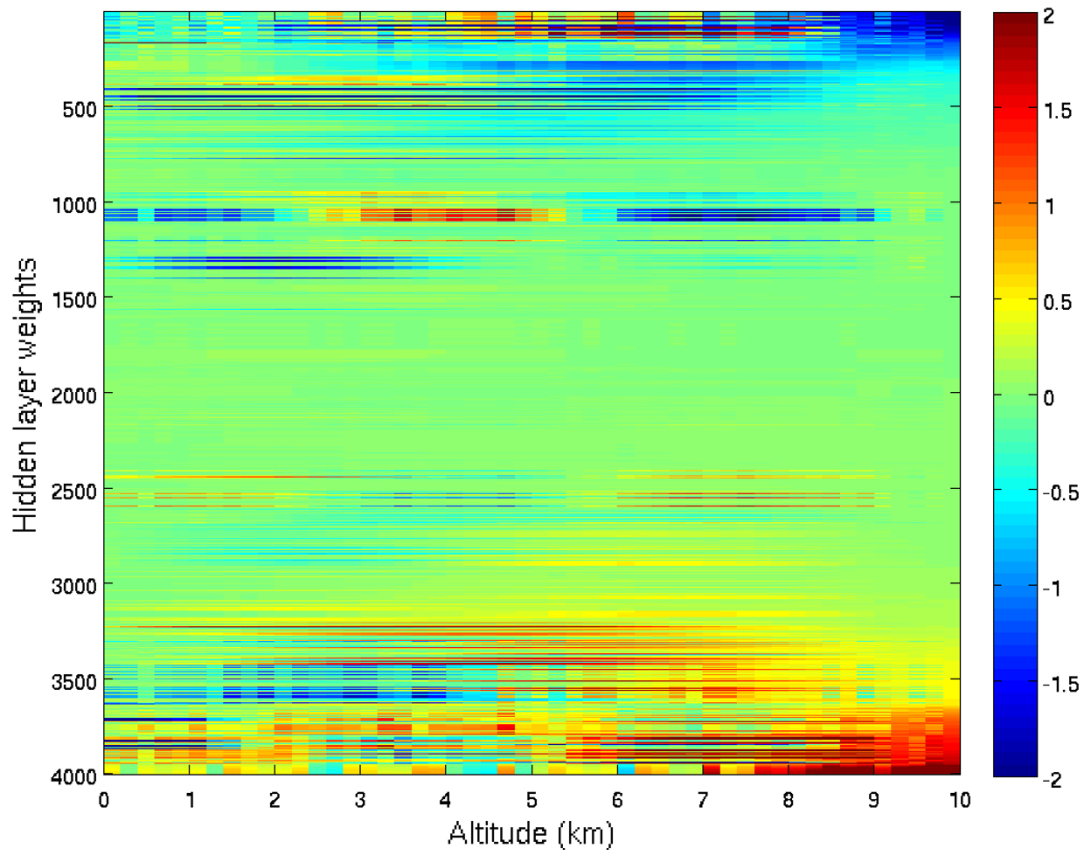
$$\nabla y_k \simeq \sum_i \frac{dy_k}{dx_i} \nabla x_i \quad (18)$$

This simple formalism allows many attributes of a remote sensing system to be evaluated by propagating the perturbations through to the outputs using the Jacobian. Furthermore, second-order analysis is facilitated by expressing the output covariance matrix as a function of the input covariance (or the covariance of the input perturbation) and the Jacobian as follows:

$$C_{\Delta Y} \approx \frac{\partial Y}{\partial X} C_{\Delta X} \left( \frac{\partial Y}{\partial X} \right)^T \quad (19)$$

Second-order analysis of this type is a very powerful tool both for diagnostic and optimization purposes.

The neural network input Jacobian for the 80-channel microwave example is shown in Figure 5a for the mean brightness temperature spectrum (the implications of other operating points will be explored in the following section). It is evident from the figure that higher



**Figure 4** The derivative of the neural network output with respect to the network weights in the hidden layer. The network outputs consist of the water vapor profile estimates from 0 to 10 km in 200 m steps.

channel numbers generally exhibit more influence on outputs corresponding to higher altitudes, as is expected, because the channels are ordered in increasing opacity. Furthermore, outputs corresponding to higher altitudes generally are more influenced by the inputs than are the low-altitude outputs.

For comparison, the analogous Jacobian image for a linear regression operator is shown in Figure 5b. There are general features common to both images. However, the smooth gradients in the linear regression Jacobian image are immediately apparent, in contrast to the high-frequency structure in the corresponding neural network Jacobian image. This structure could be evidence of nonlinear or non-Gaussian relationships between the inputs and the outputs that the neural network is exploiting.

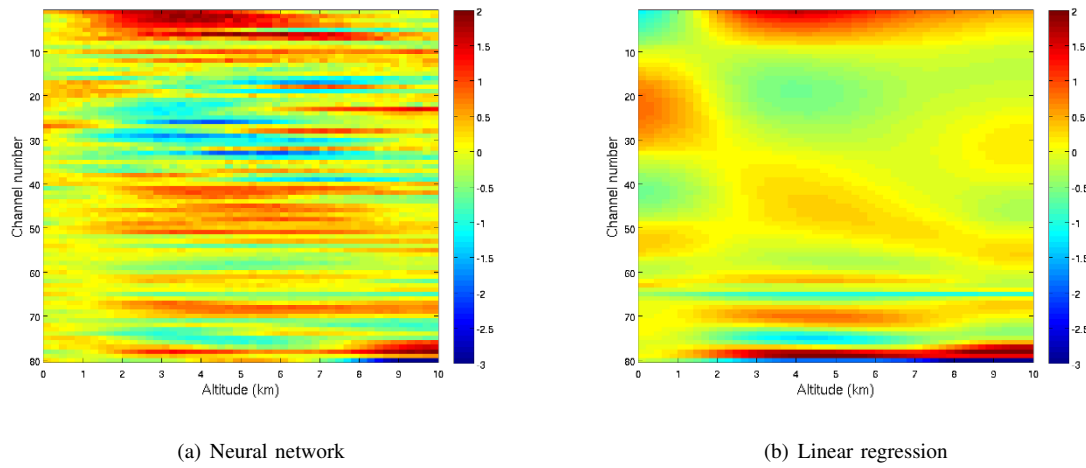
### 6. Impact of Jacobian operating point

We conclude the article with a brief analysis of the sensitivity of the Jacobian to the input operating point, that is, the point at which the tangent line is calculated ( $a_j$  in (17)). The neural network in general synthesizes a

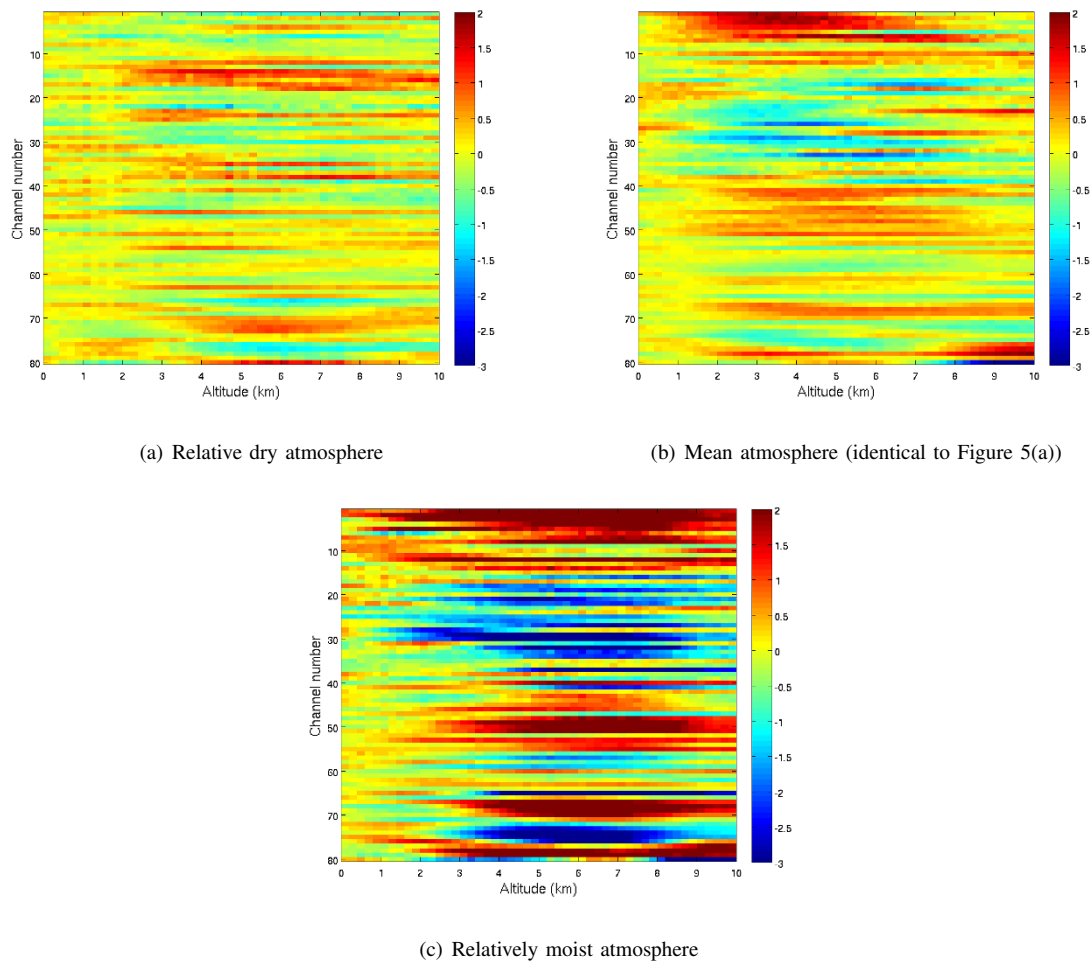
nonlinear function, and the Jacobian therefore may depend significantly on the operating point. Note in contrast that the linear regression by definition yields a linear Jacobian for all inputs.

An example is presented in Figure 6. Here, the network input Jacobian is calculated for three different cases of the water vapor retrieval, a relatively dry case with approximately 1 mm of integrated water vapor (IWV) is shown in the top panel of the figure, the mean of the training case with approximately 8 mm IWV is shown in the middle panel of the figure, and a relatively moist case with approximately 25 mm IWV is shown in the bottom panel of the figure. It is clear that the Jacobian magnitude generally increases with increasing water content (put another way, the nonlinear relationship between brightness temperature and water vapor content increases with water vapor content, a principal reason that the neural network is substantially superior to the linear regression retrieval). However, it is interesting to note that some features are strikingly different in the three images. For example, in the dry atmosphere case in the mid-troposphere, the Jacobian has a negative





**Figure 5** The derivative of the retrieval output with respect to the inputs for: (a) neural network and (b) linear regression. The outputs consist of the water vapor profile estimates from 0 to 10 km in 200 m steps.



**Figure 6** The derivative of the retrieval output with respect to the inputs for: (a) a relatively dry atmosphere (integrated water vapor, or IWV, of approximately 1 mm), (b) the mean NOAA88b atmosphere (IWV of approximately 8 mm), and (c) a relatively moist atmosphere (IWV of approximately 25 mm). The outputs consist of the water vapor profile estimates from 0 to 10 km in 200 m steps.

slope for channels 75-79, but a positive slope neighboring channels (near 70-74). The moist atmospheric case shows the opposite behavior, as a significant negative slope is seen in channels 70-74.

The implications for this result is that it is important to consider the input operating point when working with Jacobians for highly nonlinear processes such as the water vapor retrieval. Furthermore, examination of the Jacobians can be used to help assess both the nonlinearity of the underlying retrieval problem and the degree to which the nonlinearity changes with respect to the inputs.

## 7. Summary

Neural network retrievals of atmospheric water vapor were shown to substantially outperform linear regression retrievals for an 80-channel hyperspectral microwave sounding system based on a global simulation using the NOAA88b profile dataset. It was demonstrated that neural network Jacobian analysis is a powerful tool that can be used to assess a variety of performance metrics. The network Jacobians are easily calculated using analytical expressions related to network outputs, and therefore very little additional computation is required. The Jacobians can be used to determine the sensitivity of the network outputs to changes in both the network weights and the inputs. This information can provide insight into network topology optimization by identifying connections that do not significantly influence the outputs. Jacobians are also useful for perturbation analysis, where input perturbations due to sensor noise or other interfering signals can be easily propagated through the neural network, and the resulting impact on the outputs can be determined. Finally, it was shown that the Jacobians may depend significantly on the input for problems that are highly nonlinear, and care must be taken therefore to choose a suitable operating point based on the objectives of the analysis.

## Acknowledgements

The author gratefully acknowledges the contributions of many colleagues and collaborators, including: Dave Staelin, Joel Susskind, Fred Chen, Laura Bickmeier, Vince Leslie, Laura Jairam, Michael Pieper, Chuck Cho, Chinnawat Surussavadee, Jenna Samra, Christy Cull, and Adam Milstein. The comments and suggestions provided by the three anonymous reviewers substantially improved the final manuscript. This study was funded in part by NASA contract NNH08AH881. This study was supported by the National Aeronautics and Space Administration under Air Force contract FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the author and were not necessarily endorsed by the United States Government.

## Competing interests

The authors declare that they have no competing interests.

Received: 25 July 2011 Accepted: 23 March 2012

Published: 23 March 2012

## References

1. DH Staelin, AH Barrett, JW Waters, FT Barath, EJ Johnston, PW Rosenkranz, NE Gaut, WB Lenoir, Microwave spectrometer on the Nimbus 5 satellite: meteorological and geophysical data. *Science*. **182**, 1339–1341 (1973). doi:10.1126/science.182.4119.1339
2. WL Smith, Satellite techniques for observing the temperature structure of the atmosphere. *Bull Am Meteorol Soc*. **53**(11), 1074–1082 (1972). doi:10.1175/1520-0477(1972)053<1074::CO>2
3. HH Aumann, et al, AIRS/AMSU/HSB on the Aqua mission: design, science objectives, data products, and processing systems. *IEEE Trans Geosci Remote Sens*. **41**(2), 253–264 (2003). doi:10.1109/TGRS.2002.808356
4. G Chalon, F Cayla, D Diebel, IAS: An advanced sounder for operational meteorology. in *Proceedings of the 52nd Congress of IAF*, Toulouse, France, 1–5 (2001)
5. H Liu, J Li, An improvement in forecasting rapid intensification of Typhoon Sinlaku (2008) using clear-sky full spatial resolution advanced IR soundings. *J Clim Appl Meteorol*. **49**(4), 821–827 (2010). doi:10.1175/2009JAMC2374.1
6. T Pangaud, N Fourrie, V Guidard, M Dahoui, F Rabier, Assimilation of AIRS radiances affected by mid-to low-level clouds. *Mon Weather Rev*. **137**(12), 4276–4292 (2009). doi:10.1175/2009MWR3020.1
7. W Blackwell, L Bickmeier, R Leslie, M Pieper, J Samra, C Surussavadee, C Upham, Hyperspectral microwave atmospheric sounding. *IEEE Trans Geosci Remote Sens*. **49**(1), 128–142 (2011)
8. W Blackwell, FW Chen, *Neural Networks in Atmospheric Remote Sensing*, (Artech House Inc., Norwood, MA, 2009)
9. J Escobar-Munoz, A Chédin, F Chéruy, N Scott, Réseaux de neurones multicouches pour la restitution de variables thermodynamiques atmosphériques à l'aide de sondeurs verticaux satellitaires. *Comptes-Rendus de L'Academie Des Sciences; Série II*. **317**(7), 911–918 (1993)
10. WJ Blackwell, CH Chen, *Signal and Image Processing for Remote Sensing*, 2nd edn. CRC Press, Boca Raton, FL, 161–190 (2012). Chapter. 11, Hyperspectral microwave atmospheric sounding using neural networks
11. L Shi, Retrieval of atmospheric temperature profiles from AMSU-A measurement using a neural network approach. *J Atmos Oceanic Technol*. **18**, 340–347 (2001). doi:10.1175/1520-0426(2001)018<0340::CO>2
12. F Del Frate, G Schiavon, A combined natural orthogonal functions/neural network technique for the radiometric estimation of atmospheric profiles. *Radio Sci*. **33**(2), 405–410 (1998). doi:10.1029/97RS02219
13. F Del Frate, G Schiavon, Nonlinear principal component analysis for the radiometric inversion of atmospheric profiles by using neural networks. *IEEE Trans Geosci Remote Sens*. **37**(5), 2335–2342 (1999). doi:10.1109/36.789630
14. D Staelin, F Chen, Precipitation observations near 54 and 183 GHz using the NOAA-15 satellite. *IEEE Trans Geosci and Remote Sens*. **38**(5), 2322–2332 (2000). doi:10.1109/36.868889
15. F Chen, D Staelin, AIRS/AMSU/HSB precipitation estimates. *IEEE Trans Geosci Remote Sens*. **41**(2), 410–417 (2003). doi:10.1109/TGRS.2002.808322
16. C Surussavadee, DH Staelin, Millimeter-wave precipitation retrievals and observed-versus-simulated radiance distributions: Sensitivity to assumptions. *J Atmos Sci*. **64**(11), 3808–3826 (2007). doi:10.1175/2006JAS2045.1
17. C Surussavadee, DH Staelin, Global millimeter-wave precipitation retrievals trained with a cloud-resolving numerical weather prediction model, Part I: Retrieval design. *IEEE Trans Geosci Remote Sens*. **46**(1), 99–108 (2008)
18. C Surussavadee, DH Staelin, Global millimeter-wave precipitation retrievals trained with a cloud-resolving numerical weather prediction model, Part II: Performance evaluation. *IEEE Trans Geosci Remote Sens*. **46**(1), 109–118 (2008)
19. C Surussavadee, DH Staelin, Satellite retrievals of arctic and equatorial rain and snowfall rates using millimeter wavelengths. *IEEE Trans Geosci Remote Sens*. **47**(11), 3697–3707 (2009)
20. C Surussavadee, DH Staelin, Global precipitation retrievals using the NOAA/AMSU millimeter-wave channels: comparison with rain gauges. *J Clim Appl Meteorol*. **49**(1), 124–135 (2010). doi:10.1175/2009JAMC2262.1
21. C Surussavadee, DH Staelin, NPOESS precipitation retrievals using the ATMS passive microwave spectrometer. *IEEE Geosci Remote Sens Lett*. **7**(3), 440–444 (2010)
22. C Surussavadee, D Staelin, Global precipitation retrieval algorithm trained for SSMIS using a numerical weather prediction model: Design and

- evaluation. in *IEEE International Geoscience and Remote Sensing Symposium Proceedings*, Honolulu, HI, 2341–2344 (2010)
23. HE Motteler, LL Strow, L McMillin, JA Gualtieri, Comparison of neural networks and regression based methods for temperature retrievals. *Appl Opt.* **34**(24), 5390–5397 (1995). doi:10.1364/AO.34.005390
  24. WJ Blackwell, A neural-network technique for the retrieval of atmospheric temperature and moisture profiles from high spectral resolution sounding data. *IEEE Trans Geosci Remote Sens.* **43**(11), 2535–2546 (2005)
  25. WJ Blackwell, CH Chen, *Signal and Image Processing for Remote Sensing*, 1st edn. CRC Press, Boca Raton, FL, (2006). chap. 11., Neural network retrievals of atmospheric temperature and moisture profiles from high-resolution infrared and microwave sounding data
  26. F Aires, A Che'din, NA Scott, WB Rossow, A regularized neural net approach for retrieval of atmospheric and surface with the IASI instrument. *J Appl Meteorol.* **41**, 144–159 (2002). doi:10.1175/1520-0450(2002)0412.0.CO;2
  27. F Aires, WB Rossow, NA Scott, A Chédin, Remote sensing from the infrared atmospheric sounding interferometer instrument: 2. simultaneous retrieval of temperature, water vapor, and ozone atmospheric profiles. *J Geophys Res.* **107**(D22), 4620 (2002). doi:10.1029/2001JD001591
  28. F Del Frate, A Ortenzi, S Casadio, C Zehner, Application of neural algorithms for a real-time estimation of ozone profiles from gome measurements. *IEEE Trans Geosci Remote Sens.* **40**(10), 2263–2270 (2002). doi:10.1109/TGRS.2002.803622
  29. F Chen, Neural network characterization of geophysical processes with circular dependencies. *IEEE Trans Geosci Remote Sens.* **45**(10), 3037–3043 (2007)
  30. KM Hornik, M Stinchcombe, H White, Multilayer feedforward networks are universal approximators. *Neural Netw.* **4**(5), 359–366 (1989)
  31. S Haykin, *Neural Networks: A Comprehensive Foundation*, Macmillan College Publishing Company, New York, (1994)
  32. DE Rumelhart, G Hinton, R Williams, DE Rumelhart, JL McClelland, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations*, MIT Press, Cambridge, MA, (1986)
  33. MT Hagan, MB Menhaj, Training feedforward networks with the Marquardt algorithm. *IEEE Trans Neural Netw.* **5**, 989–993 (1994). doi:10.1109/72.329697
  34. PE Gill, W Murray, MH Wright, The Levenberg-Marquardt method. in *Practical Optimization*, (Academic Press, London, 1981)
  35. CM Bishop, *Neural Networks for Pattern Recognition*, (Oxford University Press, Oxford, 1995)
  36. PW Rosenkranz, Radiative transfer solution using initial values in a scattering and absorbing atmosphere with surface reflection. *IEEE Trans Geosci Remote Sens.* **40**(8), 1889–1892 (2002). doi:10.1109/TGRS.2002.802518
  37. S English, T Hewison, A fast generic millimeter-wave emissivity model. *SPIE Proc.* **3503**, 288–300 (1998)
  38. E Borbas, S Seemann, HL Huang, J Li, WP Menzel, Global profile training database for satellite regression retrievals with estimates of skin temperature and emissivity. in *Proc Int ATOVS Study Conf*, Beijing, China, 763–770 (2005)
  39. PW Rosenkranz, A Janssen, *Atmospheric Remote Sensing by Microwave Radiometry*, Wiley, New York, 37–90 (1993). Chapter 2, Absorption of microwaves by atmospheric gases
  40. HJ Liebe, MPM: an atmospheric millimeter-wave propagation model. *Int J Infrared Millimeter Waves.* **10**(6), 631–650 (1989). doi:10.1007/BF01009565
  41. HJ Liebe, PW Rosenkranz, GA Hufford, Atmospheric 60-GHz oxygen spectrum: New laboratory measurements and line parameters. *J Quant Spectrosc Radiat Transfer.* **48**(5-6), 629–643 (1992). doi:10.1016/0022-4073(92)90127-P
  42. GW Petty, KB Katsaros, The response of the SSM/I to the marine environment. Part II: A parameterization of the effect of the sea surface slope distribution on emission and reflection. *J Atmos Oceanic Technol.* **11**, 617–628 (1994). doi:10.1175/1520-0426(1994)0112.0.CO;2
  43. F Karbou, C Prigent, L Eymard, JR Pardo, Microwave land emissivity calculations using AMSU measurements. *IEEE Trans Geosci Remote Sens.* **43**, 948–959 (2005)
  44. F Aires, C Prigent, WB Rossow, Neural network uncertainty assessment using Bayesian statistics with application to remote sensing: 2. Output errors. *J Geophys Res.* **109** (2004)
  45. AG Voronovich, AJ Gasiewski, BL Weber, A fast multistream scattering-based Jacobian for microwave radiance assimilation. *IEEE Trans Geosci Remote Sens.* **42**(8), 1749–1761 (2004)

doi:10.1186/1687-6180-2012-71

Cite this article as: Blackwell: Neural network Jacobian analysis for high-resolution profiling of the atmosphere. *EURASIP Journal on Advances in Signal Processing* 2012 **2012**:71.

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---