Simultaneous Retrieval of Temperature, Water Vapor and Ozone Atmospheric Profiles from IASI: Compression, De-noising, First Guess Retrieval and Inversion Algorithms

F. Aires
Department of Applied Physics, Columbia University, NASA Goddard Institute for Space Studies, New York

W.B. Rossow
NASA Goddard Institute for Space Studies, New York

N.A. Scott, and A. Chédin
Laboratoire de Météorology Dynamique, École Polytechnique, France

Short title: ATMOSPHERIC RETRIEVALS FROM IASI USING A NEURAL NETWORK

Abstract. A fast temperature, water vapor and ozone atmospheric profile retrieval algorithm is developed for the high spectral resolution Infrared Atmospheric Sounding Interferometer (IASI) space-borne instrument. Compression and de-noising of IASI observations are performed using Principal Component Analysis. This preprocessing methodology also allows for a fast pattern recognition in a climatological data set to obtain a first guess. Then, a neural network using first guess information is developed to retrieve simultaneously temperature, water vapor and ozone atmospheric profiles. The performance of the resulting fast and accurate inverse model is evaluated with a large diversified data set of radiosondes atmospheres including rare events.

1. Introduction

The Infrared Atmospheric Sounding Interferometer (IASI), is a high resolution ($0.25\ cm^{-1}$) Fourier transform spectrometer scheduled for flight in 2005 on the European polar METeorological Operational Platform (METEOP-1) satellite funded by the EUropean organization for METeorological SATellites (EUMETSAT) and the European Space Agency (ESA) member states. This instrument is intended to replace the High Resolution Infrared Radiation Sounder (HIRS) as the operational infrared sounder and is expected to reach accuracies of 1 K in temperature and 10 % in water vapor with vertical resolutions of 1 km and 2 km respectively (cloud-free). IASI, jointly developed by the Centre National d'Études Spatiales (CNES) and EUMETSAT, provides spectral channels from 3.5 μm to 15.5 μm at considerably higher spectral resolution than HIRS and, together with the Advanced Microwave Sounding Unit (AMSU), is expected to lead to dramatic improvements in the accuracy and height resolution of remotely sensed temperature and humidity profiles and ozone amount.

The dimension (number of measurements per field-of-view) of the IASI observations is much higher than for previous instruments: 8461 channels compared to 27 for the TIROS-N Operational Vertical Sounding (TOVS) instrument. This is a major problem in the definition of retrieval algorithms. Classical approaches are often unable to deal with this amount of information. Iterative methods require a fast direct model with precise Jacobians (i.e., first derivative of observation with respect to variables to retrieve), but such a model is not available yet. Variational assimilation techniques also need a fast forward model with the Jacobians or the tangent linear operator: this approach is also unable to use the full IASI information because of the dimension of observations. To deal with this high-dimension problem, various techniques to select channel in the IASI spectrum have been developed in [*Rabier et al.*, 2000] or in [*Aires et al.*, 2000], A regularized neural net approach for retrieval of atmospheric and surface temperatures with the IASI instrument, submitted to *J. of Applied Meteorology*, 2000).

As we will see in the following, such an approach is not optimal because information is necessarily lost.

In this study, we present an inversion scheme for retrieving geophysical variables from IASI measurements. The retrieval technique should be able to deal with realistic conditions: noise in the measurements, nonlinearity of the function, non-Gaussianity of the variables involved, multicollinearities between variables, dependence of first guess errors on situation, uncertainties in the radiative transfer model, etc. Neural network techniques, and in particular the Multi-Layer Perceptron (MLP) technique, have already proved very successful in the development of computationally efficient inversion methods for satellite data and for geophysical applications [*Escobar et al.*, 1993; *Aires et al.*, 1998; *Chaboureau et al.*, 1998; *Chevallier et al.*, 1998; *Krasnopolsky et al.*, 2000; *Aires et al.*, 2001]. They are well adapted to solve nonlinear problems and are especially designed to capitalize on the inherent statistical relationships among the retrieved parameters. No assumptions are made concerning the probability distribution functions of the variables involved in the problem, so the method is able to deal with non-Gaussian distributions, which is not the case with classical inversion techniques. Furthermore, the neural network inversion method is a model of the inverse radiative transfer function in the atmosphere parameterized once and for all, where classical methods use the inversion technique for each observation.

However, for ill-conditioned problems, the use of a first guess estimate and associated error covariance matrix is essential for elaborated stand-alone retrieval schemes [*Chédin et al.*, 1985] as well as for three-dimensional/four-dimensional variational assimilation schemes since it controls the impact of the measurements on the retrieved parameters [*Thépaut et al.*, 1993]. A neural network techniques has recently been developed [*Aires et al.*, 2001] to use such a priori information (i.e., a specific state-dependent first guess estimate). This has been a major improvement of the classical neural network methods for remote sensing in particular, and for inverse problems in general.

Other great advantages of the MLP are its rapidity, small amount of memory required and accuracy of results. Fusion of information from different instruments coupled to the nonlinear habilities of the neural network model [*Prigent et al.*, 2001], can exploit more fully the relationships among the observations and among the variables that are described implicitly in the training data set. Variational techniques have to specify the covariance matrices explicitly, which is not a simple task since these matrices are dependent on atmospheric situation, latitude, etc.

We present here an application of a new neural network method to the retrieval of atmospheric temperature, water vapor and ozone profiles retrieval from IASI observations. Previous studies have used information content analysis to estimate the expected retrieval errors of IASI [*Amato and Serio*, 1997; *Prunet et al.*, 1998]; but these estimates are dependent on some assumptions (Gaussian hypothesis, independence between first guess and observation, first guess error covariance matrices often taken to be diagonal, i.e. no correlations between the first guess errors of the variables, etc), and have been applied to only for a limited number of atmospheric situations. Where our neural network model is parameterized and tested over a large number of real atmospheric situations as measured by radiosondes, taken from the Thermodynamic Initial Guess Retrieval (TIGR) database [*Chédin et al.*, 1985; *Achard*, 1991; *Escobar*, 1993b; *Chevallier et al.*, 1998].

This paper is organized as follows: the description of the IASI instrument is presented in section 2. Section 3 describes the compression and the de-noising techniques based on Principal Component Analysis (PCA) for IASI spectra. The retrieval algorithm based on a first guess-based neural network approach is presented in section 4. Temperature, water vapor and ozone atmospheric profiles retrieval results are presented in section 5. Section 6 concludes this study with some perspectives on this work.

2. IASI Instrument

2.1. Characteristics of IASI

The two major advances of the IASI instrument are: (1) the dramatically increased number of spectral channels: for each field of view, 8461 measures are available covering the spectral range from 645 to 2760 cm^{-1} with a resolution (unapodized) of 0.25 cm^{-1}, with hundreds of them sounding the atmospheric temperature. The retrieval becomes an over-constrained problem. (2) The increased resolution power: with IASI the resolution power is two order of magnitude higher than with such instrument as TOVS HIRS radiometer. So, it is expected that the vertical resolution and the accuracy of retrievals will substantially increase: the IASI mission specifications are a mean error of 1 K in atmospheric temperature and 10 % in relative humidity profiles with respectively, 1 Km and 2 Km vertical resolution. Table 1 represents some of the most important spectral regions and their associated absorbing constituent.

The IASI noise is presently simulated by a white Gaussian noise (this is a realistic assumption for an interferometer) with a $NE\Delta T$ at 280 K given in Table 2 ([*Cayla et al.*, 1995], and for more recent results, *Cayla et al.*, personal communication). The $NE\Delta T$ at 280 K represents the standard deviation $st_{280}(\nu)$ of the Gaussian noise for a given wave number ν. At any other scene brightness temperature, T', the standard deviation, $st_{T'}(\nu)$, of the Gaussian noise is computed by:

$$st_{T'}(\nu) = \frac{\frac{\partial B(Tb=280,\nu)}{\partial Tb}}{\frac{\partial B(Tb=T')}{\partial Tb}} \cdot st_{280}(\nu) \qquad (1)$$

which shows that the noise level increases when T' decreases. Figure 1 shows the IASI spectrum averaged over the whole TIGR data set (a data set of climatological situations that will be described in section 5.1) with the corresponding noise standard deviation spectrum. This figure shows that some spectral regions could have a noise standard deviation higher than 2 K for a standard atmospheric situation.

2.2. Dimension Reduction

Using *directly* (i.e. without pre-processing) all 8461 IASI channels in a retrieval algorithm is a simplistic strategy that would give poor results for practical and theoretical reasons. High-dimension data have to be reduced to limit the *curse of dimensionality* [Bishop, 1996]. The curse of dimensionality stipulates that, as the dimension M of the data space increases, the difficulty of the statistical regression procedure, consisting here to infer the Radiative Transfer Model (RTM), increases significantly and the number, E, of examples required to the regression increases exponentially with the dimension M. The curse of dimensionality, however, may remain tractable because the *intrinsic* complexity of the function to be estimated, which is really the factor controling the number of examples required, does not increase exponentially with the dimension.

However, practical problems occur. For example, the number of parameters in the regression model increases with M. This excess degrees of freedom in the regression, combined with the introduction of non-informative data (i.e., variability not linked to the desired output, like the instrumental noise or an inadequate vertical resolution), may perturb the regression process: the quality criterion used to parameterized the inverse model becomes more complex so the global minimum is harder to estimate. Furthermore, computations are longer with such a large number of parameters.

Thus, the goal of dimension reduction is to present to the regression model the most relevant information from the initial rough data. One way to reduce dimension is feature selection: only a part of the observation is selected for the regression [*Jain et al.*, 1997]. An example of such an approach is channel selection schemes [*Rodgers*, 1990]. For example Jacobian-based channel selection algorithms use the Jacobians of the RTM to investigate the information content of the instrument channels in order to select the more pertinent ones [*Rabier et al.*, 2000] and [*Aires et al.*, 2000]. But this approach does not allow a full use of all the information provided by IASI (i.e., loss of information) and it also limits the use of channel redundancy for noise reduction. This is a main

drawback for IASI since noise can be important in some spectral region, especially in the third spectral band of IASI (Figure 1). Another way to reduce the dimension of the data is by feature extraction, i.e. and operator acts on the entire observed IASI spectrum to extract its more pertinent characteristics. PCA is often used for that purpose: the dimension reduction is obtained by combining mutual information among measured brightness temperatures. As explained in next section, a compromise needs to be made between reducing data dimension and preserving the redundant information in the rough data.

3. Principal Component Analysis of IASI Spectra

Although widely used for statistical analysis, the PCA technique is also very efficient for compression purposes [*Joliffe*, 1986]. It is used here to compress and de-noise the IASI observations. In the following, all IASI spectra are the result of a RTM computation since IASI does not exist yet.

3.1. Principal Component Analysis

Let $\mathcal{D} = \{x^e \; ; \;\; e = 1, \ldots, E\}$ be a database of E spectra, x, of dimension $M = 8461$. Let Σ be the $M \times M$ covariance matrix of the $\mathcal{D}$ database. Let V be the $M \times M$ matrix with columns equal to the eigenvectors of Σ and let L be the diagonal $M \times M$ matrix with the M associated eigenvalues in decreasing order (by definition $\Sigma \cdot V = V \cdot L$).

We define the $M \times M$ filter matrix $F = L^{-1/2} \cdot V^t$. The matrix F is used to project IASI spectra, x, onto a new orthonormal base composed by the columns of F: $\{F_{\star i} \; ; \;\; i = 1, \ldots, M\}$:

$$\begin{cases} h = F \cdot x = F_{1\star} \cdot x_1 + \ldots + F_{M\star} \cdot x_M \\ x = F^{-1} \cdot h = F^t \cdot h = h_1 \cdot F_{\star 1} + \ldots + h_M \cdot F_{\star M} \end{cases} \tag{2}$$

where ^{t} is the transpose operator. The vectors $\{F_{i\star} \; ; \;\; i = 1, \ldots, M\}$ are called the filters and the normalized eigenvectors $\{F_{\star i} \; ; \;\; i = 1, \ldots, M\}$ are called the PCA base

functions. Because these eigenvectors are an orthogonal basis to represent IASI spectra, x, we will refer to them as eigen-spectra. By definition, the coordinates of the new data h are uncorrelated since:

$$< h \cdot h^t >=< F \cdot x \cdot x^t \cdot F^t >=< F \cdot \Sigma \cdot F^t >= I_{M \times M}, \quad (3)$$

where $< \cdot >$ represents the mathematical expectation.

Practically, the first step in a PCA approach is to compute the 8461 $\times$ 8461 covariance matrix $\Sigma =< (x- < x >) \cdot (x- < x >)^t >$ of the database, where x is the IASI spectrum composed of the 8461 channels. The eigenvalue matrix L and the corresponding eigenvectors V of this covariance matrix Σ are then computed using a Cholesky or a SVD decomposition.

3.2. Analysis of the Eigen-Spectrum base functions

In Figure 2, the cumulated percentage of explained variance is represented as a function of the number of components. We see that the 99 % level is attained with only 10 components. PCA uses optimally the redundant information existing in the IASI channels by adaptatively determining the principal components h_i as a weighted sum of partially redundant channels: $h_i = \sum_{j=1}^{M} F_{ij} \cdot x_j, \forall i = 1, \ldots, N$. The terms h_i can be seen as "meta channels" that have been adaptatively (in the statistical sense) determined using the $\mathcal{D}$ data set of examples.

The first ten eigen-spectra of Σ, i.e. the base functions $F_{\star i}$ ($F = L^{-1/2} \cdot V^t$), are shown in Figure 3. Each one gives particular information on the statistical dependence among the selected channels. For example, the eigen-spectrum base function 1 describes a mean general deviation from the mean spectrum. Its shape is the same as the mean spectrum in Figure 1 but it is inverted. We recognize all the absorption features described in Table 1: temperature, water vapor, ozone, CO, etc.

The eigen-spectrum base function number 2 is more related to temperature

(650-770 and 2150-2420 cm^{-1}) and ozone (1000 to 1070 cm^{-1}), less to water vapor. The eigen-spectrum base function 3 is the opposite: it gives an information more related to the 1210-2000 cm^{-1} spectral region of the water vapor. Some of higher number eigen-spectrum look more localized. For example, eigen-spectrum base function number 8 isolates at 1000 to 1070 cm^{-1} (ozone) and 2100 to 2150 cm^{-1} (CO).

The 2350-2420 cm^{-1} spectral region is dedicated to CO_2 temperature sounding. We can use the relatively "direct" link between atmospheric temperature and brightness temperature in this region to understand the behavior of the eigen-spectra. In Figure 4, the pieces of the 9 first IASI previous eigen-spectra in that spectral region are represented: the base function value is represented on the abscissa, and the wave number is represented on the ordinate. The mean spectrum of TIGR in that region is also represented (bottom right). Smaller wavenumbers are sounding high-atmosphere temperature, and larger wavenumber are sounding the near-surface temperature. It is observed that the lower-order eigen-spectra are smoother than the following ones and have a regular monotonic profile shape. For example, we see that the first eigen-spectrum is similar (but inverted) to the mean spectrum: it is then a good base function to represent the mean spectrum, i.e., a regular smooth information. The higher order eigen-spectra have more pronounced inversion(s) at different "altitudes". These base functions are used by the PCA to express the different atmospheric profiles, with an increasing amount of detail as the number of components used increases.

The interested reader can found a more detailed analysis of the eigen-spectra in [*Huang and Antonelli*, 2001] where the PCA has been used also to compress infrared high resolution spectra.

3.3. Compression of IASI spectra

Let $\bar{F}$ be the $N \times M$ truncated matrix of F. The PCA decomposition uses this truncated matrix $\bar{F}$ to project IASI spectra, x, of dimension $M = 8461$ into a space of

lower dimension N (with $N \leq M$):

$$h = \bar{F} \cdot x = F_{1\star} \cdot x_1 + \ldots + F_{N\star} \cdot x_N. \tag{4}$$

The uncompression, $\hat{x}$, of data x, given its compression h, is given by:

$$\hat{x} = \bar{F}^{-1} \cdot h \simeq h_1 \cdot F_{\star 1} + \ldots + h_N \cdot F_{\star N} \tag{5}$$

where $\bar{F}^{-1}$ is a generalized inverse matrix since $\bar{F}$ is not square. The compression error $\|x - \hat{x}\|$ is given by $\|h_{N+1} \cdot F_{\star N+1} + \ldots + h_M \cdot F_{\star M}\|$, where $\|\cdot\|$ is the Euclidean norm. PCA is optimum for the least squares errors criterion $\frac{1}{E}\sum_{e=1}^{E} \|x^e - \hat{x}^e\|^2$ [*Joliffe*, 1986].

For the compression, we only retain the N first filters, but a compromise needs to be found between a good compression level and a good compression error. Figure 5 illustrates the decreasing compression error with respect to the number of PCA components used. The more components used for compression, the lower the compression error is. With $N = 50$ filters (the 50 first principal components), the RMS compression error of IASI spectra averaged over the whole TIGR data set is close to 0.05 K, which is much lower than the averaged IASI noise which is close to 1 K.

Figure 6 shows the spectral distribution of the compression errors. The more eigen-spectra used for the compression, the lower the compression error. Taking 10 components is not enough, but with 50 components, the level of error becomes very reasonable.

A global PCA uses the same covariance (or dependency) structure, whatever the air mass, but this structure can vary with the air-mass. So a specialized PCA would be more adequate. We will not investigate this point in this paper.

3.4. De-noising of IASI Spectra

There is a possibility to suppress part of the noise during the compression process. It is assumed that the lower-order principal components $(h_1, \ldots, h_N)$ of a

PCA decomposition describe the real variability of the observations, or the signal, (here the IASI spectra) and that the remaining principal components $(h_{N+1}, \ldots, h_M)$ describe partially the instrumental noise. So, PCA representation of the spectra could advantageously be used for de-noising. Observed spectra, x, are projected into the regular subspace of the first components, describing the real variability of IASI spectra (we will comment how to choose N in the following). The PCA is performed on no-noise spectra in order that the resulting eigen-spectra are signal information and are not used to describe noise. In the compression h, the variability attributed to the instrumental noise is then partially suppressed, and the uncompression $\hat{x}$ is the resulting spectrum partially de-noised.

In Figure 7, the de-noising error (compressed and then uncompressed noisy spectrum minus no-noise spectrum) is shown with respect to the number of PCA components used for the compression. After a decrease of the error with increasing PCA number due to a better compression, the de-noising error increases. This increase of the de-noising error for an increase of number of components results from a more accurate representation of the noise. Asymptotically, the compression error would converge to zero (perfect representation of the noisy IASI spectra), but the de-noising error would converge to the instrument noise (perfect reconstruction of the noisy spectra). The optimum is given for N=30 components. This number depends not only on the spectral characteristics of the IASI observations, but also on the noise level, and on the data set (here the TIGR database) used to perform the PCA and the resulting statistics.

Figure 8 shows the spectral statistics of de-noising errors on the TIGR atmospheric database. Using only 10 components is not enough. In the first and second spectral bands, the de-noising error is still often larger than the instrumental noise. However, it is shown that the third band is already considerably de-noised (0.2 K of RMS instead of more than 2 K!). The use of 30 components for the compression/de-noising has excellent statistics: de-noising statistics for this compromise is the lowest point of the

curve. As already indicated by Figure 5, 30 components is the best compromise between the compression error, requiring a large number of components, and a de-noising error, requiring the limitation of the number of components used so as to not to represent noise.

This compromise is good, of course, only in a statistical point of view. Actually, it is interesting to note that with 200 components, some spectral regions are represented with an equivalent, or even better, de-noising level than the 30-components one (see for example 1500-1800 cm^{-1}). But statistically (i.e., on the whole spectra), the de-noising errors are higher because noise has been represented by the additional components (see first spectral band). If a spectral region is of particular interest (because of a particular constituent absorption), it is important to note that the de-noising of the entire spectrum is not necessarily the optimal solution. The particular spectral region of interest may be neglected in a statistical point of view with respect to the other channels: the compression/de-noising scheme will not sufficiently well represent this information. Control of errors for each spectral region is crucial if such particular spectral regions are of particular interest. Then, even if 30 components seem to be the perfect compromise for compression/de-noising of the whole spectrum, it might be usefull to use higher order components. Particular cases would use a combination of them. This is especially true when the regression scheme used is able to extract nonlinearly information from this components, in a non-Gaussian way. If we are interested in very localized channels, that display complex behaviour (nonlinear with respect to the amount of the absorbing constituent, unstable, etc.), a PCA, even with a high number of components, will not be ideal: it probably will use too much components to describe this complex behavior. An alternative would be to use, in that particular case, these raw specific channels.

In Figure 9, some spectral regions are represented to illustrate the compression and de-noising properties for one atmosphere. We see how our scheme is able to retrieve the signal part (i.e., no-noise spectrum) in a noisy observation (see for example the 645-650

cm^{-1} spectral region). This is particularly true for high noise-level spectral regions like 2495-2500 cm^{-1} where the scheme has used the information of flat spectrum to avoid the oscillations due to the instrument noise.

4. IASI Retrieval Method

Various inversion schemes proceed by retrieving the physical variables sequentially. In this work, we retrieve these physical variables in parallel because the inverse problem is in that case better constrained: (1) It is possible to use the nonlinear correlations or dependencies among the variables, (2) if an observation (i.e., a channel or a spectral region) is dependent simultaneously on two or more constituents, the retrieval scheme would be better suited to resolve this ambiguity, and (3) the retrieved variables will be in that case more consistent where hierarchical schemes can introduce inconsistencies. The model developed here uses a nonlinear regression of the inverse RTM in the atmosphere, obtained from a MLP neural network.

4.1. PCA-Based Pattern Recognition

For many ill-conditioned problems, the use of a first guess estimate is very important to regularize the inversion process. In the Improved Initialization Inversion (3I) retrieval scheme [*Chédin et al.*, 1985; *Scott et al., 1999*], the initial guess is found in the TIGR climate database. In the variational assimilation context, more focused on meteorology than on climatology, this first guess solution is the 6-hour prediction, see [*Prunet et al.*, 1998] for an example in the IASI context.

To retrieve such a first guess efficiently from such a data set, the Euclidean distance between observations x^0 and a spectrum from the data set, x, is often minimized

$$D_E(x^0, x) = (x^0 - x)^t \cdot (x^0 - x).$$

Another possibility is the Mahalanobis distance:

$$D_M(x^0, x) = (x^0 - x)^t \Sigma^{-1} (x^0 - x).$$

The Euclidean distance treats all variables in the same way where the Mahalanobis distance gives less weight to variables with high variance and groups highly correlated variables.

We propose here to use an Euclidean distance based on the first N PCA components, h. This distance would be equivalent to the Mahalanobis distance if we used all the PCA components ($N = M$) [*Joliffe*, 1986]. Using fewer components deletes irrelevant information and produces a faster pattern recognition step (from a distance with $M = 8461$ channels to a distance with $N = 30$ components). This distance is then used to perform a pattern recognition in a climatological data set: for each observaton x^0, the first guess is determined as the atmospheric situation of the climatological data set x such that the distance $D_E(h^0, h)$ is minimum.

Examples of pattern recognition for one TIGR atmosphere within the remaining TIGR atmospheres are presented on Figure 10 and RMS differences between first guess and real profiles are given for temperature, water vapor and ozone in Figure 11. We note that the first guess for temperature is not performant (about 4/5 K of RMS error) but this can be explained by two factors: first, the pattern recognition for one TIGR situation is made into the TIGR dataset. This can reduce by a factor of two the sampling properties of the dataset. Second, the pattern recognition is made for the whole spectra, each constituent of the atmosphere is then taken into account, and the first guess has to be a compromise between each of the variables, temperature, water vapor, ozone, etc. instead of temperature only. It is normal for the first guess error in temperature to increase with altitude since IASI has less and less information in high-level layers. A good first guess for water vapor is also difficult to obtain, the error is between 32 and 75 %, but this can be due to the fact that IASI has little or no water

vapor information for higher atmospheric layers [*Aires*, 1999]. The first guess error of the total content for water vapor is about 32.5 %. For ozone, the first guess is of good quality, between 10 and 25 %, but this is due to an insufficient representation of the ozone in the TIGR version used in this study. This point is discussed in section 5. The first guess error for ozone total content is about 10 %.

4.2. The Neural Network Model

Part of the neural network scheme developed in the next two sections is described in more detail in [*Aires et al.*, 2001]. The Multilayer Perceptron (MLP) network is a nonlinear mapping model composed of distinct layers of neurons: The first layer S_0 represents the input $X = (x_i \ ; \ \ i \in S_0)$ of the mapping. The last layer S_L represents the output mapping $Y = (y_k \ ; \ \ k \in S_L)$. The intermediate layers S_m $(0 < m < L)$ are called the "hidden layers". These layers are connected via neural links. We note W the parameters of these links. It has been demonstrated [*Hornik et al.*, 1989; *Cybenko*, 1989] that any continuous function can be represented by such a one-hidden-layer MLP.

The learning algorithm is an optimization technique that estimates the optimal network parameters W by minimizing a cost function $C_1(W)$, approaching as closely as possible the desired function. The criterion usually used to derive W is the mean squares errors in network outputs

$$C_1(W) = \frac{1}{2} \sum_{k \in S_2} \int \int D_E(\hat{y}_k(X;W), y_k)^2 \ P(X, y_k) dy_k dX, \tag{6}$$

where D_E is the Euclidean distance between y_k, the kth desired output component, and $\hat{y}_k$, the kth neural network output component, and S_2 is the output layer of the neural network.

In practice, the probability distribution function, $P(X, y_k)$, is sampled in a data set $\mathcal{B} = \{(X^e, {y_k}^e), e = 1, \ldots, E\}$ of E input/output couples, and $C_1(W)$ is then

approximated by the classical least squares criterion:

$$\tilde{C}_1(W) = \frac{1}{2E} \sum_{e=1}^{E} \sum_{k \in S_2} D_E(\hat{y}_k(X^e; W), {y_k}^e)^2. \tag{7}$$

The error back-propagation algorithm [*Rumelhart et al.*, 1986] is used to minimize $\tilde{C}_1(W)$. It is a stochastic gradient descent algorithm that is very well adapted to the MLP hierarchical architecture because the computational cost is linearly related to the number of parameters.

4.3. Learning Algorithm With First Guess

When an inverse problem is ill-posed, the solution can be nonunique and/or unstable. The use of a priori first guess information is important to reduce ambiguities: The chosen solution is then constrained so that it is physically more coherent. Statistically, this regularization avoids local minima during the learning process and speeds it up.

Introduction of a priori first guess information as part of the input to the neural network was proposed by *Aires et al.* [2001]. First, the neural transfer function becomes:

$$\hat{y} = g_W(y^b, x^o) \tag{8}$$

where $\hat{y}$ is the retrieval (i.e., retrieved physical parameters), g_W is the neural network g with parameters W, y^b is the first guess for the retrieval of physical parameters y, $x^o = RTM(y) + \eta$ are the observations, where η is the observation noise.

The learning algorithm consists of estimating the parameters W of the neural network that minimizes the mean least squares error criterion. The term "mean" depends on the probability distribution functions of the physical observation and retrieved quantities. In this experiment, the least squares criterion has the following form

$$C_2(W) = \frac{1}{2} \int \int \int D_E(g_W(y^b, x^o), y)^2 \; P(y, x^o, y^b) \tag{9}$$

$$= \frac{1}{2} \int \int \int D_E(g_W(y+\varepsilon, x+\eta), y)^2 \; P(y) P_\eta(\eta) P_\varepsilon(\varepsilon), \tag{10}$$

where $P(y)$ is the probability distribution function of the physical variables y that depends on the natural variability. $P_\eta(\eta)$ is the probability distribution function of the observation noise η. $P_\varepsilon(\varepsilon)$ is the probability distribution function of the first guess error $\varepsilon = y^b - y$.

As explained in [*Aires et al.*, 2001], the quality criterion in (9) is very similar to the quality criterion used in variational assimilation. One of the main differences is that the neural network criterion in (9) involves the distribution $P(y)$. This is due to the fact that the neural network simulates the inverse of the radiative transfer equation globally, once and for all, and uses the distribution $P(y)$ for this purpose. The neural network model is then valid for all observations (i.e., global inversion). The variational assimilation model has to compute an estimator for each observation (i.e., local inversion).

To minimize the criterion of Eq. (9), we create a data set $\mathcal{B} = \{(y^e, x^{oe}, y^{be});\ e = 1, \ldots, E\}$ that samples as well as possible all the probability distribution functions in (9). Then, the practical criterion used during the learning stage is given by:

$$\tilde{C}_2(W) = \frac{1}{2E} \sum_{e=1}^{E} D_E(g_W(y^{be}, x^{oe}), y^e))^2. \tag{11}$$

First, to sample the probability distribution function, $P(y)$, we select geophysical states (y^e) that cover all natural combinations and their correlations and by calculating $x^e = RTM(y^e)$ with the physical model (i.e., *physical* inversion). Alternatively we could obtain these relationships from a "sufficiently large" set of collocated and coincident values of x and y (i.e., *empirical* inversion). For sampling P_η, we need a priori information about the measurement noise characteristics; a physical noise model could be used, but if all we have is an estimation of the noise magnitude, then we have to assume Gaussian distributed noise η that is not correlated among the measurements (i.e., the hypothesis made for IASI, see section 2.1). To sample the first guess variability with respect to state y (i.e., sampling $P(y^b|y)$), we use a first guess

data set $\{y^{b^e};\ \ e = 1, \ldots, E\}$: this data set can be a climatological data set or a 6-hour prediction (would have better statistics of errors, but would add model dependencies). The balance between reliance on the first guess and the direct measurements is then made automatically and optimally by the neural network during the learning stage.

As for the PCA-based pattern recognition (section 4.1), the effect of using PCA components , h, instead of the raw IASI spectra, x, is that the method is faster because of the dimension reduction, and uses observation with a reduced noise level. Furthermore, the learning stage is faster since the network has less inputs and less parameters to estimate. The quality criterion in (11) is simpler because inputs are decorrelated and there are less degrees of freedom in the model and so it is easier to minimize, with less probability to become trapped in a local minimum. For a more detailed description of PCA-based regression, the reader is referred to [*Joliffe*, 1986].

4.4. Weighting in the Quality Criterion

The inputs and outputs of the network are not homogeneous, i.e. different types of variables have different dynamic ranges. As we will see in the following, solving this problem necessitates to diagnose the learning step, and control correctly the system, in contradiction with the black-box conception often associated to neural networks. The range of values, which is different for temperature or water vapor, is not the true issue here since, traditionally, the data are normalized to unity as inputs and as outputs of the neural network. The true concern is the too different dynamical range of values for the same variable. For example, the range of the water vapor path per layer can go from zero to more than 5 *cm*, with an exponential decrease with altitude. Using these physical values as outputs of the network can be misleading: an error of 0.1 *cm* in a wet situation with a total of 5 *cm* would have the same weight, during the learning stage, as an error of 0.1 *cm* in a wet situation with a total of 0.2 *cm*. So depending on the observed situation, an error of 2 % would have the same weight than an error of 50 % !

Absolute value errors are inadequate in this case. To resolve this problem, we equalize the importance of the different values. There are two possibilities: using the logarithm of the water vapor content or using a percentage error criterion

$$D(g_W(y^{be}, x^{oe}), y^e)) = \sqrt{\sum_i \left(\frac{g_W(y^{be}, x^{oe})_i - x^{oe}{}_i}{x^{oe}{}_i} \right)^2},$$

instead of the absolute RMS error

$$D_E(g_W(y^{be}, x^{oe}), y^e)) = \sqrt{\sum_i \left(g_W(y^{be}, x^{oe})_i - x^{oe}{}_i\right)^2}$$

in (11). In other words, for a global analysis, the percentage error is a more pertinent criterion than the absolute error that would have over emphasized wet atmospheric situations. We will use, in the following, the percentage of error instead of the absolute error for the water vapor and the ozone values. The counterpart of this new criterion is that the percentage of error could be exagerated for values very close to zero. We will comment this effect during the presentation of the results.

The atmospheric temperature is described by 39 output values (i.e., the 39 atmospheric levels) in the neural network where water vapor and ozone are each described by only 8 values (i.e., the 8 atmospheric layers). In order to give the same importance to each of these 3 physical variables, we use an additional weight in the criterion for each of the neural outputs: 1 for each of the temperature and 39/8 for each of the water vapor and ozone values.

5. Results For the Retrieval of Temperature, Water Vapor and Ozone Atmospheric Profiles

5.1. Data Set

Our neural network model is trained and tested using a large number of real atmospheric situations measured by radiosondes, taken from the TIGR database [*Chédin et al.*, 1985; *Achard*, 1991; *Escobar*, 1993b; *Chevallier et al.*, 1998; 2000]. We use

the TIGR3 database composed of 2311 atmospheres: 872 in tropical air-mass, 388 in mid-latitude type 1, 354 in mid-latitude type 2, 104 in polar type 1 and 593 in polar type 2. These atmospheres are described by their temperature and gas concentration profiles. For the retrieval scheme, we have used the discretization described in Table 3. The discretization in temperature is the same as the one used by ECMWF, except for the 3 near-surface levels (37, 38, and 39) that are each the combination of two ECMWF levels which we consider too thin for IASI (the first ECMWF level is at 60 m). For water vapor, we take layers of about 2 km, which follows the recommendation for IASI. The ozone discretization is not regular; it emphasizes the layers near 30 hPa where the ozone abondance is maximum. The water vapor and ozone discretizations are kept as a sub-discretizations of the ECMWF scheme.

The TIGR atmospheres, selected from a collection of more than 150,000 radiosonde measurements, include a large number of rare events. Not only is the range of variability occasionally extreme, but also the occurence and strength of inversions in the profiles imtroduces complicated structures that are very challenging to any retrieval method. These very complex profiles are much more irregular than reanalysis data or model output data. The data set represents, as much as possible, all kinds of possible atmospheric situations. This complexity represents a higher variability than that encountered under operational conditions where model output is used as the first guess, so our estimate of the retrieval errors could be an over-estimate. However the use of a large and complex climatological data set allows the inversion model to be calibrated globally including rare events.

The ozone variability representation is not sufficient in this version of TIGR. So, it is expected that in our results, the retrieval error for ozone be probably an under-estimate of the correct error level for IASI. A new data set is presently being developed to improve the ozone representation.

The RTIASI forward radiative transfer algorithm developed at ECMWF [*Matricardi*

and Saunders, 1999] has been used to compute the IASI brightness temperatures associated with the TIGR atmospheres for clear conditions over the sea. For each 2311 atmospheres of TIGR, we have simulated 5 noise realization using the specifications for the instrument, Table 2 and equation (1). Our training and testing data set is then composed of 11,555 examples. We have specialized a neural network, NN1, for wet atmospheres (precipitable water amount larger than 1 cm) and another one, NN2, for dry atmospheres (precipitable water amount lower than 1 cm). We have 5,775 examples in the first case and 5,780 for the second case. The choice of dry or wet configuration can be made using the first guess.

5.2. Wet Atmosphere Configuration

The $E = 5,775$ wet examples have been randomly separated into two subsets: a training set of 5000 examples and a testing set of 775 examples. We take 100 PCA components (i.e. more than the optimal 30 components for de-noising) as inputs for the IASI observation part since the NN is able to use only the information that it needs for the desired retrieval: $x^0 = h$. It is possible that between the 30th and the 100th PCA components, there are information on specific spectral region, not statistically important on the whole spectrum, that is useful for the NN retrieval.

The architecture of the network NN1 is a MLP (Figure 12) with 155 inputs coding the $M = 100$ PCA components, $x^0 = h$, and the first guess, x_b (39 temperature, 8 water vapor and 8 ozone values), 80 neurons in the hidden layer, and 55 neurons in the output layer coding the retrieval, x. The number of neurons in the hidden layer is estimated by a heuristic procedure that monitors the generalization errors of the neural network as the configuration is varied: for a too small number of neurons in the hidden layer, the generalization of the neural network is insufficient because of the lack of complexity of the neural architecture to represent the desired model (i.e., bias error). For a number of neurons too large, the complexity of the neural network is too rich

compared to the desired model and the overfitting problem, where the learning error is small, but the generalization error is big, appear (i.e. variance error). This dilema is called bias/variance dilema [*Geman et al.*, 1992]. In practice, we use different number of neurons in the hidden layer and the smaller generalization error indicates the best compromise.

Figure 13 represents the learning and the testing curves of some of the retrieved quantities during the learning stage. The purpose of this figure is to show how the inhomogeneity of the outputs in the neural network can be a problem. The water vapor is much more complex to retrieve than ozone or temperature: the curve has plateaus which correspond to local minima, where the error can not be decreased, and error increases, when the learning overshoots the local minima. To control this kind of problems, it is important to give an uniform weight to each of the variables, this is the reason why we have modified our quality criterion as explained in section 4.4. Even with this new criterion, the water vapor can still be trapped at some stages, while other variables (like the temperature or ozone) continue to improve. However, at some time, the constraints between water vapor and temperature or ozone are so strongly violated that the optimization algorithm changes the water vapor to bypass this local minima: first, an increase of the error and, then, a decrease of the error. This shows how it can be advantageous to retrieve in parallel more than one physical variable, the problem being better constrained.

Figure 14 presents three examples of retrievals. We see, in each case, a major improvement of the temperature profile retrieval over the first guess: true profile and the noisy retrieval are difficult to distinguish in this figure. This is also true for the water vapor retrieval. Ozone is also very good, but the first guess was already very close to the correct solution. Consequently, the retrieval statistics for wet atmospheres, represented in Figure 15, are good. The RMS temperature error is mostly below 1 K, being in the 0.5-0.7 K range for level between 900 and 250 hPa. We have already

shown in [*Aires*, 1999] that the fusion of the information from AMSU would improve significantly the temperature retrieval above 200 hPa. The retrieval of water vapor is very good: 5 % error for total water vapor path, 10 % for the first 3 atmospheric layers, then errors in the range 10-20 %, except for the layer near 300 hPa. The peack error in the test retrieval of water vapor at 300 hPa is probably due to an insufficient number of atmospheres in the training data set. Ozone retrieval is very good, but this retrieval is too optimistic because of insufficient variability in the data set.

5.3. Dry Atmosphere Configuration

The $E = 5,780$ dry examples have been randomly separated in two subsets: a training set of 5000 examples and a testing set of 780 examples. The architecture of the network NN2 (Figure 12) is the same as NN1: 155 inputs, 80 neurons in the hidden layer, and 55 neurons in the output layer.

Figure 16 presents three examples of retrievals. The same comments as for wet conditions hold: the overall retrieval of temperature, water vapor and ozone seems good. However, we see some small error in the retrieval of atmospheric temperature above 200-100 hPa (see for example temperature profile B). Also, errors can appear when the true profile posseses a too strong inversion (see profile C at level 100 hPa). Water vapor is well retrieved, a small over-estimate can be observed for atmosphere B. Retrieval errors for ozone are small; even when the first guess is already close to the true profile, like atmosphere C, the retrieval scheme still improves the retrieval.

Figure 17 shows the RMS retrieval errors for temperature, water vapor and ozone for the dry condition neural network. The retrieval of temperature is more difficult in dry condition than in wet conditions (Figure 15). The RMS error is < 1 K, except for near-surface levels, due to near-surface inversions, and near 200 hPa. The total water vapor content is retrieved with 7 % mean percentage error and only three atmospheric layers (around 300 hPa) are above 15 % mean error. It is important at this point to note

that the percentage error is not a perfect measure of the errors: for zero or near-zero content, the percentage error has no significance. For example, retrieving a content of 0.0002 *cm* for an actual true value of 0.0001 *cm* would produce a percentage error of 300 %, even if the absolute error is very small. Furthermore, the physical limitations of the IASI instrument, in terms of signal to noise ratio, will not allow a good retrieval for very low water vapor contents. Figure 17(B) shows the mean percentage error without the contribution of the low water vapor content cases (less than 0.01 *cm*); percentage mean error becomes uniform with height at 15 %, which is a good result for these dry situations.

6. Conclusion and Perspectives

We have developed a PCA-based method for compressing, de-noising, and first guess retrieval for the high-resolution interferometer IASI. Our approach allows for a more complete exploitation of the information in the IASI spectra. In particular using the redundancy among channels for noise reduction and the nonlinear correlations to provide more strongly constrained retrieval. These pre-processing steps (compression, de-noising, first guess retrieval) are a crucial step in the neural network retrieval, but this approach is completely general and does not depend on retrieval method. For example, our compression/de-noising approach could be used in a variational assimilation scheme: dimension of data is smaller, noise is reduced, and variables are decorrelated. This would simplify calculations and speed the scheme.

We also developed a neural network retrieval scheme which uses first guess information. This additional information has the advantage of better constraining the inverse problem, improving retrieval results. This neural network approach does not need Jacobians as in classical inversion algorithms. The simultaneous retrieval of many variables is also a crucial point, since it allows us to exploit the complex inter-dependencies among the observations, among the variables and between observations

and variables for a better constrained inverse problem.

Our experiments were made with the TIGR database, i.e. a vast and complex set of real atmospheric situations (from radiosonde measurements which are much more irregular than model output) with rare events. This fact provides a global applicability of our method. The retrieval errors are good: temperature is retrieved with an error under or close to 1 K, total amount of water vapor has a mean percentage error between 5 and 7 %. Amospheric water vapor layers is retrieved with error between 10 and 15 % most of the time. Statistics of ozone retrieval are too much optimistic due to a lack of representation of ozone variability of our data set.

It is important to note that the results obtained for the IASI retrieval are entirely dependent on the complexity of the data set used to perform the statistics. Thus, it has been demonstrated, in this work, that with our complex atmospheric situations, the potentialities of the IASI instrument allows reaching the WMO specifications on realistic conditions. This new instrument will be a clear advance compared to the previous instruments. It has been shown also that the MLP inversion technique is a pertinent method for the processing of IASI observations. It is flexible enough to introduce a priori information in the retrieval scheme, it is robust to noise, and it is very fast and accurate. This new scheme is then a privilegiate candidate for the processing of IASI observations.

There are various perspectives for this work. First, a more optimal de-noising approach would be to perform a PCA for each air mass. In effect, using a global PCA, the same statistical structure of dependencies is used for each air-mass, which can be non-optimal. A specialized PCA is expected to even better describe the natural variability on IASI observations. A new TIGR is under development where the ozone variability is improved. Another advantage of our approach is that it can easily accomodate nonlinear relationships between the information from other instruments [*Prigent et al.*, 2001]. This is a particularly interesting feature since IASI results on

high-altitude atmosphere temperature is expected to be improved by AMSU [*Aires*, 1999]. Our algorithm needs also to be extended to land, and to take into account cloudy conditions; for that purpose, we will capitalize on our work on the SSM/I instrument [*Aires et al.*, 2001].

Notation

y	vector of physical variables to retrieve.
$\hat{y}$	estimate of y.
y^b	first guess a priori information for x.
ε	$= y^b - y$, first guess error.
$RTM(y)$	radiative transfer model for the physical variables y (also a vector).
x^o	IASI brightness temperature spectrum observations.
M	dimension of the IASI spectrum x.
η	IASI instrumental noise.
P	generic probability measure.
$P_\eta(\eta)$	probability distribution function of η.
$P_\varepsilon(\varepsilon)$	probability distribution function of ε.
E	number of samples in the data set.
h	PCA compression of the IASI spectrum x.
N	dimension of the compression h ($N \leq M$).
Σ	$M \times M$ covariance matrix of spectra x.
V	$M \times M$ matrix of eigenvectors of Σ.
L	$m \times M$ diagonal matrix of eigenvalues of Σ.
F	$M \times M$ filter matrix.
I	Identity matrix.
$< \cdot >$	expectation operator.
σ	sigmoid function of the neural network.
g_W	neural network model, or transfer function for our application.
W	$= \{w_{ij}\}$, the set of the parameters of the

References

Achard, V., Trois problèmes clés de l'analyse tridimensionelle de la structure thermodynamique de l'atmosphère par satellite : mesure du contenu en ozone, classification des masses d'air, modélisation hyper-rapide du transfert radiatif, PhD thesis, Université Pierre et Marie Curie (Paris VI), 1991.

Aires, F., R. Armante, A. Chédin, and N.A. Scott, Surface and atmospheric temperature retrieval with the high resolution interferometer IASI, *Proc. Am. Meteorol. Soc., 98,* 181–186, 1998.

Aires, F., Problèmes inverses et réseaux de neurones: Application à l'interféromètre haute résolution IASI et à l'analyse de séries temporelles, *PhD Thesis*, 220 pp., Université Paris IX/Dauphine, 1999.

Aires, F., M. Schmitt, N.A. Scott, and A. Chédin, The weight smoothing regularization for Jacobian stabilization, *IEEE Trans. Neural Networks, 10*(6), 1502–1510, 1999.

Aires, F., A. Chédin, N.A. Scott, and W.B. Rossow, A regularized neural net approach for retrieval of atmospheric and surface temperatures with the IASI instrument, submitted to *J. of Applied Meteorology*, 2000.

Aires, F., C. Prigent, W.B. Rossow, and M. Rothstein, A new neural network approach including first guess for retrieval of atmospheric water vapor, cloud liquid water path, surface temperature and emissivities over land from satellite microwave observations, *J. of Geophys. Res.*, in press, 2001.

Amato, U., and C. Serio, The impact of random noise on the performance of a new infrared atmospheric sounding interferometer (IASI) configuration, *Int. J. Remote Sensing, 18*(15), 3,135–3,143, 1997.

Cayla, F., B. Tournier, and P. Hebert, Performance budgets of IASI options, Technical Report IA-TN-0B-5476-CNE, CNES, 1995.

Chaboureau, J.-P., A. Chédin, and N.A. Scott, Remote sensing of the vertical

distribution of atmospheric water vapor from the TOVS observations. Method and validation, *J. Geophys. Res., 103*, 8743–8752, 1998.

Chédin, A., N.A. Scott, C. Wahiche, and P. Moulinier, The Improved Initialization Inversion method: a high resolution physical method for temperature retrievals from Tiros-n series, *J. Climate Appl. Meteor., 24*, 128–143, 1985.

Chevallier, F., F. Chéruy, N.A. Scott, and A. Chédin, Neural network approach for a fast and accurate computation of the LongWave radiation budget, *J. Appl. Meteor., 37*, 1385–1397, 1998.

Chevallier, F., J.-J. Morcrette, F. Chéruy, and N.A. Scott, Use of a neural network-based longwave radiative transfer scheme in the ECMWF atmospheric model, *Quart. J. Roy. Meteor. Soc., 126*, 761-776, 2000.

Cybenko, G., Approximation by superpositions of a sigmoidal function, *Math. Control Signals Syst., 2*, 303–314, 1989.

Escobar, J., A. Chédin, F. Chéruy, and N.A. Scott, Réseaux de neurones multicouches pour la restitution de variables thermodynamiques atmosphériques à l'aide de sondeurs verticaux satellitaires, *C. R. Acad. Sci. Paris, 317*(2), 911–918, 1993.

Escobar, J., Base de données pour la restitution de paramètres atmosphériques à l'échelle globale ; étude sur l'inversion par réseaux de neurones des données des sondeurs verticaux atmosphériques satellitaires présents et à venir, *PhD Thesis*, Université Denis Diderot (Paris VII), 1993b.

Eyre, J.R., G.A. Kelly, A.P. MacNelly, E. Anderson, and A. Persson, Assimilation of TOVS radiance information through one-dimensional variational analysis, *Q. J. R. Meteorol. Soc., 119*, 1427–1463, 1993.

Geman, S., E. Bienenstock, and R. Doursat, Neural networks and the bias-variance dilema, *Neural Computation, 1*(4), 1–58, 1992.

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

Huang, H.-L., P. Antonelli, Application of Principal Component Analysis to High-Resolution Infrared Measurement Compression and Retrieval, *J. Climate Appl. Meteor.*, *40*, 365–3388, 2001.

Jain, A. and D. Zongker, Feature selection: evaluation, application, and small sample performance, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, *19*(2), 153–158, 1997.

Jolliffe, I.T., Principal component analysis, *Springer series in statistics*, 1986.

Krasnopolsky, V.M., G.H. Gemmill, and L.C. Breaker, A neural network multiparameter algorithm for SSM/I ocean retrievals: Comparisons validations, *Remote Sens. Environ.*, *73*, 133–142, 2000.

Matricardi, M., and R. Saunders, Fast radiative transfer model for simulation of infrared atmospheric sounding interferometer radiances, *Appl. Opt.*, *38*(27), 5679–5691, 1999.

Prunet, P., J.-N. Thépaut, and V. Cassé, The information content of clear sky IASI radiances and their potential for numerical weather prediction, *Q. J. R. Meteorol. Soc.*, *124*, 211–241, 1998.

Prigent, C., A. Aires, W.B. Rossow, and E. Matthews, Joint characterization of the vegetation by satellite observations from visible to microwave wavelengths: A sensitivity analysis, *J. Geophys. Res.*, in press, 2001.

Rabier, F., D. Chafai and R. Abboudi, Comparison of channel selection methods for IASI, *Proc. of the Eleventh International TOVS Study Conference, 00*, Budapest, 2000.

Rodgers, C.D., Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation, *Rev. Geophys.*, *14*, 609-624, 1976.

Rodgers, C.D., Characterizatioon and error analysis of profiles retrieved from remote sounding measurements, *J. Geophys. Res.*, *95*, 5587–5595, 1990.

Rumelhart, D.E., G.E. Hinton, and R.J. Williams, Learning Internal Representations

by Error Propagation, in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. I, Foudations*, edited by D.E. Rumelhart, J.L. McClelland, and the PDP Research Group, pp. 318–362, MIT Press, Cambridge, 1986.

Schlussel, P., and M. Goldberg, Retrieval of temperature and water vapour from IASI measurements in partly cloudy situations, *Advances in Space Research (Cospar Information Bulletin)*, submitted, 2001.

Scott, N.A., and A. Chédin, A fast line-by-line method for atmospheric absorption computations: the automatized atmospheric absorption atlas, *J. Appl. Meteor.*, *20*, 20802–20812, 1981.

Thépaut, J.-N., R.N. Hoffman, and P. Courtier, Interactions of dynamics and observations in a four-dimensional variational assimilation, *Mon. Weather Rev.*, *121*(112), 3393–3414, 1993.

F. Aires, Department of Applied Physics, Columbia University, NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025 (faires@giss.nasa.gov)

W.B. Rossow, NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025 (wrossow@giss.nasa.gov)

N.A. Scott and A. Chédin, Laboratoire de Météorologie Dynamique, École Polytechnique, 91128 Palaiseau Cedex, France (chedin@jungle.polytechnique.fr, scott@araf1.polytechnique.fr)

Received ... .., 2001; revised ...., 2001; accepted ...., 2001.

To appear in the *Journal of Geophysical Research*, 2001.

Table 1. IASI Spectral Information

Spectral Region (cm^{-1})	Variable
650 to 770	CO_2 temperature sounding
770 to 980	surfaces and cloud properties
1000 to 1070	O_3 sounding
1080 to 1150	surfaces and cloud properties
1210 to 1650	water vapour sounding; N_2O and CH_4 column amounts
2100 to 2150	CO column amount
2150 to 2250	CO_2 temperature sounding; N_2O column amount
2350 to 2420	CO_2 temperature sounding
2420 to 2700	surfaces and cloud properties
2700 to 2760	CH_4 column amount

Table 2. $NE\Delta T$ Noise Characteristics of IASI at 280 K

ν in cm^{-1}	$NE\Delta T$ in K
650	0.419
700	0.157
750	0.145
800	0.145
850	0.150
900	0.150
950	0.165
1000	0.165
1050	0.176
1100	0.200
1150	0.200
1200	0.095
1250	0.096
1300	0.098
1350	0.100
1400	0.105
1450	0.105
1500	0.111
1550	0.116
1600	0.125
1650	0.137
1700	0.160
1750	0.170
1800	0.200
1850	0.224
1900	0.250
1950	0.240
2000	0.130
2050	0.135
2100	0.141
2150	0.151
2200	0.172
2250	0.200
2300	0.239
2350	0.287
2400	0.351
2450	0.400
2500	0.700
2550	0.900
2600	1.100
2650	1.300

Table 2. (continued)

ν in cm^{-1}	$NE\Delta T$ in K
2700	1.600
2750	1.935

Table 3. Temperature Levels, Water Vapor and Ozone Layers for the IASI Retrieval Scheme

Layer or Level Number	Temperature Levels (hPa)	Water Vapor Layers (hPa)	Ozone Layers (hPa)
1	0.100	0.10 to 167.95	0.10 to 0.69
2	0.290	167.95 to 253.71	0.69 to 2.61
3	0.690	253.71 to 358.28	2.61 to 20.40
4	1.420	358.28 to 478.54	20.40 to 45.29
5	2.611	478.54 to 610.60	45.29 to 69.97
6	4.407	610.60 to 795.09	69.97 to 102.05
7	6.950	795.09 to 1013.25	102.05 to 1013.25
8	10.370	0.10 to 1013.25	0.10 to 1013.25
9	14.810	...	...
10	20.400	...	...
11	27.260	...	...
12	35.510	...	...
13	45.290	...	...
14	56.730	...	...
15	69.970	...	...
16	85.180	...	...
17	102.050	...	...
18	122.040	...	...
19	143.840	...	...
20	167.950	...	...
21	194.360	...	...
22	222.940	...	...
23	253.710	...	...
24	286.600	...	...
25	321.500	...	...
26	358.280	...	...
27	396.810	...	...
28	436.950	...	...
29	478.540	...	...
30	521.460	...	...
31	565.540	...	...
32	610.600	...	...
33	656.430	...	...
34	702.730	...	...
35	749.120	...	...
36	(795.090+839.950)/2.	...	...
37	(882.800+922.460)/2.	...	...
38	(957.440+985.880)/2.	...	...
39	(1005.430+1013.25)/2.	...	...

Figure 1. Mean IASI spectrum (left) and corresponding standard deviation of IASI instrumental noise (right). Principal spectral absortion regions are indicated, as in Table 1

Figure 2. Cumulated explained variance percentage of IASI spectra with respect to the number of PCA components

Figure 3. First 10 IASI eigen-spectrum base functions

Figure 4. Interpretation of IASI eigen-spectra for temperature, in the 2350-2420 cm^{-1} spectral region

Figure 5. Compression error with respect to the number of PCA components used

Figure 6. Statistics of compression in the 3 spectral bands of IASI for 10 components (upper line) and 200 components (lower line), intrumental noise standard deviation is represented in grey for comparison purpose

Figure 7. De-noising error (continuous line), and overall instrumental noise (dashed line), with respect to the number of PCA components used in the compression

Figure 8. Statistics of de-noising errors in the 3 spectral bands of IASI using 10, 30 and 200 PCA components, for the TIGR situations, instrument noise (red line) is shown for comparison purpose

Figure 9. Comparison of one noise-free spectrum (dotted line with points), the same spectrum with noise (continuous line), and the corresponding de-noised spectrum using 30 PCA components (dashed line)

Figure 10. First guess retrieval examples: (A) tropical, (B) temperate, and (C) polar situations. Near surface values for water vapor and ozone represent the total vertical content

Figure 11. RMS error of the first guess for (A) temperature, (B) water vapor, and (C) ozone. Near surface values for water vapor and ozone represent the total vertical content

Figure 12. Architecture of a MLP neural network with first guess a priori information: y^b is the climatological first guess, x^o is the IASI observation (brightness temperature spectrum compressed and de-noised by PCA), and y is the neural network retrieval

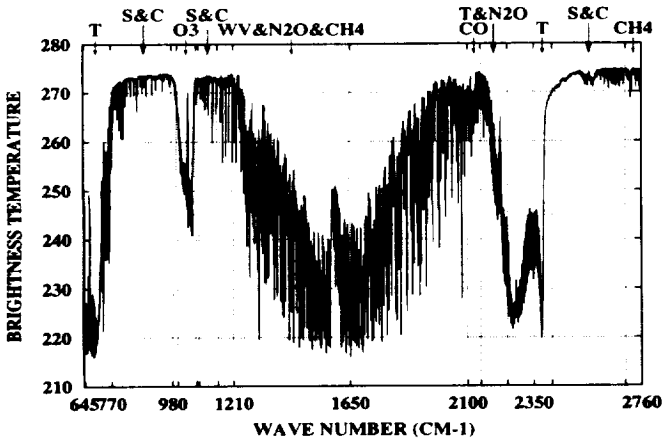
Figure 13. Learning curves for (A) temperature at 817 hPa, (B) water vapor between 358 and 478 hPa, and (C) ozone between 20 and 45 hPa

Figure 14. Three examples (A, B, and C) of temperature, water vapor, and ozone atmospheric profiles retrieval in the wet atmospheres configuration. Near surface values for water vapor and ozone represent the total vertical content

Figure 15. Error profile for the retrievals in the learning set (continuous line) and in the generalization set (discontinuous line) for temperature (A), water vapor (B), and ozone (C): Wet atmospheres configuration. Near surface values for water vapor and ozone represent the total vertical content

Figure 16. Three examples (A, B, and C) of temperature, water vapor, and ozone atmospheric profiles retrieval in the dry atmospheres configuration. Near surface values for water vapor and ozone represent the total vertical content

Figure 17. Error profile for retrievals in the learning set (continuous line) and in the generalization set (discontinuous line) for temperature (A), water vapor (B), and ozone (C): Dry atmospheres configuration. Near surface values for water vapor and ozone represent the total vertical content


T
S&C
O3
S&C
WV&N2O&CH4
T&N2O
CO
T
S&C
CH4
BRIGHTNESS TEMPERATURE
280
270
260
250
240
230
220
210
645770
980
1210
1650
2100
2350
2760
WAVE NUMBER (CM-1)

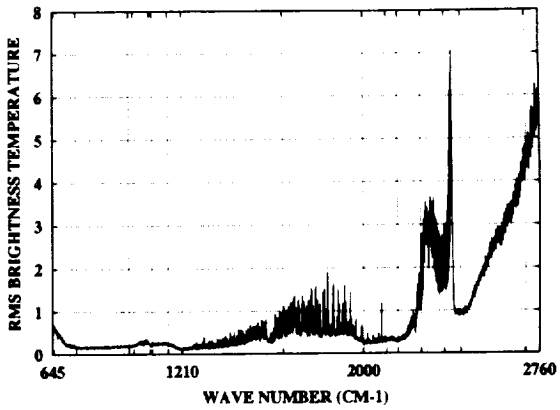

RMS BRIGHTNESS TEMPERATURE
8
7
6
5
4
3
2
1
0
645
1210
2000
2760
WAVE NUMBER (CM-1)

Figure 1.

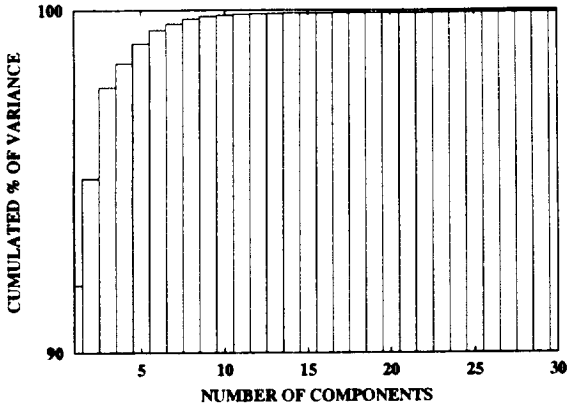

100
90
CUMULATED % OF VARIANCE
5
10
15
20
25
30
NUMBER OF COMPONENTS

Figure 2.

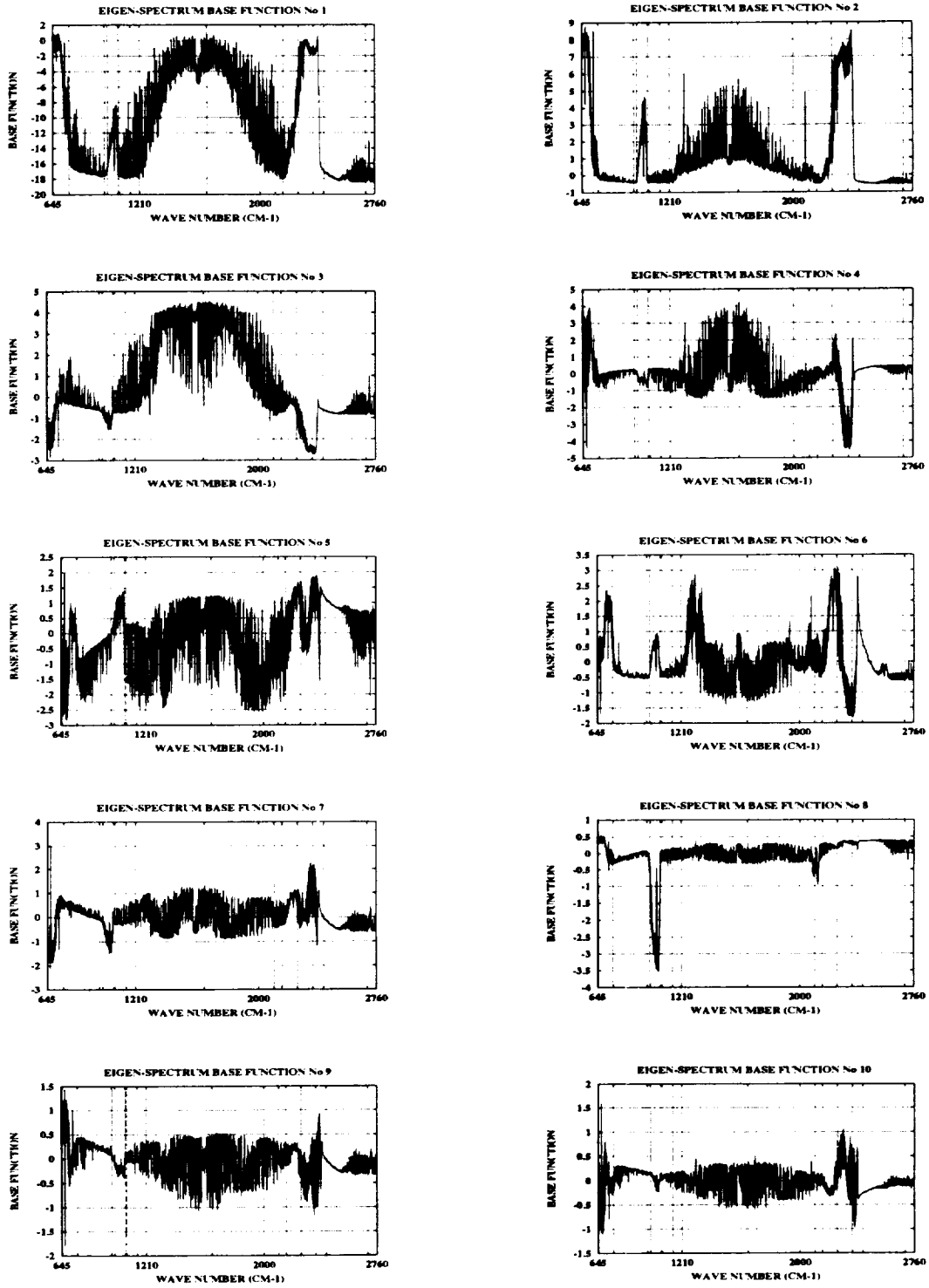

EIGEN-SPECTRUM BASE FUNCTION No 1
EIGEN-SPECTRUM BASE FUNCTION No 2
EIGEN-SPECTRUM BASE FUNCTION No 3
EIGEN-SPECTRUM BASE FUNCTION No 4
EIGEN-SPECTRUM BASE FUNCTION No 5
EIGEN-SPECTRUM BASE FUNCTION No 6
EIGEN-SPECTRUM BASE FUNCTION No 7
EIGEN-SPECTRUM BASE FUNCTION No 8
EIGEN-SPECTRUM BASE FUNCTION No 9
EIGEN-SPECTRUM BASE FUNCTION No 10
BASE FUNCTION
WAVE NUMBER (CM-1)
645
1210
2000
2760

Figure 3.

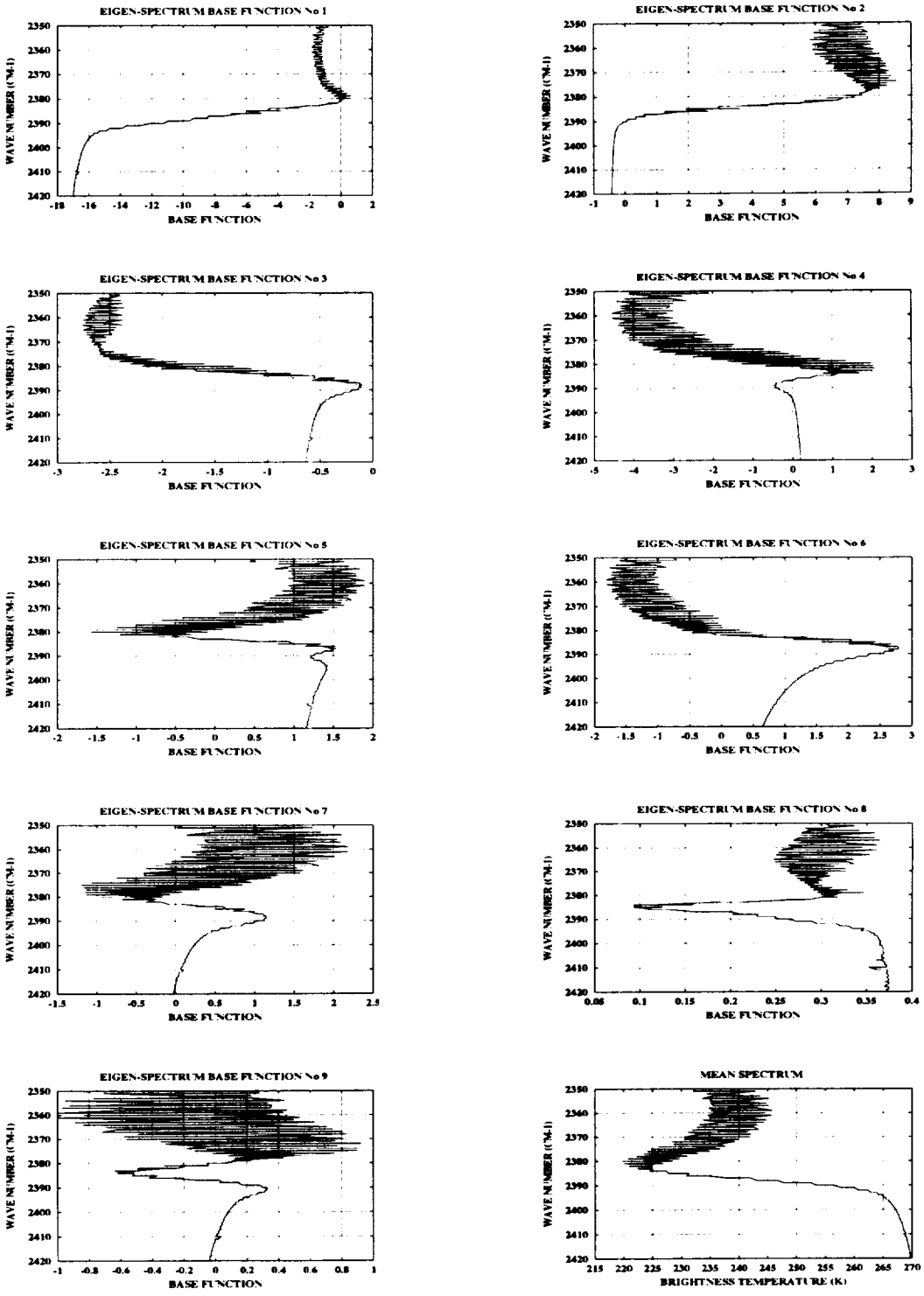

EIGEN-SPECTRUM BASE FUNCTION No 1
EIGEN-SPECTRUM BASE FUNCTION No 2
EIGEN-SPECTRUM BASE FUNCTION No 3
EIGEN-SPECTRUM BASE FUNCTION No 4
EIGEN-SPECTRUM BASE FUNCTION No 5
EIGEN-SPECTRUM BASE FUNCTION No 6
EIGEN-SPECTRUM BASE FUNCTION No 7
EIGEN-SPECTRUM BASE FUNCTION No 8
EIGEN-SPECTRUM BASE FUNCTION No 9
MEAN SPECTRUM
WAVE NUMBER (CM-1)
BASE FUNCTION
BRIGHTNESS TEMPERATURE (K)

Figure 4.

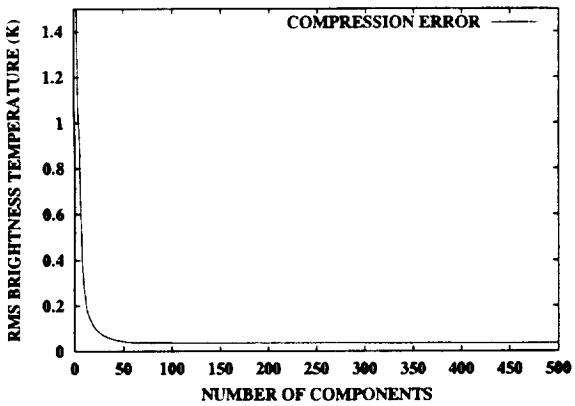
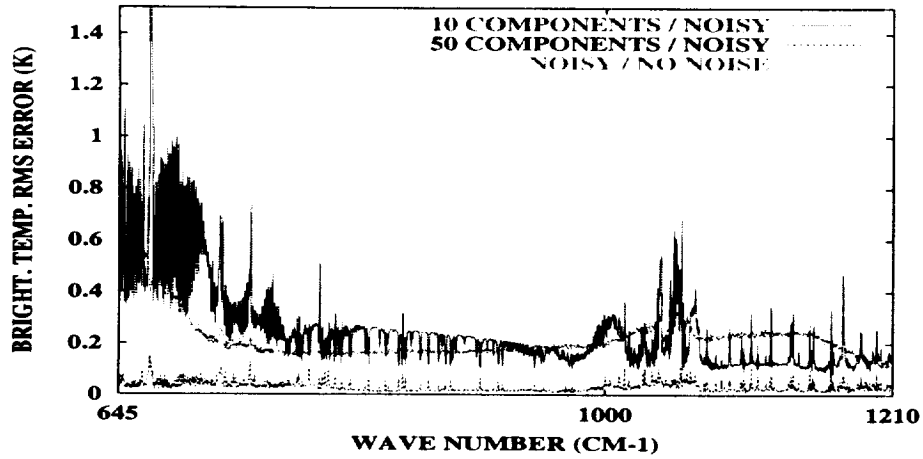
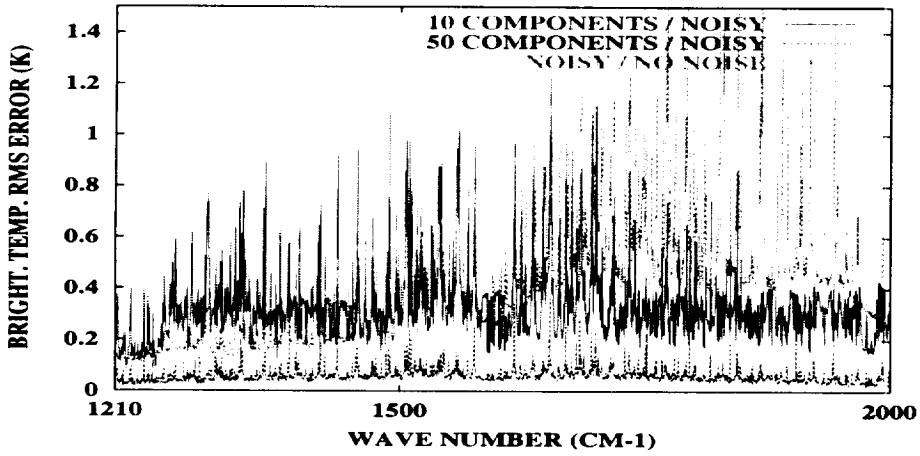

COMPRESSION ERROR
RMS BRIGHTNESS TEMPERATURE (K)
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
50
100
150
200
250
300
350
400
450
500
NUMBER OF COMPONENTS

Figure 5.


10 COMPONENTS / NOISY
50 COMPONENTS / NOISY
NOISY / NO NOISE
BRIGHT. TEMP. RMS ERROR (K)
1.4
1.2
1
0.8
0.6
0.4
0.2
0
645
1000
1210
WAVE NUMBER (CM-1)


10 COMPONENTS / NOISY
50 COMPONENTS / NOISY
NOISY / NO NOISE
BRIGHT. TEMP. RMS ERROR (K)
1.4
1.2
1
0.8
0.6
0.4
0.2
0
1210
1500
2000
WAVE NUMBER (CM-1)

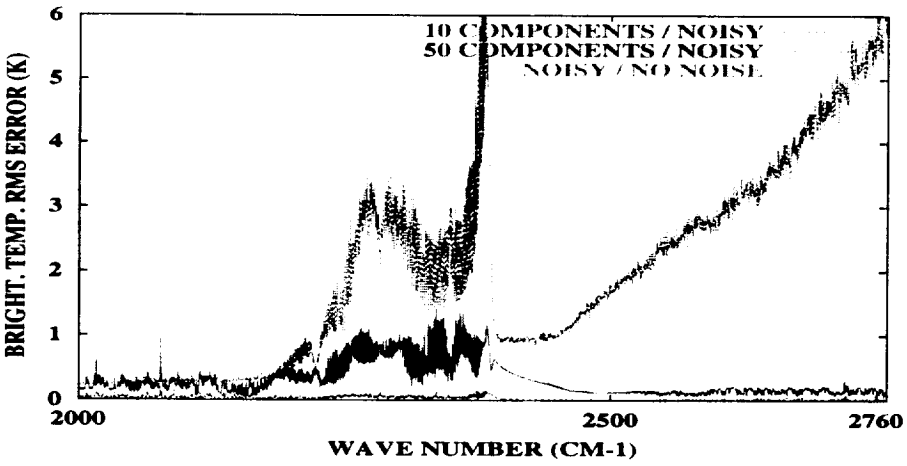

10 COMPONENTS / NOISY
50 COMPONENTS / NOISY
NOISY / NO NOISE
BRIGHT. TEMP. RMS ERROR (K)
6
5
4
3
2
1
0
2000
2500
2760
WAVE NUMBER (CM-1)

Figure 6.

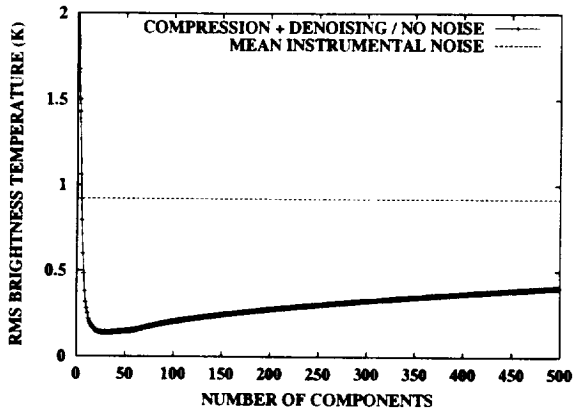
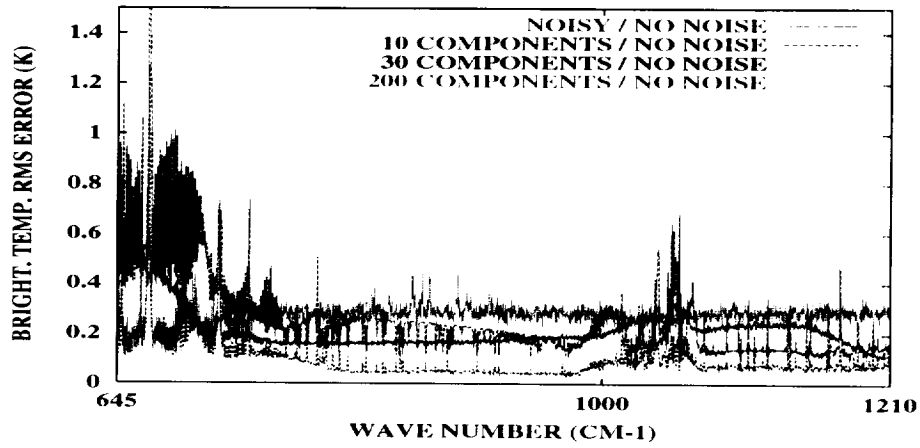
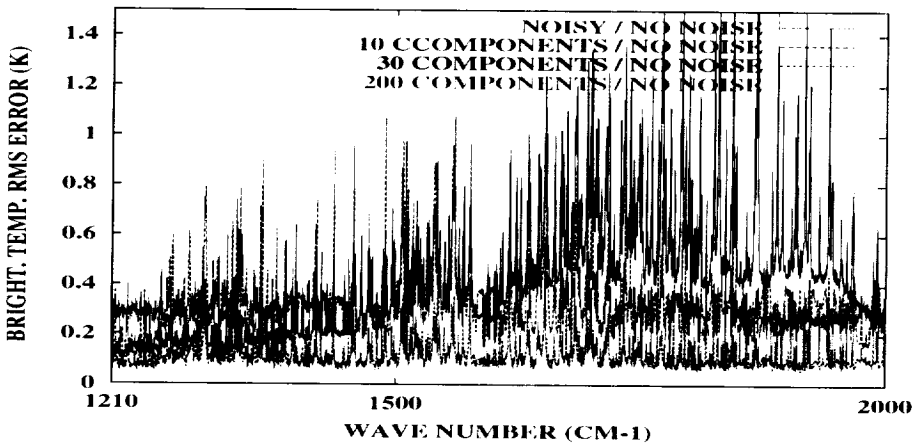

COMPRESSION + DENOISING / NO NOISE
MEAN INSTRUMENTAL NOISE
RMS BRIGHTNESS TEMPERATURE (K)
NUMBER OF COMPONENTS
0
0.5
1
1.5
2
0
50
100
150
200
250
300
350
400
450
500

Figure 7.


NOISY / NO NOISE
10 COMPONENTS / NO NOISE
30 COMPONENTS / NO NOISE
200 COMPONENTS / NO NOISE
BRIGHT. TEMP. RMS ERROR (K)
1.4
1.2
1
0.8
0.6
0.4
0.2
0
645
1000
1210
WAVE NUMBER (CM-1)


NOISY / NO NOISE
10 CCOMPONENTS / NO NOISE
30 COMPONENTS / NO NOISE
200 COMPONENTS / NO NOISE
BRIGHT. TEMP. RMS ERROR (K)
1.4
1.2
1
0.8
0.6
0.4
0.2
0
1210
1500
2000
WAVE NUMBER (CM-1)

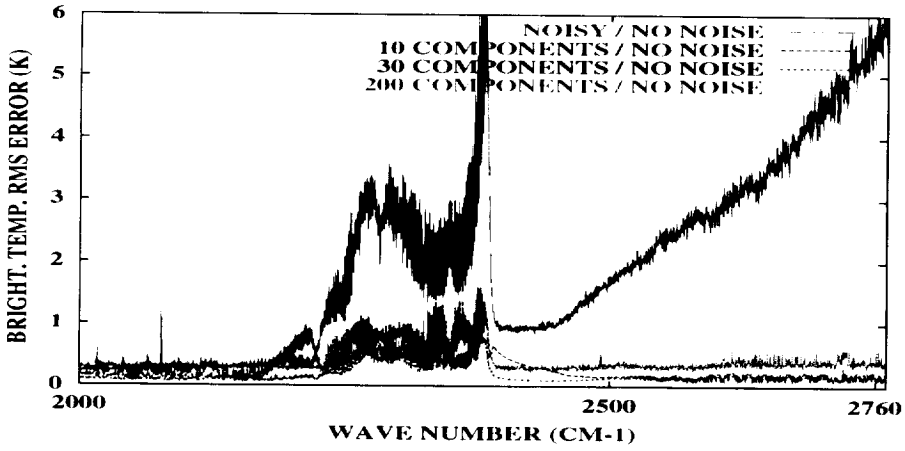

NOISY / NO NOISE
10 COMPONENTS / NO NOISE
30 COMPONENTS / NO NOISE
200 COMPONENTS / NO NOISE
BRIGHT. TEMP. RMS ERROR (K)
6
5
4
3
2
1
0
2000
2500
2760
WAVE NUMBER (CM-1)

Figure 8.

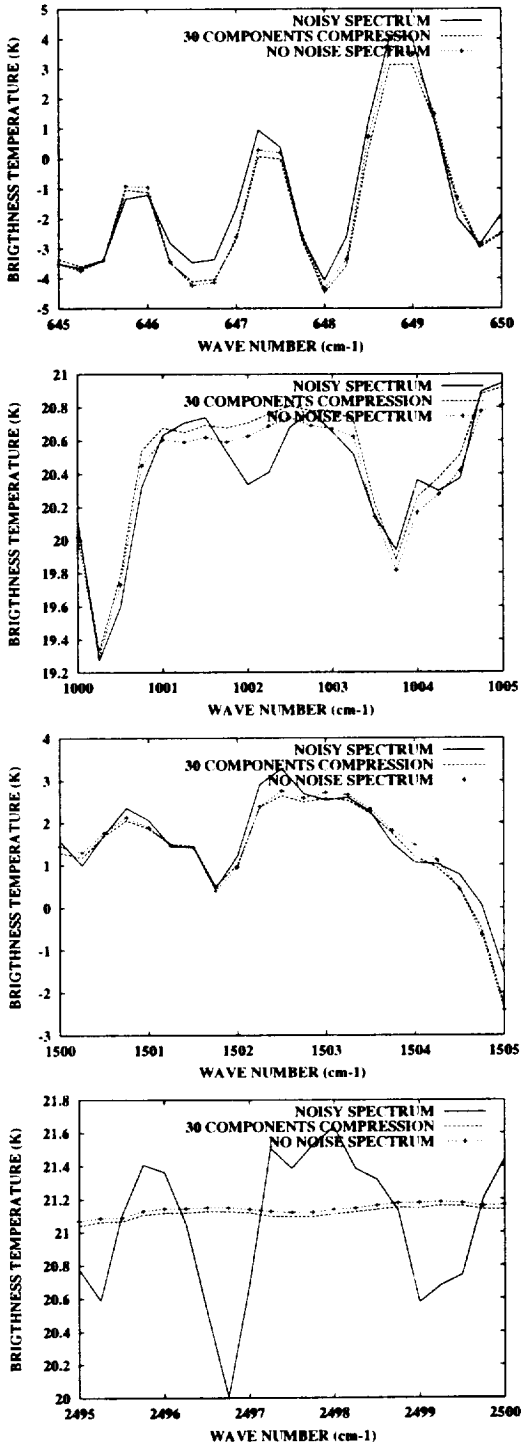

NOISY SPECTRUM
30 COMPONENTS COMPRESSION
NO NOISE SPECTRUM
BRIGTHNESS TEMPERATURE (K)
WAVE NUMBER (cm-1)
5
4
3
2
1
0
-1
-2
-3
-4
-5
645
646
647
648
649
650
21
20.8
20.6
20.4
20.2
20
19.8
19.6
19.4
19.2
1000
1001
1002
1003
1004
1005
4
3
2
1
0
-1
-2
-3
1500
1501
1502
1503
1504
1505
21.8
21.6
21.4
21.2
21
20.8
20.6
20.4
20.2
20
2495
2496
2497
2498
2499
2500

Figure 9.

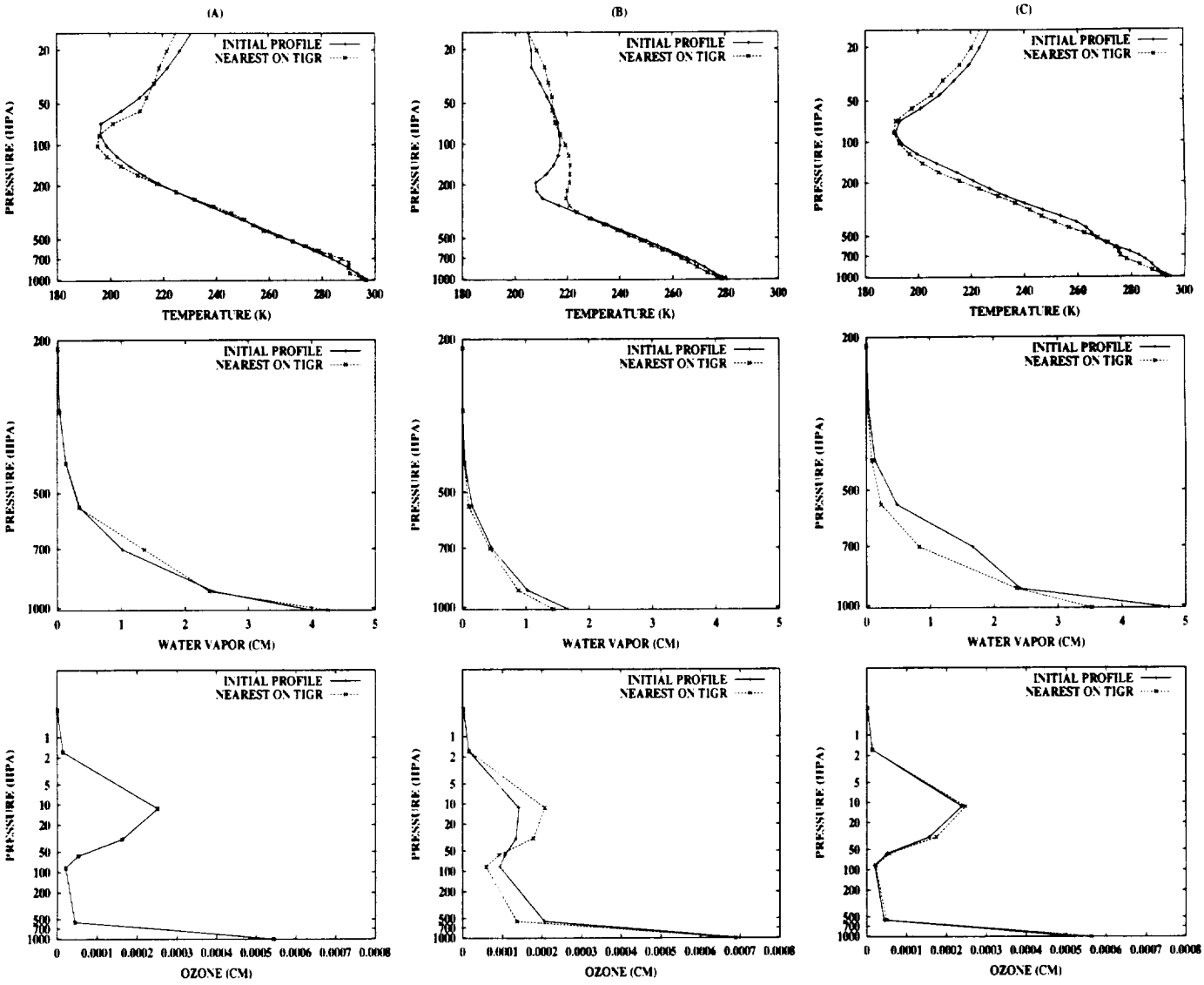

(A)
(B)
(C)
INITIAL PROFILE
NEAREST ON TIGR
PRESSURE (HPA)
TEMPERATURE (K)
WATER VAPOR (CM)
OZONE (CM)

Figure 10.

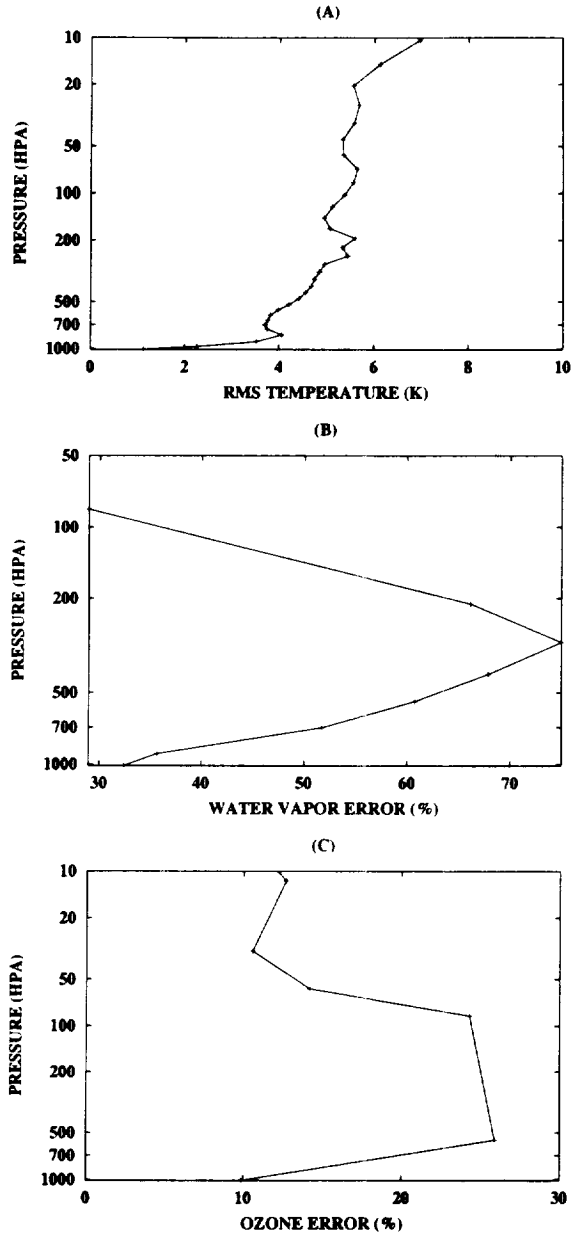

(A)
PRESSURE (HPA)
10
20
50
100
200
500
700
1000
0
2
4
6
8
10
RMS TEMPERATURE (K)
(B)
PRESSURE (HPA)
50
100
200
500
700
1000
30
40
50
60
70
WATER VAPOR ERROR (%)
(C)
PRESSURE (HPA)
10
20
50
100
200
500
700
1000
0
10
20
30
OZONE ERROR (%)

Figure 11.

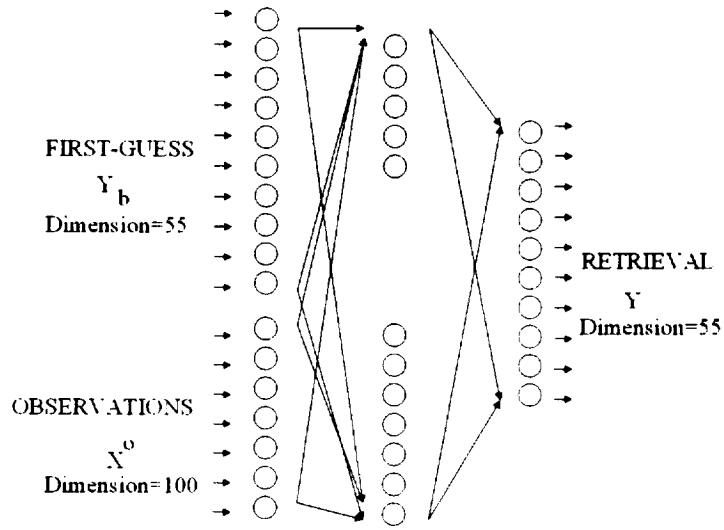

FIRST-GUESS
Y_b
Dimension=55
OBSERVATIONS
X^o
Dimension=100
RETRIEVAL
Y
Dimension=55

Figure 12.

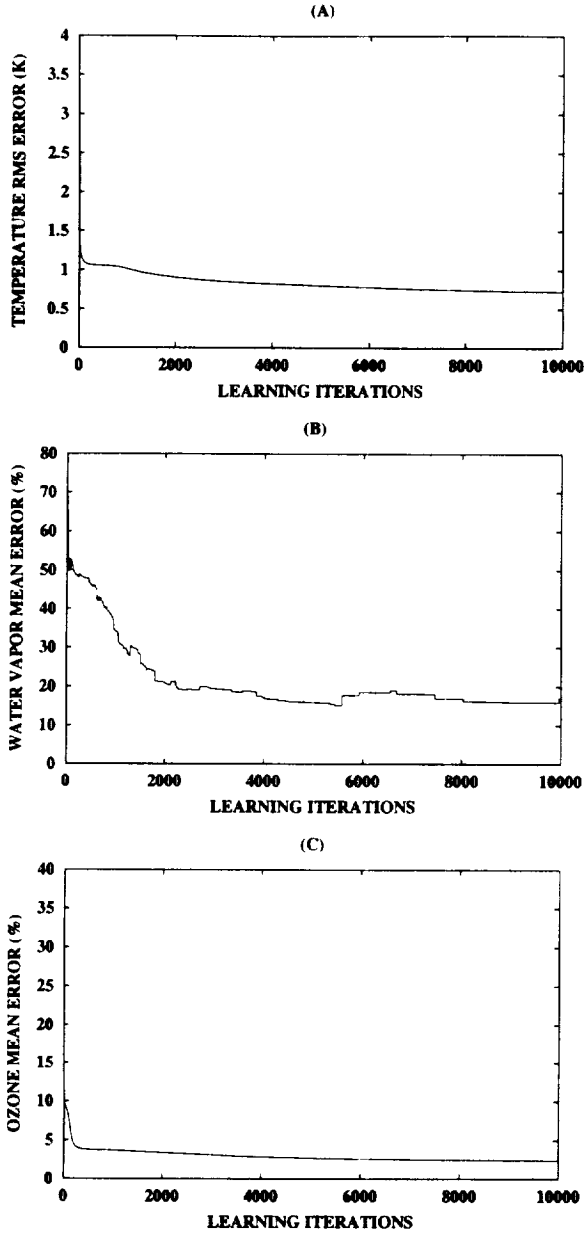

(A)
TEMPERATURE RMS ERROR (K)
LEARNING ITERATIONS
(B)
WATER VAPOR MEAN ERROR (%)
LEARNING ITERATIONS
(C)
OZONE MEAN ERROR (%)
LEARNING ITERATIONS

Figure 13.

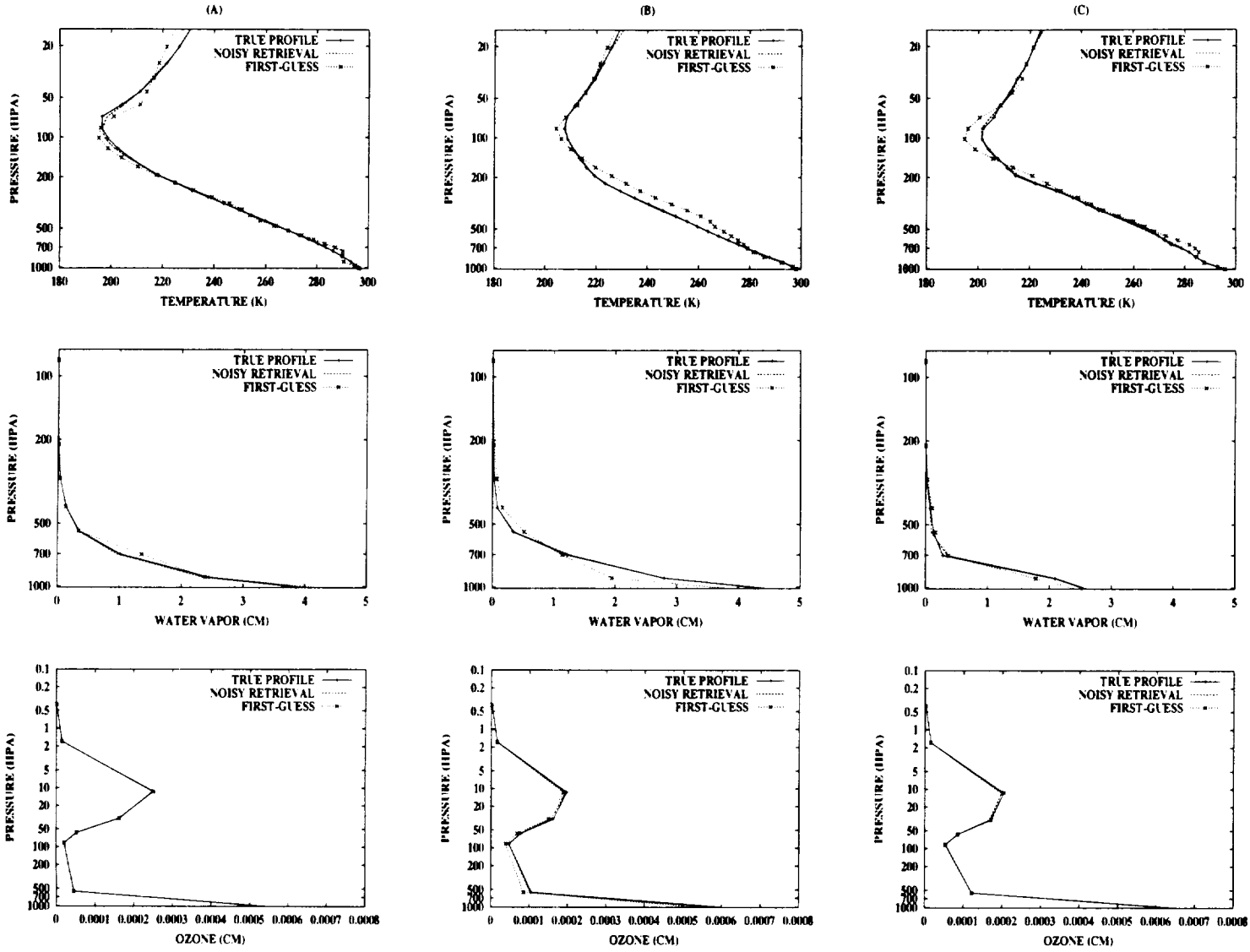

(A)
(B)
(C)
TRUE PROFILE
NOISY RETRIEVAL
FIRST-GUESS
PRESSURE (HPA)
TEMPERATURE (K)
WATER VAPOR (CM)
OZONE (CM)

Figure 14.

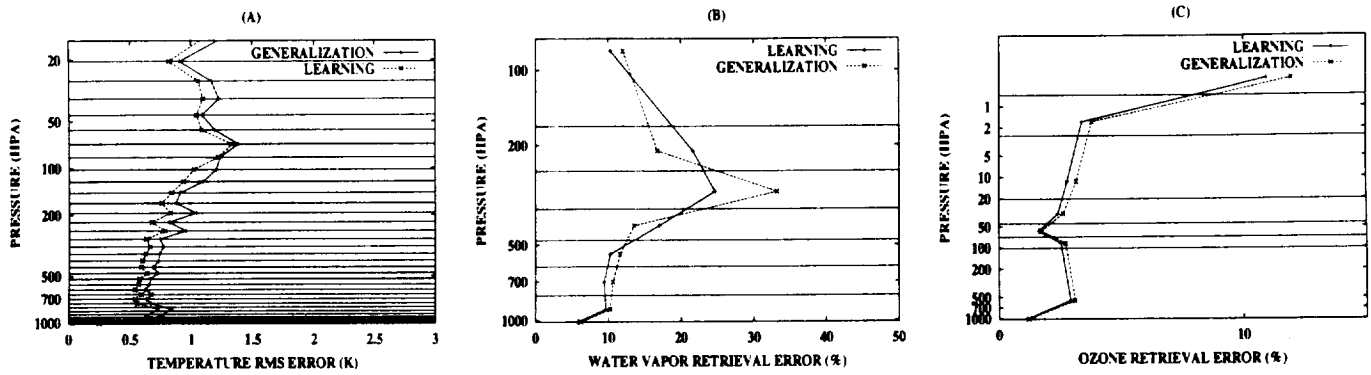

(A)
GENERALIZATION
LEARNING
PRESSURE (HPA)
TEMPERATURE RMS ERROR (K)
(B)
LEARNING
GENERALIZATION
PRESSURE (HPA)
WATER VAPOR RETRIEVAL ERROR (%)
(C)
LEARNING
GENERALIZATION
PRESSURE (HPA)
OZONE RETRIEVAL ERROR (%)

Figure 15.

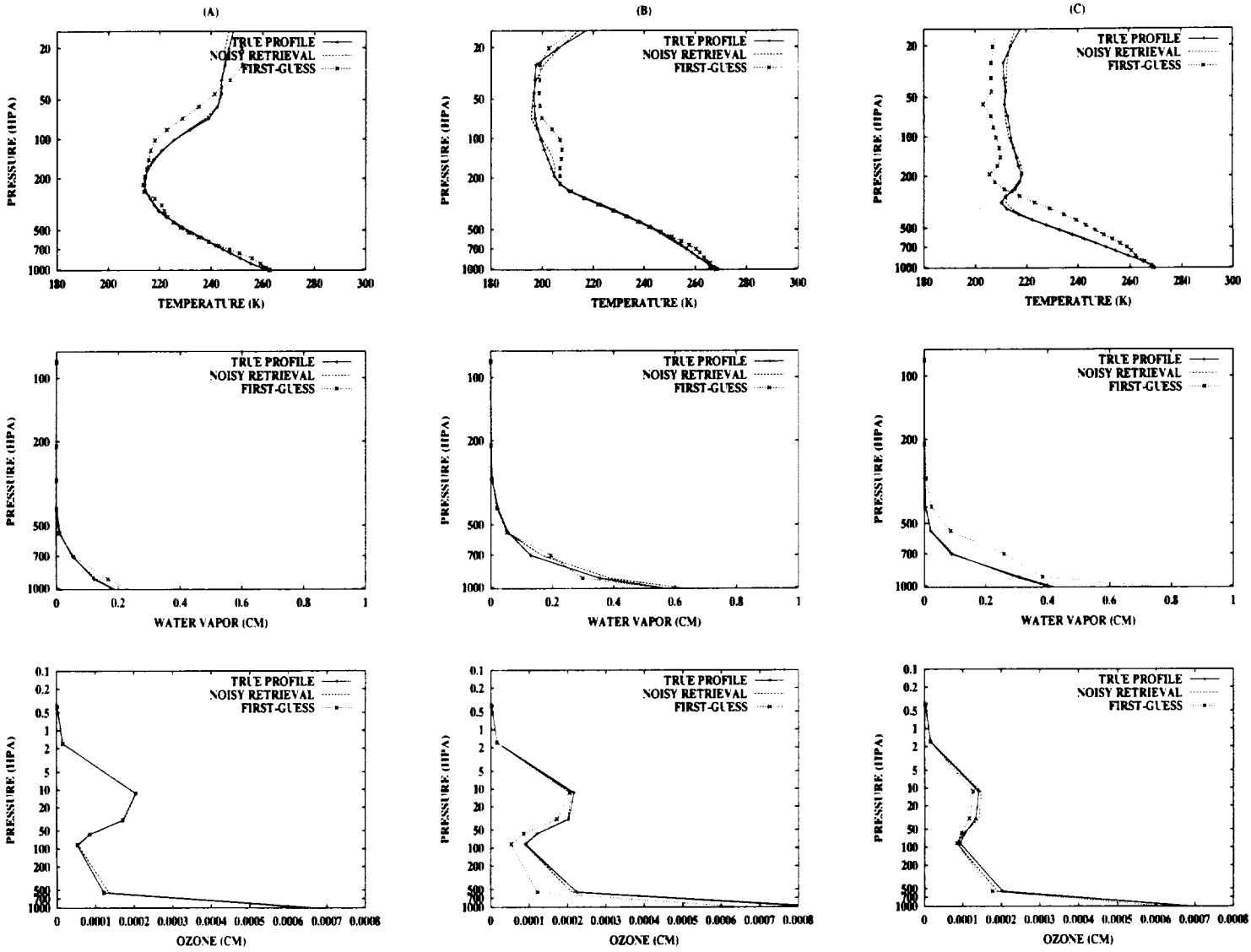

(A)
(B)
(C)
TRUE PROFILE
NOISY RETRIEVAL
FIRST-GUESS
PRESSURE (HPA)
TEMPERATURE (K)
WATER VAPOR (CM)
OZONE (CM)

Figure 16.

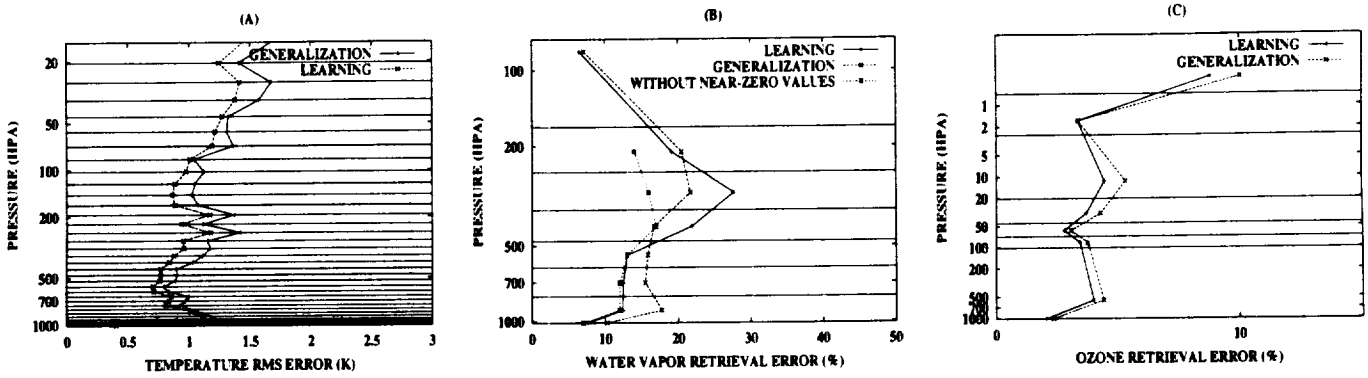

(A)
GENERALIZATION
LEARNING
PRESSURE (HPA)
20
50
100
200
500
700
1000
0
0.5
1
1.5
2
2.5
3
TEMPERATURE RMS ERROR (K)
(B)
LEARNING
GENERALIZATION
WITHOUT NEAR-ZERO VALUES
PRESSURE (HPA)
100
200
500
700
1000
0
10
20
30
40
50
WATER VAPOR RETRIEVAL ERROR (%)
(C)
LEARNING
GENERALIZATION
PRESSURE (HPA)
1
2
5
10
20
50
100
200
500
700
1000
0
10
OZONE RETRIEVAL ERROR (%)

Figure 17.