# Remote sensing from the infrared atmospheric sounding interferometer instrument
# 2. Simultaneous retrieval of temperature, water vapor, and ozone atmospheric profiles

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

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

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

Received 26 December 2001; revised 4 February 2002; accepted 5 February 2002; published 19 November 2002.

[1] A fast algorithm is developed to retrieve temperature, water vapor, and ozone atmospheric profile from the high spectral resolution Infrared Atmospheric Sounding Interferometer spaceborne instrument. Compression, denoising, and pattern recognition algorithms have been developed in a companion paper [*Aires et al.*, 2002b]. A principal component analysis neural network using this a guess information is developed here to retrieve simultaneously temperature, water vapor and ozone atmospheric profiles. The performance of the resulting fast and accurate inverse model is evaluated with a climatological data set including rare events: temperature is retrieved with an error $\leq$1 K, and total amount of water vapor has a mean percentage error between 5 and 7%. Atmospheric water vapor layers are retrieved with an error between 10 and 15% most of the time. The statistics of the ozone retrieval are too optimistic due to a lack of representation of ozone variability in our test data set. *INDEX TERMS:* 0399 Atmospheric Composition and Structure: General or miscellaneous; 1640 Global Change: Remote sensing; 3260 Mathematical Geophysics: Inverse theory; 3337 Meteorology and Atmospheric Dynamics: Numerical modeling and data assimilation; 3360 Meteorology and Atmospheric Dynamics: Remote sensing; *KEYWORDS:* IASI, infrared interferometer, neural networks, principal component analysis

**Citation:** Aires, F., W. B. Rossow, N. A. Scott, and A. Chedin, 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, doi:10.1029/2001JD001591, 2002.

## 1. Introduction

[2] 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 (METOP-1) satellite (see *Aires et al.* [2002b] for a more detailed description of IASI). The dimension (number of measurements per field-of-view) of IASI observations is much higher than for previous instruments: 8461 channels compared to 19 for high-resolution infrared radiometer sounder (HIRS) on TIROS-N operational vertical sounder. This is a major problem in the definition of retrieval algorithms. To deal with this high-dimension problem, various techniques to select channels in the IASI spectrum have been developed: See *Rabier et al.* [2002] or *Aires et al.* [2002a].


0148-0227/02/2001JD001591$09.00

[3] We have developed a method based on principal component analysis (PCA) for compressing, denoising, and first-guess retrieval for IASI in the *Aires et al.* [2002b] paper. Our approach, similar to the analysis of *Huang and Antonelli* [2001], allows for a more complete exploitation of all channels in the IASI spectra. The compression step allows to reduction of the dimension of the data used. The denoising process, using the redundancy information among channels, reduces considerably the instrument noise in IASI observations. The instrumental noise in the overall IASI spectrum goes from 0.9K to 0.2K after denoising. The PCA representation of the IASI spectra allows also for a fast and multivariate first-guess retrieval. This information is important for the development of inversion approaches.

[4] In this study, we are interested in a nonlinear 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 inverse radiative transfer equation, 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 multilayer 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. Furthermore, the neural network inversion method provides 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. This type of inversion scheme is called “global” inversion, as opposed to “local” inversion.

[5] 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 abilities of the neural network model [*Prigent et al.*, 2001a, 2001b], 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. In the neural network, the requirement takes the form of good sampling for the learning data set (this has to be done anyway in the variational assimilation to estimate the covariance matrices). But once this sampling is done, no choice about how many covariance matrices to estimate, and where, needs to be done. All the samples are used in the learning stage.

[6] However, for ill-conditioned problems, the use of a first-guess estimate and associated error covariance matrix is essential for elaborate 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.

[7] In a previous study [*Aires et al.*, 2002a] the authors have used a neural network approach for the IASI retrieval of surface temperature and atmospheric temperature profile. This retrieval scheme was regularized by introducing a priori information into the neural network scheme in many ways. But only a part of IASI information was used, using a channel selection based on the physical Jacobians, and no first-guess information was used.

[8] 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. We suppose in this study that the IASI spectra have been selected to be cloud-free (see section comments for a discussion on our future cloud scheme) [*Schlussel and Goldberg*, 2001]. 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 information content estimates are dependent on some theoretical assumptions: Gaussian distributed quantities, independence between first guess and observation, first-guess error covariance matrices sometimes not fully determined (i.e., no correlations between the first-guess errors of different variables), etc. Our neural network model is parameterized and tested without these assumptions and 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*, 1993; *Chevallier et al.*, 1998]. using RTIASI (Radiative Transfer model for IASI) [*Matricardi and Saunders*, 1999]. Even if this study concerns specifically the IASI instrument, the algorithms developped here can easily be adapted to other instruments, in particular such as the atmospheric infrared sounder (AIRS) on board the Aqua spacecraft.

[9] This paper is organized as follows: The retrieval algorithm based on a first-guess-based PCA-neural network approach is presented in section 2. Temperature, water vapor and ozone atmospheric profiles retrieval results are presented in section 3. Section 4 concludes this study with some perspectives on this work.

## 2. IASI Retrieval Method

[10] 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 whereas hierarchical schemes may introduce inconsistencies. The model developed here uses a nonlinear regression of the inverse RTM in the atmosphere obtained from a MLP neural network.

### 2.1. Neural Network Model

[11] Part of the neural network scheme developed in the next two sections is described in more detail by *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 $Y = (y_i; i \in S_0)$ of the mapping. The last layer $S_L$ represents the output mapping $X = (x_k; k \in S_L)$. The intermediate layers $S_m$ $(0 < m < L)$ are called the “hidden layers”. These layers are connected via neural links. We denote by $W$ the parameters of these links. It has been demonstrated [*Hornik et al.*, 1989; *Cybenko*, 1989] that any continuous function can be represented by a one-hidden-layer MLP.

[12] A convenient feature of the neural network model is that the analytical Jacobians (i.e., first derivatives of each output with respect to each input of the model) can be easily computed. These quantities allow for an a posteriori analysis of the fitted model [*Aires et al.*, 1999, 2001].

[13] 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 squared error in network outputs

$$C_1(W) = \frac{1}{2}\sum_{k\in S_2}\int\int D_E(\hat{x}_k(Y;W),x_k)^2\, P(Y,x_k)dx_k dY, \quad (1)$$

where $D_E$ is the Euclidean distance between $x_k$, the $k$th desired output component, and $\hat{x}_k$, the $k$th neural network output component, and $S_2$ is the output layer of the neural network.

[14] In practice, the probability distribution function, $P(Y, x_k)$, is sampled in a data set $\mathcal{B} = \{(Y^e, x_k^e), e = 1,\ldots, E\}$ of $E$ input/output couples, and $C_1(W)$ is then approximated by the classical least squares criterion:

$$\tilde{C}_1(W) = \frac{1}{2E}\sum_{e=1}^{E}\sum_{k\in S_2} D_E\left(\hat{x}_k(Y^e;W),x_k^e\right)^2. \quad (2)$$

[15] 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. Traditional gradient descent algorithms use all the samples of the data set $\mathcal{B}$ to compute a mean Jacobian of the criterion $\tilde{C}_1(W)$ in equation (2). These algorithms are called deterministic gradient descent. The major inconvenience of this approach is that the descent can be trapped in local minima. In the present application, a stochastic gradient descent algorithm is adopted: It uses the gradient descent formula iteratively for a unique random sample in the data set. With some constraints not discussed here, the stochastic character of this optimization algorithm theoretically allows the optimization technique to reach the global minimum of the criterion instead of a local minimum [*Duflo*, 1996].

[16] The more neurons in the hidden layer, the better is the fit to the learning data set. But the learning fit error is not a good criterion to constrain the neural architecture because too many neurons produces the overfitting problem: the network fits the learning data set very well, but is bad for generalization (i.e., the fit error on an independent data set of observations is large). For too few 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 too many neurons, the complexity of the neural network is too rich compared to the desired model and the overfitting problem appear (i.e., variance error). This dilemma is called the bias/variance dilemma [*Geman et al.*, 1992]. Thus the number of neurons in the hidden layer can be estimated by a heuristic procedure that monitors the generalization fit errors of the neural network as the configuration is varied: we vary the number of neurons in the hidden layer until the smallest generalization error is found.

### 2.2. Learning Algorithm With First Guess

[17] 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.

[18] 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:

$$\hat{x} = g_W\left(x^b, y^o\right), \quad (3)$$

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

[19] The learning algorithm consists of estimating the parameters $W$ of the neural network that minimize 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:

$$C_2(W) = \frac{1}{2}\int\int\int D_E\left(g_W\left(x^b,y^o\right),x\right)^2 P\left(x,y^o,x^b\right) \quad (4)$$

$$= \frac{1}{2}\int\int\int D_E(g_W(x+\varepsilon,y+\eta),x)^2 P(x)P_\eta(\eta)P_\varepsilon(\varepsilon), \quad (5)$$

where $P(x)$ is the probability distribution function of the physical variables $x$ that depends on the natural variability. $P_\eta(\eta)$ is the probability distribution function of the observation noise $\eta$. $P_\varepsilon(\varepsilon)$ is the probability distribution function of the first-guess error $\varepsilon = x^b - x$. This noise is presently simulated for IASI as white Gaussian noise, but it is important to see that we can use any model for the noise as long as we simulate this noise model during the learning stage. Using noisy data in the inputs of the neural network during the learning process is called "Input Perturbation" and it is a powerful regularization technique [*Aires et al.*, 2002a]. The Input Perturbation constrains the solution to be smooth, suppressing degrees of freedom. One can also see that the use of different noise simulations each time that we use a sample during the learning is a way of increasing the number of samples in the learning data set: from 5000 samples to 5000 times the number of iterations (typically thousands) during the learning stage.

[20] As explained by *Aires et al.* [2001], the quality criterion in equation (4) is very similar to the quality criterion used in variational assimilation. One of the main differences is that the neural network criterion in equation (4) involves the distribution $P(x)$. 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 $P(x)$ 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).

[21] To minimize the criterion of equation (4), we create a data set $\mathcal{B} = \{(x^e, y^{oe}, x^{b^e}); e = 1,\ldots, E\}$ that samples as well as possible all the probability distribution functions in (4). Then, the practical criterion used during the learning stage is given by:

$$\tilde{C}_2(W) = \frac{1}{2E}\sum_{e=1}^{E} D_E\left(g_W\left(x^{b^e}, y^{o^e}\right), x^e\right)^2. \quad (6)$$

[22] First, to sample the probability distribution function, $P(x)$, we select geophysical states ($x^e$) that cover all natural combinations and their correlations and by calculating $y^e = RTM(x^e)$ 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 $y$ and $x$ (i.e., *empirical* inversion).

[23] The quality of the resulting data set is a prerequisite for a good retrieval method: the sampling should be sufficiently dense so that interpolation between the samples by the neural network is accurate, and the sampling should represent any case that can occur in operational conditions so that the neural network does not need to extrapolate the behavior in the training data set to outlying cases.

[24] For sampling $P_\eta$, 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 $\eta$ (see section 2.1 of companion paper, same issue). To sample the first-guess variability with respect to state $x$ (i.e., sampling $P(x^b|x)$), we use a first-guess data set $\{x^{b^e}; e = 1,\ldots, E\}$: this data set can be a climatological data set or a 6-hour prediction (which would have better error statistics, 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.

[25] The factorization between equations (4) and (5) uses the hypothesis that the geophysical parameters $x$, the first-guess error $\varepsilon$ and the instrument noise $\eta$ are independent. If this is not the case, it is still possible to use the criterion in equation (4) instead of using the easier one in equation (5). The principle consists of introducing this dependency into the learning data set by using coupled atmospheric states $x$ and first-guess $x^b$. The dependency structure used during the learning stage should reflect the dependency structure expected during the operational use of the neural network retrieval. We will see in the next section that this is the strategy adopted in this work. This possibility of introducing a complex structure of dependencies between the observations and the first-guess errors is another illustration of the flexibility of the neural network approach.

[26] What is the behavior of the neural retrieval when the first-guess is far from the actual solution? In principle, the nonlinearity of the neural network allows it to have different weights on the observations and first-guess information, depending on the situation. For example, if first guesses are better in tropical cases than in polar cases, the neural network would have inferred this behavior during the learning stage, and then will give less emphasis to the first guess when a polar situation is analyzed. This supposes, once again, that the training data set has been correctly sampled. We are presently working on new tools to diagnose bad cases (bad first guess, incoherent measurement, situations not contained in the training data set, uncertainties of the neural network on the possible retrievals, etc.). These tools would take the form of a posteriori probability distributions for the neural network retrieval that can be used to define confidence intervals on the retrieved quantities.

## 2.3. General PCA Regression

[27] As for the PCA-based pattern recognition, the practical benefits of using PCA components, $h$, instead of the raw IASI spectra, $y$, are that the method is faster because of the significant input dimension reduction and uses observations with a reduced noise level. Furthermore, the learning stage is faster since the network has less inputs and less parameters to estimate.

[28] The fact that the dimension of inputs is reduced decreases also the number of parameters in the regression model (i.e., weights in the neural network), and consequently decreases the number of degrees of freedom in the model, which is good for all regression techniques. The variance in determining the actual values of the neural weights is reduced, which also improves the retrieval. The combination of PCA and neural network has been used, for example, by *Huang* [1999] where a generalized Hebbian algorithm (i.e., a different species of neural network architecture than the MLP used in this study) has been used to perform a classification of seismic data.

[29] The quality criterion in equation (6) is simpler because the inputs are decorrelated. Correlated inputs in a regression are called multicollinearities and they are well-known to disturb the model fit [*Gelman et al.*, 1995]. Suppressing these multicollinearities makes the minimization of the quality criterion more efficient: it is easier to minimize, with less probability to become trapped in a local minimum. It has the general effect of suppressing variance in the determination of the parameters of the model and, consequently, reducing the uncertainties in the retrievals. For a more detailed description of PCA-based regression, the reader is referred to *Jolliffe* [2002].

[30] How many PCA components should the regression use? We have seen in section 4.2 of *Aires et al.* [2002b] that the optimum compromise between the best spectral fit and denoising in terms of global statistics for the whole IASI spectrum is 30 PCA components. For some spectral regions, more PCA components would have obtained better denoising results, for others fewer components would have been better: the 30 PCA components is the best compromise for the overall IASI spectrum. Moreover, the retrieval technique, especially if it's a nonlinear method, can use nontrivial information in the higher-order components. For example, *Huang and Antonelli* [2001] found 150 components as the optimal number of components for the denoising, but 250 components were optimal for the retrieval process. No theoretical results exist to define the optimal PCA number for regression, it entirely depends on the problem to be solved. Various tests have to be performed. Our experience with the neural network technique has shown that if the problem is sufficiently well regularized, once the available information has been provided to the neural inputs, adding more PCA components does not have a big impact on the

retrieval results, it is just more practically difficult because of the increase in the data dimension. We then recommend being conservative and taking more PCA components than the denoising optimum would indicate.

### 2.4. Weighting in the Quality Criterion

[31] The inputs and outputs of the network are not homogeneous; that is, different types of variables have different dynamic ranges. As we will see in the following, solving this problem necessitates diagnosing the learning step and controlling correctly the system, in contradiction to the black-box conception often associated with neural networks. The range of values, which is different for temperature and 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 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 dry 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

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

instead of the absolute RMS error

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

in equation (6). 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 error instead of an absolute error for the water vapor and the ozone values. The counterpart of this new criterion is that the percentage error could be exaggerated for values very close to zero. We will comment on this effect during the presentation of the results.

[32] The atmospheric temperature is described by 39 output values (i.e., 39 atmospheric levels) in the neural network where water vapor and ozone are each described by only 8 values (i.e., 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.

## 3. Results for the Retrieval of Temperature, Water Vapor, and Ozone Atmospheric Profiles

[33] 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 (see section 3.2 in companion paper, same issue). The choice of dry or wet configuration can be made using the first guess.

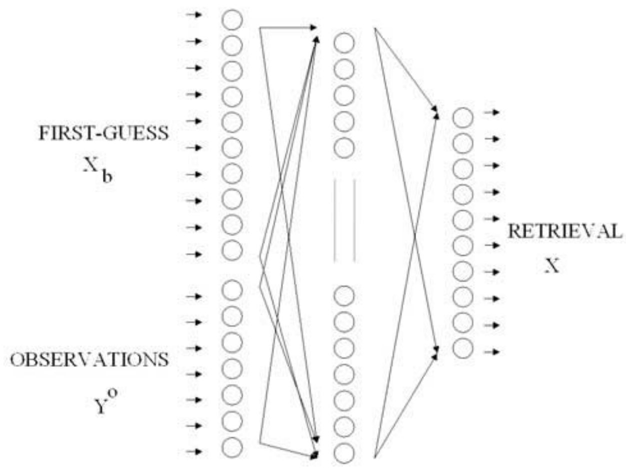



**Figure 1.** Architecture of a multilayer perceptron neural network with first guess a priori information: $x^b$ is the climatological first guess, $y^o$ is the Infrared Atmospheric Sounding Interferometer observation (brightness temperature spectrum compressed and denoised by PCA), and $x$ is the neural network retrieval.

### 3.1. Wet Atmosphere Configuration

[34] 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 denoising) as inputs for the IASI observation part since the NN is able to use only the information that it needs for the desired retrieval: $y^0 = h$. It is possible that between the 30th and the 100th PCA components, there is information on a specific spectral region, not statistically important for the whole spectrum, that is useful for the NN retrieval (see section 2.3). *Huang and Antonelli* [2001], found 250 components to be optimal for the retrieval process, but 150 components were already needed for the denoising; these differences can be explained by differences in the instrument and its noise (section 4.2 of companion paper, same issue).

[35] The architecture of the network NN1 is an MLP (Figure 1) with 155 inputs coding the $M$ = 100 PCA components, $y^0 = h$, 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 (section 2.1): For too small a 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 too large number of neurons, 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, appears (i.e.,

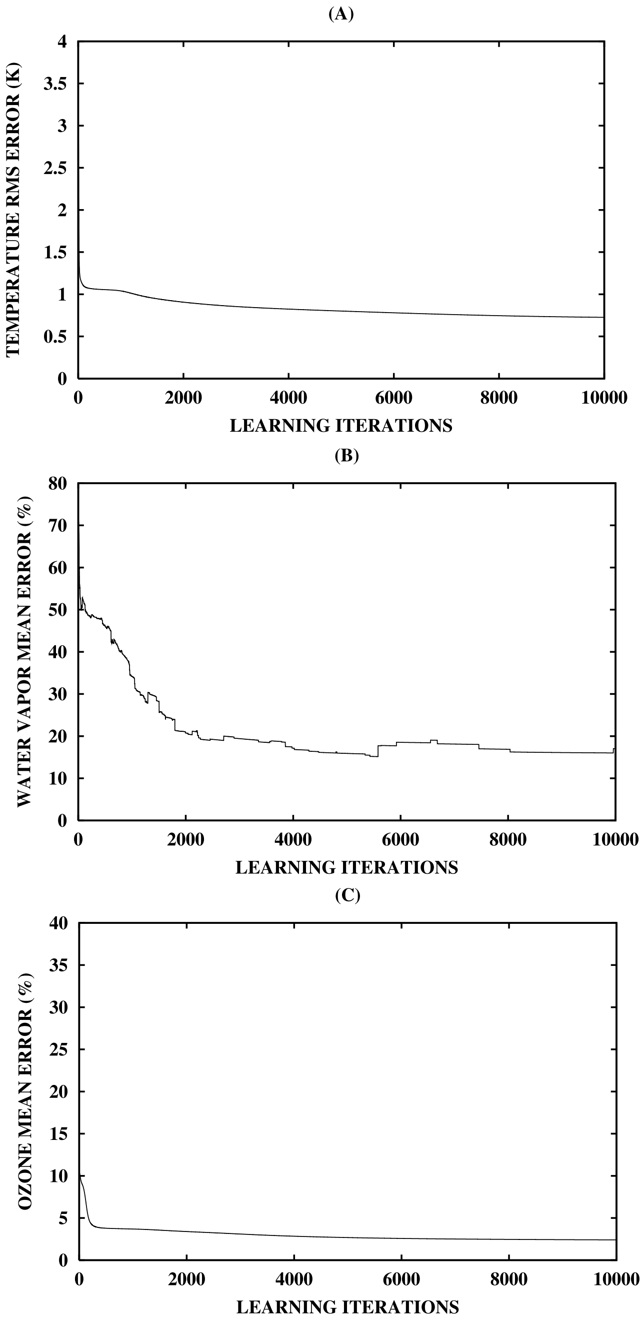



**Figure 2.** Learning curves for (a) temperature at 817 hPa, (b) water vapor between 358 and 478 hPa, and (c) ozone between 20 and 45 hPa.

variance error). In practice, we use a different number of neurons in the hidden layer and smaller generalization error indicates the best compromise.

[36] Figure 2 represents the learning and the testing curves of some of the retrieved quantities during the learning stage. The purpose of Figure 2 is to show how the inhomogeneity of the outputs in the neural network can be a problem. 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 problem, it is important to give uniform weight to each of the variables, this is the reason why we have modified our quality criterion as explained in section 2.4. 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 times, 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: indicated by 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.

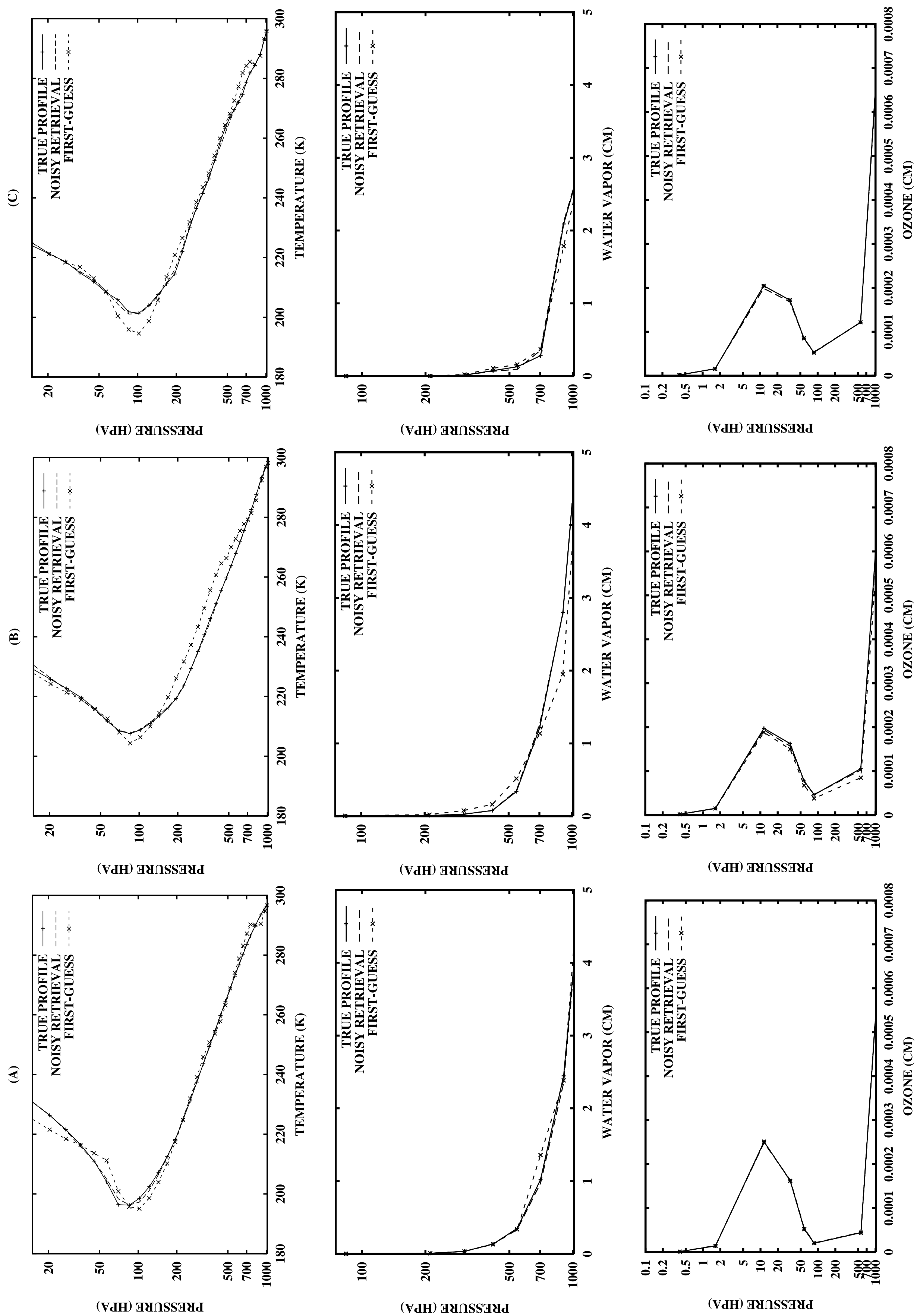
[37] Figure 3 presents three examples of retrievals. We see, in each case, a major improvement of the temperature profile retrieval over the first guess: the true profile and the noisy retrieval are difficult to distinguish in Figure 3. This is also true for the water vapor retrieval. The ozone is also very good, but the first guess was already very close to the correct solution. Consequently, the retrieval statistics for wet atmospheres, represented in Figure 4, 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 [*Aires*, 1999] that the fusion of the information from the advanced microwave sounding unit (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 peak 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 test data set.

### 3.2. Dry Atmosphere Configuration

[38] 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 1) is the same as NN1: 155 inputs, 80 neurons in the hidden layer, and 55 neurons in the output layer.

[39] Figure 5 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 possesses too strong an inversion (see profile C at level 100 hPa). Water vapor is well retrieved, a

**Figure 3.** (opposite) Columns A–C: three examples of (from top to bottom) 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.


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

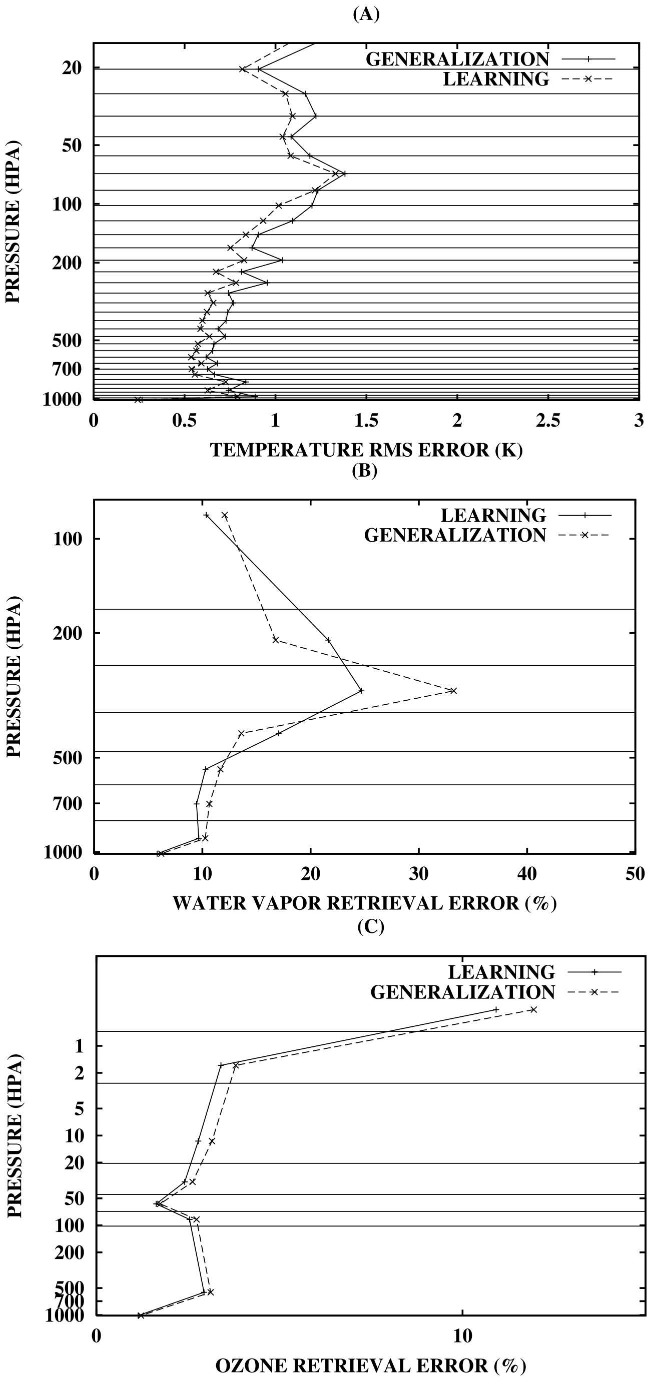



**Figure 4.** Error profile for the retrievals in the learning set (solid line) and in the generalization set (dashed line) for (a) temperature, (b) water vapor, (c) and ozone: wet atmospheres configuration. Near-surface values for water vapor and ozone represent the total vertical content.

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.

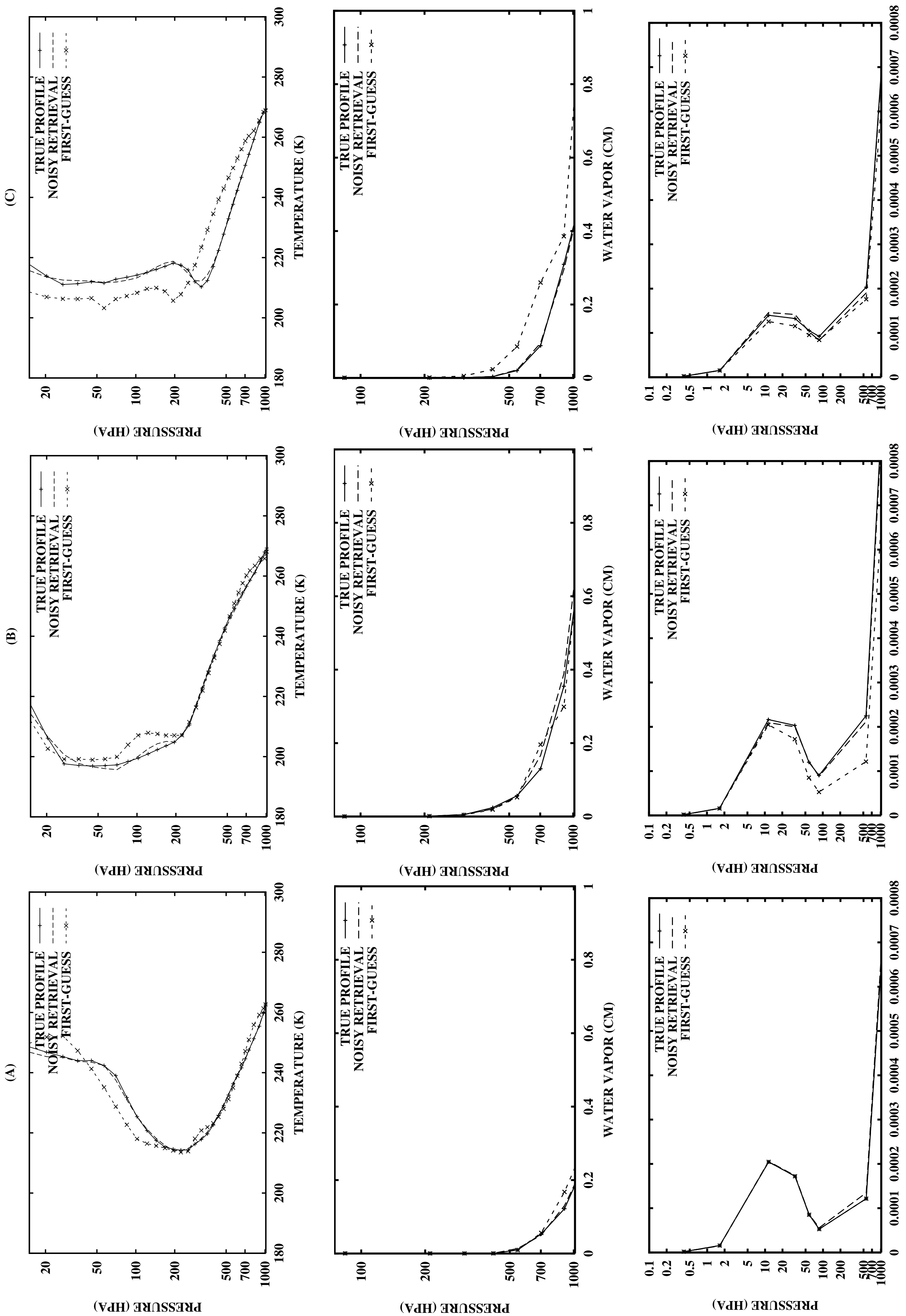
[40] Figure 6 shows the RMS retrieval errors for temperature, water vapor and ozone for the dry condition neural network. The retrieval of temperature is more difficult in dry condition than in wet conditions (Figure 4). The RMS error is still <1 K, except for near-surface levels, due to near-surface 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 6b 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.

### 3.3. Additional Comments

[41] There are four fields-of-view for each IASI sample, covering an area with a diameter of 9 to 12 km at nadir. Assuming homogeneous meteorological conditions, an average of the four pixel measures can be used to perform the retrievals: these four field-of-view provide redundant measurements that can be averaged to reduce noise. We have shown [*Aires et al.*, 2002a] that because IASI has thousands of channels, reducing the instrument noise by pixel averaging has little impact on the IASI retrieval. Furthermore, the atmospheric temperature, water vapor or ozone channels have a radiative transfer function-Jacobian that is vertically broad, which means that the information that they provide is ambiguous (i.e., limited vertical resolution). It is then normal that the retrieval of these atmospheric profile is more sensitive to fit error and less to instrument noise error. This was true without denoising techniques, and it is even more true after our denoising method has reduced the overall IASI noise from 0.9 to 0.2K. The regularization used to avoid noise effects in our approach, by the input perturbation method, is also sufficiently efficient that the reduction of noise by pixel-averaging has little impact on the quality of the retrievals. This means that our method is able to provide good results for each pixel to maximize the horizontal resolution or to perform scene selection. So pixel-averaging is not recommended.

[42] The neural network scheme that we have developed in this study uses a direct radiative transfer model. This sort of inversion approaches is often referred to as "physical" inversion in contrast to an algorithm that uses a data set of colocated and coincident in situ measurements, called an "empirical" inversion. Both approaches can be used with the neural network. The later cannot be generalized outside the observed variability of the initial data set, whereas the use of a physical direct model allows adaptative improve-

**Figure 5.** (opposite) Columns A–C: three examples of (from top to bottom) 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.


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

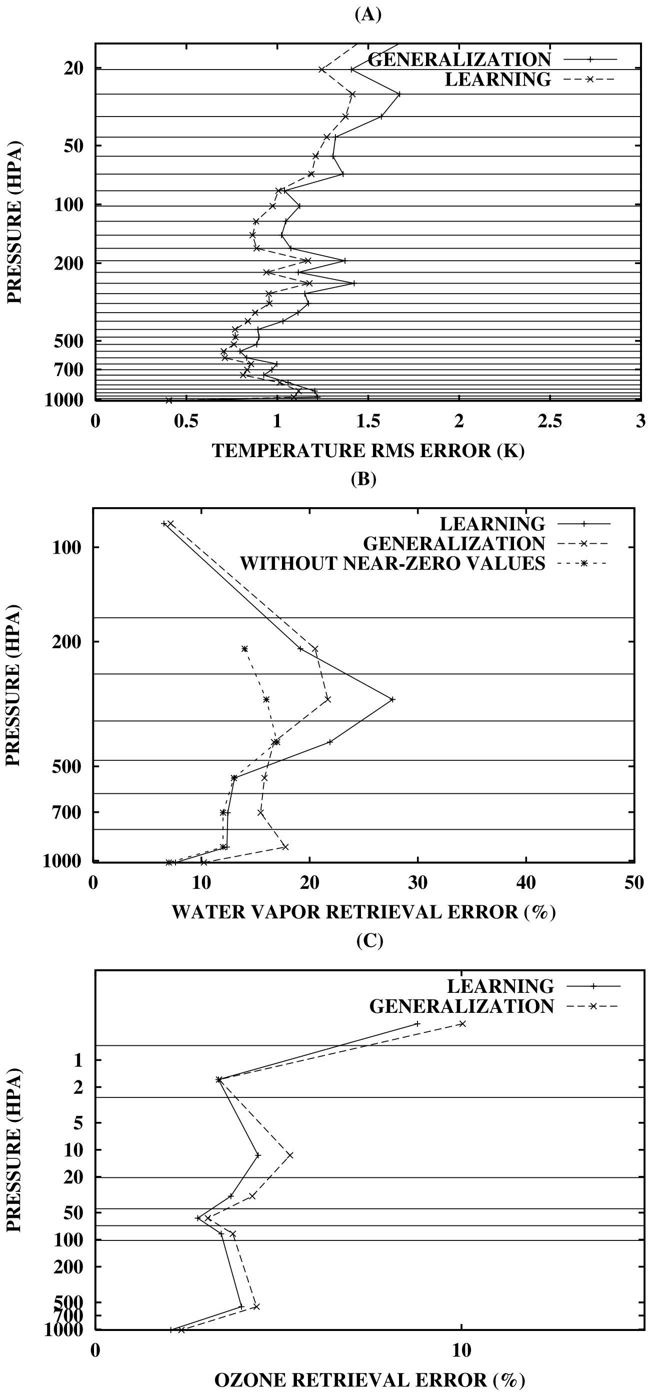



**Figure 6.** Error profile for retrievals in the learning set (solid line) and in the generalization set (dashed line) for (a) temperature, (b) water vapor, and (c) ozone: Dry atmospheres configuration. Near-surface values for water vapor and ozone represent the total vertical content.

ment based on the representativity of the data set. The drawback of the physical inversion scheme is that the relationship between observations and geophysical parameters is dependent on the direct model errors. But if these errors have been characterized by a validation process, it is possible to use these characteristics of model error during the learning. This would be done by adding a simulated model error to the brightness temperatures during the training stage. This is similar to the application of instrumental noise to brightness temperatures during the training (section 2.2). The similarity in the treatment of these two errors is close to what is done in variational assimilation where the covariance matrix of the direct model and the instrument noise are simply added.

[43] A major concern for any IASI retrieval scheme is how to handle cloudy cases. The traditional approach is to perform cloud clearing: cloudy situations have to be detected in a first step, and then, information from a microwave instrument (i.e., AMSU in the METOP case) is used to remove in the IASI infrared observations the effects of the clouds. Cloud-free observations are then used by an inversion algorithm. In this work, we have supposed that the IASI observation were previously cloud-cleared. However, there is another possibility that needs to be investigated. We could use the ability of the neural network methods for fusion of information from different instruments. Using infrared and microwave information together in a single analysis might allow the retrieval of cloud properties along with the temperature, water vapor and ozone. Since the cloud parameters have different effects on the infrared and microwave spectra it should be possible to infer these cloud parameters from the differences of the two instruments. For that purpose, we need a good first guess for clouds such as the International Satellite Cloud Climatology Project (ISCCP) data\-base [*Rossow and Schiffer*, 1999].

[44] The simultaneous retrieval of many variables is an important aspect of this work. As it has been shown, for example, by *Krasnopolsky et al.* [1996, 1999], retrieving more than one variable allows the use of the correlation structure between these variables. This constrains better the inverse problem which becomes less ambiguous. *Aires et al.* [2001] retrieved the surface temperature and surface emissivities using the special sensor microwave imager (SSM/I) and the ISCCP cloud data. This retrieval is difficult and ambiguous since their variability is mixed in the observations. In this case, the retrieval of both quantities is necessary to reduce these ambiguities, but first-guess information is also needed. We can use this same approach for the retrieval of surface temperature and emissivities for IASI.

## 4. Conclusion and Perspectives

[45] Together with a PCA-based method for compressing, denoising, and first-guess retrieval that has been described in a companion paper (same issue), we have developed a PCA-neural network retrieval scheme which uses first-guess information. Such a neural network approach has several similarities to the variational assimilation technique. Although their practical implementation can appear very different, *Aires et al.* [2001] has shown that the two techniques share very basic aspects. They are both statistical inversion methods that minimize a quality criterion, using a priori first-guess information and a radiative transfer (physical) model.

[46] The three major advances of this work over *Aires et al.* [2002a] on IASI are: the PCA-based compression, denoising, and first-guess retrieval that decrease instrument noise effects using redundancy information and reduce

also the dimension of the neural network, which allows the faster retrieval to be applied at higher spatial resolution; the introduction of a first-guess information that adds more information than only the IASI observations, and that constrains better the inversion process, improving retrieval results; and the simultaneous retrieval of three atmospheric variables (temperature, water vapor and ozone). This 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. *Krasnopolsky et al.* [2000] have shown that the multivariate aspect of the inversion scheme reduces the variance in the retrievals. All these three improvements had a considerable impact on the obtained results. They are also an important step for defining an operational inversion scheme for IASI in realistic conditions, providing many ways of improving the algorithm in the future.

[47] 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 to our method. The retrieval errors are small: temperature is retrieved with an error $\leq$1 K, total amount of water vapor has a mean percentage error between 5 and 7%. Atmospheric water vapor layers are retrieved with error between 10 and 15% most of the time. Statistics of ozone retrieval are too optimistic due to a lack of representation of ozone variability of our data set.

[48] 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 even with complex atmospheric situations, the potentialities of the IASI instrument allow achieving the World Meteorological Organization specifications. 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 to the retrieval scheme, it is robust to noise, and it is very fast and accurate. This new scheme is then an excellent candidate for the processing of IASI observations.

[49] There are various perspectives on this work. We have shown previously [*Aires et al.*, 2002a] that the specialization of the neural inversion scheme improves the results. An advantage of our approach is that it can easily accommodate nonlinear relationships between the information from other instruments [*Prigent et al.*, 2001a, 2001b]. This is a particularly interesting feature since IASI results for high-altitude atmospheric temperature are expected to be improved by AMSU [*Aires*, 1999]. This should be beneficial to the retrieval results because more information is added, and also because the structure of correlation between the different instruments constrains better the inverse problem, reducing the uncertainties. 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]. We plan also to develop the same chain of algorithms for the AIRS instrument on the Aqua platform.

[50] It would be highly interesting to combine this neural network retrieval with the variational approach: this could be done by using the neural network retrieval as an independent first guess for the variational assimilation model, but it could also be done assimilating the neural network retrieved products instead of the IASI observations.

## Notation

$x$ vector of physical variables to retrieve.
$\hat{x}$ estimate of $x$.
$x^b$ first guess a priori information for $y$.
$\varepsilon$ $= x^b - x$, first-guess error.
$RTM(x)$ radiative transfer model for the physical variables $x$ (also a vector).
$y^o$ IASI brightness temperature spectrum observations.
$\eta$ IASI instrumental noise.
$E$ number of samples in the data set.
$h$ PCA compression of the IASI spectrum $y$.
$\sigma$ 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 neural network.
$y_i$ neural network input value on neuron $i$.
$x_k$ neural network output value on neuron $k$.
$S$ number of neurons in a neural network layer.
$L$ number of hidden layers in the neural network.
$\mathcal{B}$ data set sampling the probability distribution functions.
$D$ generic distance.
$D_E$ Euclidean distance.
$D_M$ Mahalanobis distance.
$P$ generic probability measure.
$P_\eta(\eta)$ probability distribution function of $\eta$.
$P_\varepsilon(\varepsilon)$ probability distribution function of $\varepsilon$.
$C_1(W)$ theoretical quality criterion for classical neural network learning phase.
$\tilde{C}_1(W)$ practical quality criterion for classical neural network learning phase.
$C_2(W)$ theoretical quality criterion for classical neural network learning phase with first-guess information.
$\tilde{C}_2(W)$ practical quality criterion for classical neural network learning phase with first-guess information.

[51] **Acknowledgments.** We would like to thank the ECMWF for providing the direct radiative transfer model RTIASI. This work was partly supported by special NASA Climate and Radiation Program funding provided by Robert J. Curran and now by Donald Anderson. The authors thank three anonymous referees for their fruitful comments and suggestions.

## 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, Ph.D. thesis, Univ. Pierre et Marie Curie (Paris VI), Paris, 1991.

Aires, F., Problèmes inverses et réseaux de neurones: Application à l'interféromètre haute résolution IASI et à l'analyse de séries temporelles, Ph.D. thesis, Univ. Paris IX/Dauphine, Paris, 1999.

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., 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., 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. Geophys. Res.*, *106*(D14), 14,887–14,907, 2001.

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, *J. Appl. Meteorol.*, *41*, 144–159, 2002a.

Aires, F., A. Chédin, N. A. Scott, and W. B. Rossow, Remote sensing from the infrared atmospheric sounding interferometer instrument, 1, Compression, denoising, first-guess retrieval algorithms, *J. Geophys. Res.*, *107*(DX), 10.1029/2001JD000955, in press, 2002b.

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 Sens.*, *18*(15), 3135–3143, 1997.

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*(D8), 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. Clim. Appl. Meteorol.*, *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. Meteorol.*, *37*, 1385–1397, 1998.

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

Duflo, M., Algorithmes stochastiques, in *Mathématiques et Applications*, Springer-Verlag, New York, 1996.

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, Ph.D. thesis, Univ. Denis Diderot (Paris VII), Paris, 1993.

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.

Gelman, A. B., J. S. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, Chapman and Hall, New York, 1995.

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

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

Huang, H.-L., and P. Antonelli, Application of principal component analysis to high-resolution infrared measurement compression and retrieval, *J. Clim. Appl. Meteorol.*, *40*, 365–388, 2001.

Huang, K., Neural network for seismic principal components analysis, *IEEE Trans. Geosci. Remote Sens.*, *37*, 297–311, 1999.

Jolliffe, I. T., *Principal Component Analysis, Springer Ser. Stat.*, 2nd ed., Springer-Verlag, New York, 2002.

Krasnopolsky, V. M., G. H Gemmill, and L. C. Breaker, A new transfer function for SSM/I based on an expanded neural network architecture, *Tech. Note, OMB Contrib. 137*, Natl. Cent. for Environ. Prediction, Washington, D. C., 1996. (Available at http://polar.wwb.noaa.gov/omb/papers/tn137/tnombnn3.html)

Krasnopolsky, V. M., G. H. Gemmill, and L. C. Breaker, A multi-parameter empirical ocean algorithm for SSM/I retrievals, *Can. J. Remote Sens.*, *25*(5), 486–503, 1999.

Krasnopolsky, V. M., G. H. Gemmill, and L. C. Breaker, A neural network multiparameter algorithm for SSM/I ocean retrievals: Comparisons and 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.

Prigent, C., F. 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.*, *106*(D18), 20,665–20,685, 2001a.

Prigent, C., E. Matthews, F. Aires, and W. B. Rossow, Remote sensing of global wetland dynamics with multiple satellite data sets, *Geophys. Res. Lett.*, *28*(24), 4631–4634, 2001b.

Rabier, F., N. Fourrié, D. Chafä, and P. Prunet, Channel selection methods for infrared atmospheric sounding interferometer radiances, *Q. J. R. Meteorol. Soc.*, *128*, 1011–1027, 2002.

Rossow, W. B., and R. A. Schiffer, Advances in understanding clouds from ISCCP, *Bull. Am. Meteorol. Soc.*, *80*(11), 2261–2287, 1999.

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. 1, *Foundations*, edited by D. E. Rumelhart, J. L. McClelland, and the PDP Research Group, pp. 318–362, MIT Press, Cambridge, Mass., 1986.

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

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

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