A Regularized Neural Net Approach for Retrieval of Atmospheric and Surface Temperatures with the IASI Instrument

FILIPE AIRES

Department of Applied Physics, Columbia University, and NASA Goddard Institute for Space Studies, New York, New York

ALAIN CHÉDIN AND NOËLLE A. SCOTT

Laboratoire de Météorologie Dynamique, CNRS, École Polytechnique, Palaiseau, France

WILLIAM B. ROSSOW

NASA Goddard Institute for Space Studies, New York, New York

ABSTRACT

In this paper, a fast atmospheric and surface temperature retrieval algorithm is developed for the high-resolution Infrared Atmospheric Sounding Interferometer (IASI) spaceborne instrument. This algorithm is constructed on the basis of a neural network technique that has been regularized by introduction of information about the solution of the problem that is in addition to the information contained in the problem (a priori information). The performance of the resulting fast and accurate inverse radiative transfer model is presented for a large diversified dataset of radiosonde atmospheres that includes rare events. Two configurations are considered: a tropical-airmass specialized scheme and an all-airmasses scheme. The surface temperature for tropical situations yields an rms error of 0.4 K for instantaneous retrievals. Results for atmospheric temperature profile retrievals are close to the specifications of the World Meteorological Organization, namely, 1-K rms error for the instantaneous temperature retrieval with 1-km vertical resolution.

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 the Exploitation of Meteorological Satellites (EUMETSAT) and the European Space Agency 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 and 2 km, respectively. IASI, jointly developed by the Centre National d'Études Spatiales (CNES) and EUMETSAT, provides spectral coverage from 3.5 to 15.5 μm at considerably higher spectral resolution than that of 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 goal of this study is to present an inversion algorithm that retrieves geophysical variables from IASI measurements. We are confronted, in this work, with problems related to the ill-posed character of the inverse problem, the sensitivity to noise and, specific to IASI, the data dimension. The multilayer perceptron (MLP) technique is particularly interesting to solve this kind of problem. Such an approach has already been developed by the Atmospheric Radiation Analysis (ARA) group of Laboratoire de Météorologie Dynamique for HIRS coupled with the Microwave Sounding Unit (MSU; Escobar et al. 1993), for the Special Sensor Microwave Temperature Sounder and Water Vapor Profiler (SSM/T-1) instrument on the Defense Meteorological Satellite Program satellites (Rieu et al. 1996), and even for the high-resolution infrared spectrometer known as the Advanced Infrared Radiation Sounder (AIRS) of the National Aeronautics and Space Administration for the coming Earth Observation System Aqua satellite (Escobar et al. 1993) or for the IASI instrument (Aires et al. 1998). The great advantages of MLP are the rapidity, the small amount of memory required, and accuracy of

Corresponding author address: Dr. Filipe Aires, Dept. of Applied Physics, Columbia University, NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025.
E-mail: faires@giss.nasa.gov

results (Aires 1999). The MLP model is nonlinear, which is a crucial point for the regression fit to the inverse radiative transfer equation (RTE). Furthermore, assumptions such as the linearity of the RTE or the Gaussian assumption for stochastic variables are not required for MLP.

In this paper, it is demonstrated that the inversion procedure may be regularized by introducing various kinds of outside information about the physical problem to the neural method. This approach may be achieved within the three components of the neural network technique: the architecture of the network, the learning algorithm, and the learning database.

We present here an application to the problem of surface temperature and the atmospheric temperature profile retrievals with IASI. Previous studies have used information content analysis to estimate the expected retrieval errors for IASI (Amato and Serio 1997; Prunet et al. 1998), but this kind of estimate is dependent on some assumptions (Gaussian hypothesis; independence of first guess and observation; first-guess error covariance matrices often taken to be diagonal, i.e., no correlations among the first-guess errors of the variables; etc.) and on the limited number of atmospheric situations that have been examined.

Our neural network model is trained and tested using a large number, 3500, 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, 2000). These atmospheric situations include very complex temperature profiles that are often much more irregular than either reanalysis data or model output data. Rare situations are also included so that the dataset represents, as much as possible, all kinds of possible atmospheric situations (initially for a pattern-recognition purpose). This complexity represents a higher variability than that encountered in operational conditions with model output data, so our estimation of the retrieval errors could be an overestimate. The use of a large and complex climatological dataset allows the inversion model to be calibrated globally and even for rare events. Furthermore, our analysis of the retrieval error is made for realistic instrumental noise conditions. Contrary to other approaches, no assumptions about the physical problem, such as the linear or the Gaussian assumptions, are used.

This paper is organized as follows. The physical problem associated with our application is presented in section 2. The neural network approach is described in section 3. The databases used in this study are presented in section 4. Two applications of our neural technique are then presented: the surface temperature retrieval (section 5) and the atmospheric temperature profile retrieval (section 6). Conclusions and perspectives are given in section 7.

2. Sounding the atmosphere with the IASI instrument

a. Radiative transfer in the atmosphere

The radiance measured by an instrument at the top of the atmosphere depends on the atmospheric and surface physical properties. This dependence is described by the RTE:

$$I(\nu) = \varepsilon_s B(T_s, \nu)\tau_{s\nu}$$
$$+ \int_{P_s}^{P_{top}} B[T(P), \nu]\frac{\partial \tau_\nu}{\partial \ln P}(P)\, d\ln P, \quad (1)$$

where ν is the wavenumber (cm^{-1}); ε_s is the earth's surface emissivity, which may be a function of wavenumber; $B[T(P), \nu]$ is the Planck function, which indicates the radiance emitted by a blackbody at temperature T and atmospheric pressure P; and τ_ν is the atmospheric transmission between the satellite and the pressure level P. Also, $\partial \tau_\nu/\partial \ln P$ is termed the weighting function since, as is seen in Eq. (1), it weights the Planck radiance contribution to the column radiance.

To retrieve atmospheric profile variables from radiative measurements at the top of the atmosphere, the inverse of Eq. (1) has to be solved. The analytical inversion of this equation is not possible; only an inference approach can be used (Twomey 1977). Contrary to the direct problem, which advantageously may be estimated with high precision by a physical algorithm, the inverse problem needs a method of resolution based on a statistical representation of the (unknown) inverse equation. Two general approaches exist: using an inversion scheme for each observation (we call this approach the *local* inversion) or modeling the inverse RTE once and for all (we call this approach the *global* inversion). The local inversion generally requires a good initial guess to constrain the solution and a rapid and accurate direct transfer model (Rodgers 1976). Even if global inversion models use a first guess (Aires et al. 2001), this is not required, and no direct model is required during operational use. Although global inversion does not have these two limitations, it is a more ambitious problem.

b. Instrumental characteristics

There are two major advances of the IASI instrument. First is 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 overconstrained problem (more observations than degrees of freedom). Second, the resolving power is increased: with IASI the resolving power is about $\lambda/d\lambda$ of 1200, where λ is wavelength. The resolving power of the Television and Infrared Observation Satellite—Next Generation (TIROS-N) Ob-

Table 1. NEΔT noise characteristics of IASI at 280 K.

ν (cm^{-1})	NEΔT (K)	ν (cm^{-1})	NEΔT (K)
650	0.28	1660	0.34
770	0.28	2090	0.5
790	0.34	2100	0.5
980	0.34	2420	0.5
1000	0.28	2430	0.6
1070	0.28	2500	0.77
1080	0.34	2600	1.1
1200	0.34	2700	1.58
1210	0.28	2760	1.97
1650	0.28		

servational Vertical Sounding (TOVS) radiometer is presently between 50 and 100.

The IASI noise is simulated (Cayla et al. 1995) by a white Gaussian noise (this is a realistic assumption for interferometers) with a noise equivalent temperature (NEΔT) at 280 K (Table 1). NEΔT at 280 K represents the standard deviation $\mathrm{st}_{280}(\nu)$ of the Gaussian noise for a given wavenumber ν. At a different scene brightness temperature T', the standard deviation $\mathrm{st}_{T'}(\nu)$ of the Gaussian noise is computed by

$$\mathrm{st}_{T'}(\nu) = \frac{\dfrac{\partial B(\mathrm{Tb} = 280, \nu)}{\partial \mathrm{Tb}}}{\dfrac{\partial B(\mathrm{Tb} = T')}{\partial \mathrm{Tb}}}\, \mathrm{st}_{280}(\nu), \qquad (2)$$

which shows that the noise level increases as T' decreases. Figure 1 illustrates the standard deviation of noise at different T'. It is expected that these charac-

teristics are an overestimation of the actual noise level for the instrument. Figure 2 shows the IASI spectrum averaged over the TIGR dataset with the corresponding noise standard deviation spectrum. Note that some spectral regions could have a noise standard deviation larger than 2 K on average.

There are four fields-of-view for each IASI sample, covering an area with a diameter of 9–12 km at nadir. Assuming homogeneous meteorological conditions, an average of the four pixel measures can be used to perform the retrievals: these four fields-of-view provide redundant measurements that can be averaged to reduce noise.

3. The neural network inversion approach

Various neural inversion techniques have been developed, such as the "iterative inversion" (Kindermann and Linden 1990), the "distal learning" (Jordan and Rumelhart 1992), or the distal learning optimized by a Monte Carlo algorithm (Hidalgo and Gómez-Treviño 1996). We have chosen to use the "direct inversion" approach for two reasons: it performs a global inversion and it is possible to introduce "a priori" information into the method. The a priori knowledge is any information about the solution of the problem that is in addition to the information contained in the observations. In usual statistical techniques (such as regression), overcoming the "black box" modeling conception (no assumptions about the physical problem) improves results. Therefore, we have combined three approaches: the structural stabilization of the network, regularization of the learning algorithm by the input perturbation tech-
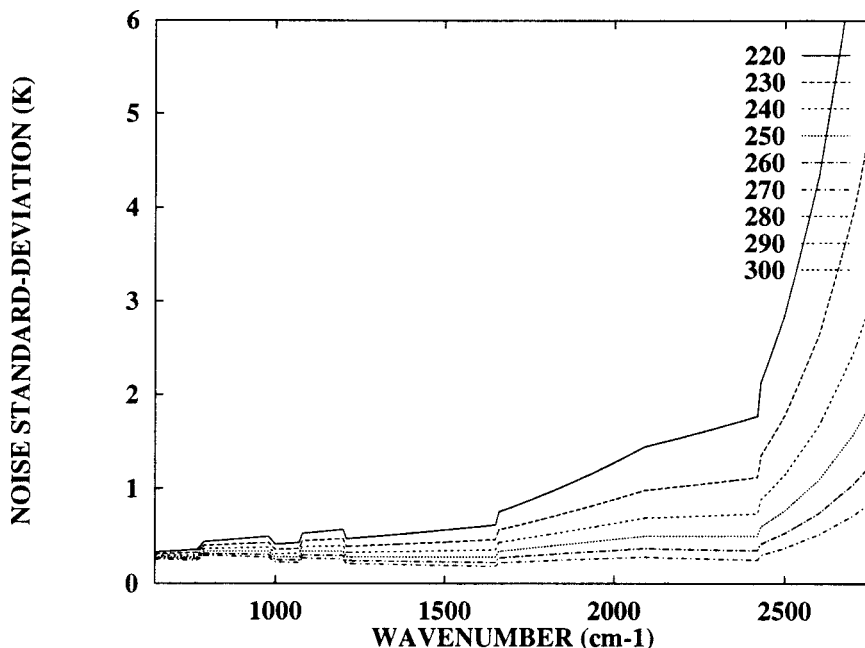

Fig. 1. Standard deviation of IASI instrument noise for different brightness temperature measurements T'.
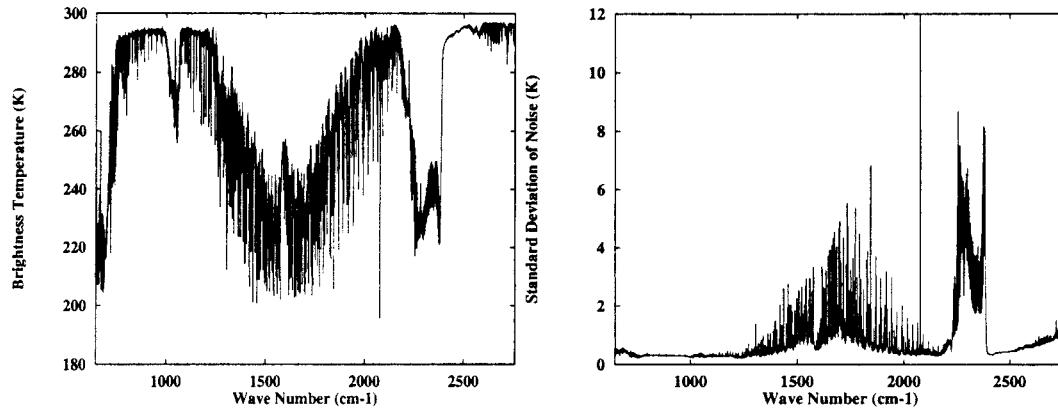
FIG. 2. (left) Mean IASI spectrum and (right) corresponding standard deviation of IASI instrumental noise.

nique, and a physically optimized feature selection process in the IASI data. Our numerical experiments have shown that the introduction of this kind of a priori information is very useful and makes training possible with relatively few data.

a. Global inversion

In the direct inversion technique, an MLP neural network is used to estimate directly the mapping between the IASI observations and retrieved geophysical variables. In effect, the "trained" MLP is a statistical model of the inverse RTE, providing once and for all a global inversion. The learning algorithm (the more expensive computational part) is performed offline only once. Then, the application of the neural network model for the inversion of IASI observations is quasi immediate in the operational stage: no regressions and no Jacobian computations are required.

Another advantage over classical physico-statistical techniques is that a good initial condition for the inversion is not needed. Moreover, the required memory storage is very small. There is also no need for a rapid direct model (necessary in iterative inversion algorithms) in which the speed is usually obtained by linearizing the RTE and assuming uncorrelated Gaussian errors.

b. MLP and structural stabilization of the architecture

The MLP network is a mapping model composed of parallel processors called "neurons." These processors are organized in distinct layers: the first layer (number 0) represents the input $X = (x_i; \ 0 \leq i \leq m_0)$ of the mapping, where m_0 is the number of neurons in layer 0. The last layer (number L) represents the output of the mapping $Y = (y_k; \ 0 \leq k \leq m_L)$. The intermediate layers $(0 < m < L)$ are called the "hidden layers." These layers are connected via neuronal links (Fig. 3): two neurons i and j between two consecutive layers have synaptic connections associated with a synaptic weight ω_{ij} (we pose W as the set of all synaptic weights ω_{ij}). A neuron executes two simple operations: first, it makes a weighted sum of the inputs and then transfers this signal to its output through a so-called transfer or ac-
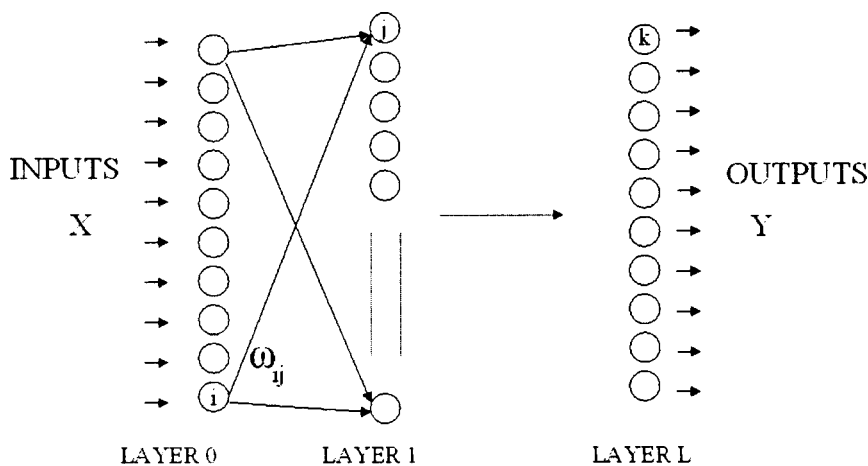

FIG. 3. Architecture of an MLP neural network with L layers, with inputs X and outputs Y.

tivation function such as $\sigma(a) = \tanh(a)$. The neuron j of a hidden-layer has an output z_j given by $z_j = \sigma(\Sigma_{l=0}^{m_0} \omega_{ij} z_l)$. For regression problems, the output units generally have a transfer function that is identity. For example, in a one-hidden-layer MLP, the kth output y_k of the network is defined as

$$y_k(y) = \sum_{j \in S_1} \omega_{jk} \sigma(a_j) = \sum_{j \in S_1} \omega_{jk} \sigma\left(\sum_{i \in S_0} \omega_{ij} x_i\right), \quad (3)$$

where σ is the sigmoid function, a_j is the activity of neuron j, and S_i is the ith layer of the network (with $i = 0$ for the input layer). We have deliberately omitted the usual bias term in this formula to simplify notation. It has been demonstrated (Hornik et al. 1989; Cybenko 1989) that any continuous function may be represented by a one-hidden-layer MLP.

The neuron acts, in its entire input space, as a "fuzzy" linear discriminant: a neuron j cuts its input space into two half subspaces separated by a plane orthogonal to the vector of its input weights $\{\omega_{ij}; l \in \text{Inputs}(j)\}$. On one side of the "decision boundary" the response of the neuron is 0, on the other side the response is 1, and in the fuzzy region the response of the neuron is quasi linear (corresponding to the approximately linear part of the transfer function). So, the MLP network, like linear regression, is adapted very well to high-dimensional data because its neurons act in the entire data space and not in a partition of this space as with some methods (radial basis function, splines interpolators, etc.).

How is the neural network structure defined? First, the number of inputs and outputs in the neural network is fixed entirely by the problem.

Second, the number of layers has to be defined. It has been demonstrated theoretically (Sontag 1992) that any inverse problem may be resolved by a two-hidden-layer MLP network because such neural networks can take into account discontinuities and extremely nonlinear variations (often present in inverse problems), in contrast to one-hidden-layer MLPs that approximate continuous functions. In practice, the answer is different. We have observed in our experiments that, with noise-corrupted data, a one-hidden-layer network may be sufficient. Furthermore, our experiments show that smooth solutions are obtained using just one hidden layer. This limitation in the number of hidden layers is a structural stabilization of the solution. The resulting reduction of the number of free parameters (the synaptic weights W) regularizes the neural estimation, producing a functional equivalence between the desired function (the inverse of the RTE) and its estimation (the trained neural network). Because we have noisy observations and relatively smooth behavior of the functions (we prefer smooth retrieved profiles over irregular ones), it is necessary to regularize the inverse problem. One way is to constrain the solution to be smooth (this kind of regularization is used also in ridge regression or in varia-

tional assimilation); we use a one-hidden-layer network to accomplish this smoothness.

Third, the user has to specify the number of neurons in the hidden layer. The more neurons in the hidden layer, the better is the fit to the learning dataset. However, the learning fit error is not a good criterion to constrain the neural architecture because having too many neurons produces the overfitting problem: the network fits the learning dataset very well but is bad for generalization (i.e., the fit error on an independent dataset 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 when compared with the desired model, and the overfitting problem appears (i.e., variance error). This dilemma is called 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.

c. Learning algorithm and regularization by input perturbation

Given an architecture (number of layers, input and output nodes, and interconnections), all the information for the network is contained in the weights W. The learning algorithm is the optimization technique that estimates the optimal network parameters $W = \{\omega_{ij}\}$ by minimizing a loss function $C(W)$ so that the neural mapping approaches as closely as possible the desired function. The most frequently used criterion to adjust W is the mean-square error in network outputs:

$$C(W) = \frac{1}{2} \sum_{k=1}^{m_L} \int\int [y_k(x; W) - t_k]^2 P(t_k|x)P(x) \, dt_k \, dx, \quad (4)$$

with t_k the kth desired output component, y_k the kth neural output component, and $P(x)$ the probability density function of input data x. In practice, $C(W)$ is approximated by

$$\overline{C}(W) = \frac{1}{2N} \sum_{e=1}^{N} [y_k(x; \omega) - t_k]^2. \quad (5)$$

The error back-propagation algorithm (Rumelhart et al. 1986) is used to minimize $C(W)$. It is a stochastic steepest-descent method very well adapted to this neural architecture because the computational cost is linearly related to the number of parameters.

To reduce the estimation sensitivity to input noise in the data, we use the input perturbation (IP) technique. It is a heuristic method to control the effective com-

plexity of the neural network mapping. The technique consists, during the learning step, of adding to each input a random vector representing the instrumental noise. It has been demonstrated (Bishop 1996) that, under certain conditions (low noise assumption), training with noise is closely related to regularization (or smoothing) technique. In the IP method, the usual error function $C(W)$ [Eq. (4)] takes the form

$$\tilde{C}(W) = \frac{1}{2} \sum_{k=1}^{m_L} \int \int \int [y_k(x + \eta; W) - t_k]^2 P(t_k|x) P(x) P(\eta) \, dt_k \, dx \, d\eta. \tag{6}$$

If the noise η is sufficiently small, we may expand the network function $y_k(x + \eta; \omega)$ to first order. Then, we obtain the relationship

$$\tilde{C}(W) \simeq C(W) + \nu \Omega(W), \tag{7}$$

where ν is the noise variance, and

$$\Omega(W) = \frac{1}{2} \sum_{i=1}^{m_0} \sum_{k=1}^{m_L} \int \left(\frac{\partial y_k}{\partial x_i}\right)^2 P(x) \, dx \tag{8}$$

is a Tikhonov penalty term (i.e., stabilizer) that avoids solutions with high gradients (rapid variations of the neural function). So, the minimization of this new criterion $C(W)$ constrains the solutions to be smooth. This regularization technique limits the number of degrees of freedom in the neural network to bring its complexity nearer to the desired function. This limitation reduces the class of possible solutions and makes the solution of the problem unique.

d. Feature selection for dimension reduction

An MLP neural network can, in principle, be used to map any input vector space to any output vector space; however, in practice, the data representation significantly affects the quality of the final results. In particular, care must be exercised to avoid an overemphasis on the noise component. Dimension reduction techniques may be used to present not only a more compact representation but also more pertinent information to the input of the neural network.

The ''curse of dimensionality'' stipulates that it is hard to apply a statistical technique to high-dimension-space data. We have seen in section 3b that the MLP is a well-adapted technique in this kind of problem, but practical problems still occur for high-dimensional data; for example, the number of parameters (the weights W in the MLP neural network) increases with the number of inputs. This can allow excessive of degrees of freedom in the neural interpolator, which, when combined with the introduction of noninformative data (i.e., noise or spectral information nonrelated to retrieved quantities), may distort the learning process: the quality criterion is more difficult to minimize and the computations are longer.

Thus, the goal of dimension reduction is to present to the neural network the most relevant information

from initial raw data (i.e., noisy physical measurements). There exist two ways to reduce the dimensionality of the input data (Jain and Zongker 1997): feature extraction (a transformation, linear or not, of raw data) and feature selection (selection of specific channels in input data; Bishop 1996). Feature selection is chosen here (Rabier et al. 2001). For the retrieval of one geophysical variable, we select channels that are, as far as possible, uniquely sensitive to this one atmospheric parameter. By studying the RTE Jacobians (derivatives of the transmittances with respect to each geophysical parameter), it is possible to analyze mutual information between measured brightness temperatures and geophysical variables (Chéruy et al. 1993). However, we need to make a compromise between reducing data dimensionality and preserving the redundant information in the raw data to alleviate the effects of noise.

4. Radiosonde-based learning and test datasets

a. Construction of an IASI learning dataset: The TIGR database

We use in our application the three TIGR databases of the ARA group: TIGR1 (861 atmospheres; Chédin et al. 1985), its 1990 revised version TIGR2 (1761 atmospheres: 322 in tropical air mass, 388 in midlatitude type 1, 354 in midlatitude type 2, 104 in polar type 1, and 593 in polar type 2; Achard 1991; Escobar et al. 1993), and its 1997 extended version TIGR3 (2311 atmospheres: same as TIGR2 but with an extended tropical air mass of 872 atmospheres; Chevallier et al. 1998). All of these datasets are constituted from more than 150 000 radiosonde measurements, sampled for their diversity and described by their temperature and gas concentration profiles with a discretization of the atmosphere into 40 layers (see Table 2). The sample includes a large number of rare events. The final database is composed of 3494 complex atmospheres. The minimum and maximum envelopes of the TIGR3 atmospheric temperature profiles are represented in Fig. 4 to illustrate the large range of variability that the radiosonde measurements represent. Not only is the range of variability extreme, but also inversion in the vertical profiles can produce complicated structures that are very challenging to all retrieval methods.

The Automatized Atmospheric Absorption Atlas (4A)

TABLE 2. The 40 pressure levels of the 4A algorithm.

Level	Pressure (hPa)	Altitude (km)	Level	Pressure (hPa)	Altitude (km)
1	0.05	68.4	21	131.20	14.1
2	0.09	64.3	22	161.99	12.6
3	0.17	59.9	23	200.00	11.1
4	0.30	56.0	24	222.65	10.4
5	0.55	51.8	25	247.90	9.7
6	1.00	47.7	26	275.95	8.9
7	1.50	44.9	27	307.20	8.2
8	2.23	42.2	28	341.99	7.4
9	3.33	39.4	29	380.73	6.7
10	4.98	36.6	30	423.85	6.0
11	7.43	33.9	31	471.86	5.2
12	11.11	31.1	32	525.00	4.5
13	16.60	28.3	33	584.80	3.7
14	24.79	25.6	34	651.04	3.0
15	37.04	22.8	35	724.73	2.3
16	45.73	21.3	36	800.00	1.6
17	56.46	19.9	37	848.69	1.2
18	69.71	18.4	38	900.33	0.8
19	86.07	17.0	39	955.12	0.4
20	106.27	15.5	40	1013.00	0.0

line-by-line forward radiative transfer algorithm (Scott and Chédin 1981; Tournier 1994) has been used to compute the IASI brightness temperatures associated with these 3494 atmospheres for clear conditions over the sea. The 4A algorithm allows for an analytical computation of the physical Jacobians (Chéruy et al. 1995). An illustration of such Jacobians versus pressure is given in Fig. 5 for the spectral region 650–800 cm^{-1} (15.5–12.5 μm). The vertical integration of the atmospheric information is illustrated in Fig. 6 in which Jacobians for six wavenumbers in the 15.5–12.5-μm spectral region are shown. Channels with a limited extent (mostly in the lower atmosphere), in terms of vertical resolution, provide more precise information than the others (in the top of the atmosphere) because a flat Jacobian indicates ambiguities in the retrieved profile. The spacing of the peaks is also important to reduce ambiguities. The concept of vertical resolution depends on both the width and the spacing of the channel's Jacobians (Rodgers 1990).

b. Improved representation of the surface temperature in TIGR

In the current TIGR database, the surface temperature Ts has been set equal to the temperature of the 40th (lowest) atmospheric level T40. This does not represent the actual situation, especially over land, where the surface skin temperature can differ significantly from the near-surface air temperature in systematic ways with time of day, latitude, season, and location (e.g., Rossow et al. 1989). For a better representation, we statistically generate, for each atmosphere, a set of 10 different Ts using the T40 information, based on the statistical distribution (i.e., mean and standard deviation) of T40 − Ts in a database of 150 000 radiosonde measurements. Thus, for every atmosphere, knowing T40, we choose

randomly 10 Ts with the estimated density probability. For example, in the tropical air mass, we obtain a Ts database of 3220 atmospheres (322 × 10).

5. Surface temperature retrieval

This study is limited to clear-sky oceanic situations and to the tropical airmass case, and emissivity is set equal to 1.0; Ts in the tropical air mass is very important to climatological analyses.

a. Jacobian-based channel selection

There are two spectral regions sensitive to the surface characteristics in the IASI spectral domain: 12.5–10.2 μm ($\simeq$800–980 cm^{-1}) and 4.0–3.6 μm ($\simeq$2500–2750 cm^{-1}). Note that the second spectral region may be contaminated by solar radiation during the day. However, in these regions, some wavelengths are contaminated by other atmospheric constituents. To eliminate the corrupted channels and to reduce the dimensionality (as explained in section 3d), we use a channel selection process based upon an analysis of the wavelength sensitivity of radiance to Ts variations. We define sensitivity as the mean variation $I(\nu)$ for 1-K change of Ts [see Eq. (1)]. We select, in these two windows, all channels with a sensitivity higher than a fixed threshold (Fig. 7), where the sensitivity of a channel is defined as the percent of repercussion of the channel measurement when surface temperature is increased by 1 K. Three hundred fifty-seven channels are obtained in the first window (with a threshold of 70%, which realizes a good compromise) and 262 are obtained in the second window (with a threshold of 85%, because channels are more sensitive to surface temperature in this window). These two thresholds have been chosen heuristically because they realize a good compromise between the dimension reduction of observations and the use of redundant and highly correlated channels for noise reduction.

b. Network learning and testing

The TIGR database (section 4b) is divided into a learning base of 3000 atmospheres to make the regression and a base of 220 atmospheres to test the generalization ability of the trained neural mapping.

To retrieve the variable Ts, we use a one-hidden-layer MLP neural network (see section 3b for structural stabilization). For the first window (800–980 cm^{-1}), the neural structure notated as 357–20–1 is selected in which 357 neurons are in the input layer (357 selected brightness temperatures), 20 neurons are in the hidden layer, and 1 neuron is in the output layer (representing Ts). For the second window (2500–2750 cm^{-1}), the structure selected is 262–20–1. The number of neurons in the hidden layer, 20, has been determined using the generalization test: we use different values and select the number of hidden neurons corresponding to the min-
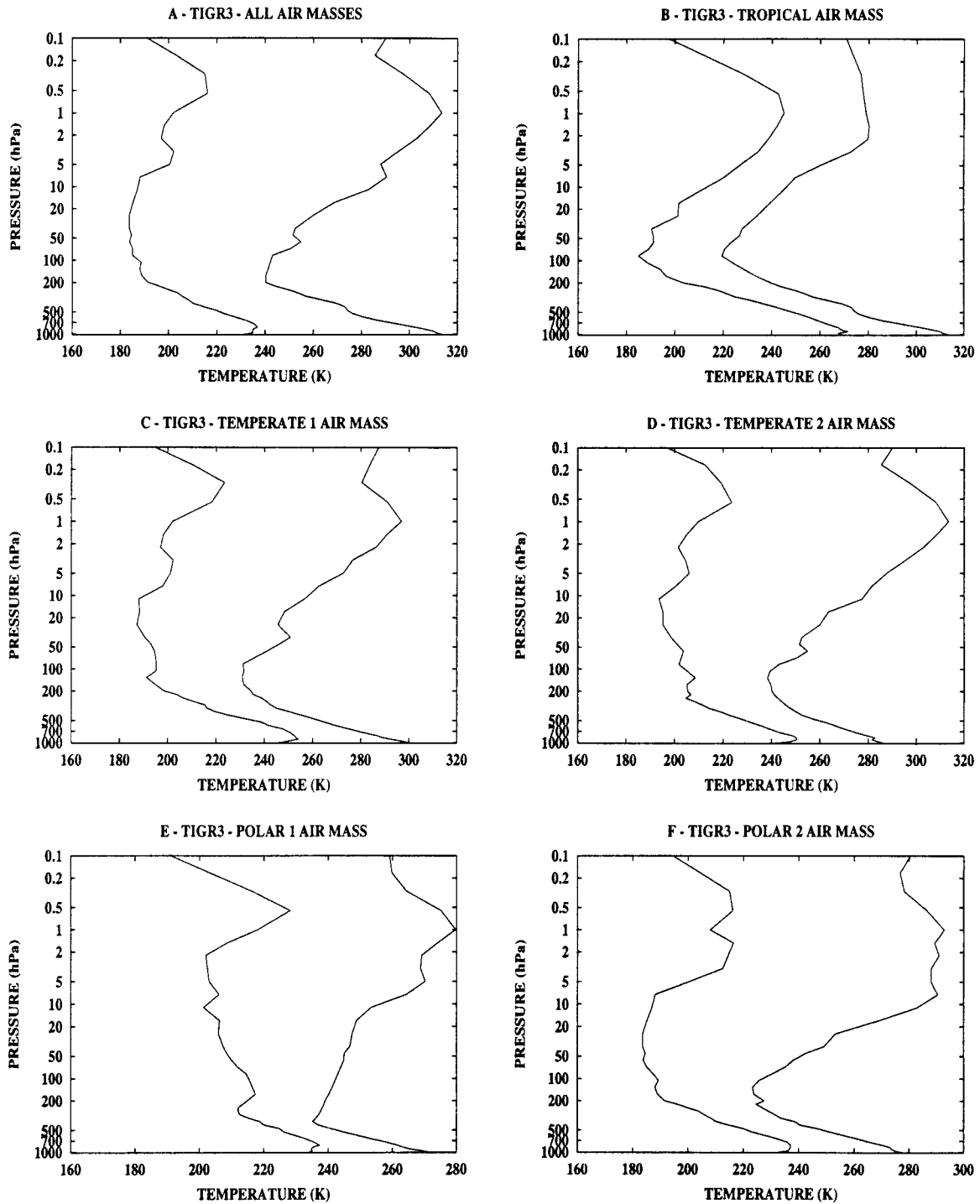
FIG. 4. Minimal and maximal envelope of TIGR3 atmospheric temperature profiles for (a) all air masses, (b) tropical air mass, (c) temperate-1 air mass, (d) temperate-2 air mass, (e) polar-1 air mass, and (f) polar-2 air mass.

imum generalization error. In this way, we have a neural network that is a good compromise between a small learning error and a small generalization error (to avoid the overfitting problem).

This neural mapping is trained by the error back-propagation algorithm on the learning base. The IP regularization technique is used: simulated noise (according to the NEΔT specifications) is added to the input data during the learning step. The generalization ability of our model was then tested on noisy data computed
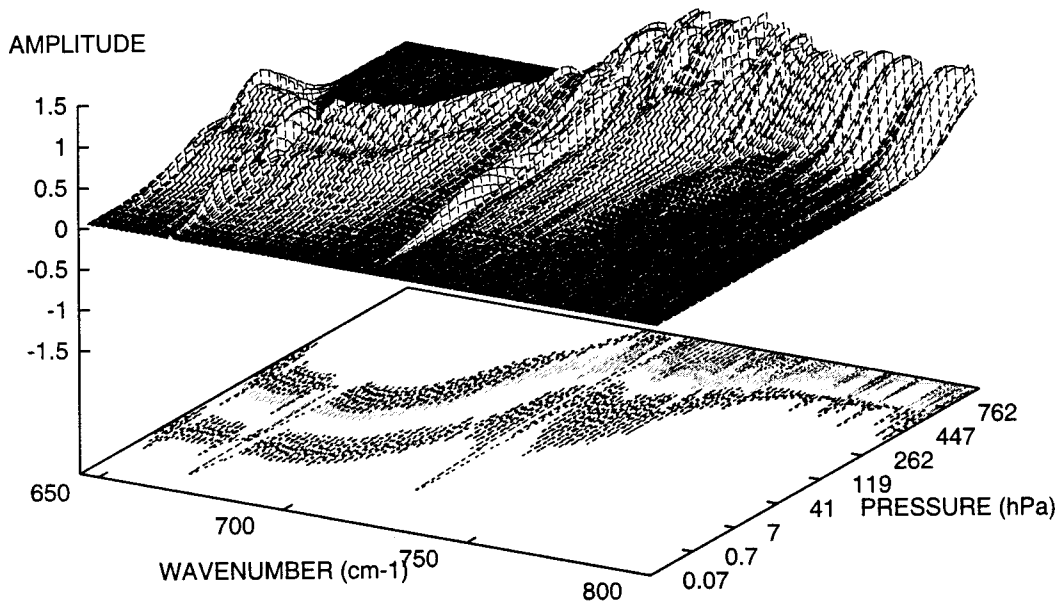
FIG. 5. Mean (for TIGR3 atmospheres) temperature Jacobian in the 15.5–12.5-μm spectral region. The temperature Jacobian is the derivative of radiative transfer function in Eq. (1) with respect to temperature $T(P)$ at pressure P.

on the 220 test atmospheres. The instantaneous retrieval of Ts from noisy data gives a generalization rms of approximately 0.4 K. Similar results are obtained using only the second spectral window. Without noise, the rms error is less than 0.3 K. This means that the retrieval error is significantly affected by measurement error: 1/4 of the rms is due to instrument noise and 3/4 is due to the neural regression fit.

6. Atmospheric temperature profile retrieval

The 40 layers of 4A (see Table 2) were used to compute the brightness temperature spectrum for clear conditions over ocean, but for the retrieval, the vertical discretization of the atmosphere has been changed (from 4A levels to 1-km levels) to match IASI specifications. The objective of this section is then to retrieve the 32 lower atmospheric temperatures of the 1-km-layer profiles.

a. Channel selection

The choice of the channels for the retrieval of temperature profiles is made so that they are, as much as possible, sensitive to only one constant-concentration gas; then, variations of $I(\nu)$ in Eq. (1) result mainly from temperature variations. Thus, the "CO_2 or NO_2 (or both) absorbing spectral regions" are used for the retrieval of atmospheric temperature profiles: the 15.5–12.5-μm ($\approx$645–800 cm^{-1}) and the 4.7–4.0-μm ($\approx$2100–2500 cm^{-1}) spectral regions.

To present the most relevant information to the neural network inputs (section 3d), we use a channel selection process. The feature selection method is based on the study of the Jacobians in order to define the sensitivity

of a channel to atmospheric temperature. The mean Jacobian in TIGR3 indicates the sensitivity relation between atmospheric layers and channels. The standard deviation of the Jacobian (around the mean) is negligible except near the surface; this means that the mean Jacobian is robust to the atmospheric situation except in the lower atmospheric layers.

The channel selection process has two steps. First, channels are selected that are satisfying some quality criteria (see Fig. 8), that is, specifying an information as unambiguously as possible: 1) the half width half height (or Jacobian extent), which characterizes the vertical resolution and the channel integration (in terms of the area below the Jacobian), has to be smaller than fixed threshold; 2) the half width half maximum of channel (Jacobian width at midmaximum) has to be smaller than fixed threshold so that the channels selected give a more vertically localized information; 3) the Jacobian center of a channel is near surface (with a threshold of two atmospheric layers); and 4) the channel Jacobian has a single peak. The results have been taken into account to determine the previous thresholds. We obtain 442 channels among the nominal 621. For the 15.5–12.5-μm spectral region, we have selected 442 channels from the 621 channels in the spectral range (645–800 cm^{-1} with 0.25-cm^{-1} resolution).

The second step chooses a vertically uniform subset of the channels that meet the quality criteria. The IASI instrument gives little information below 10 hPa, so our retrievals will be limited to the pressure range 1013–10 hPa (32 layers with discretization of 1 km). We have chosen nine channels for each of 30 layers (the previous 32 layers minus the two lowest layers sensitive to surface temperature) between 1013 and 10 hPa. The final
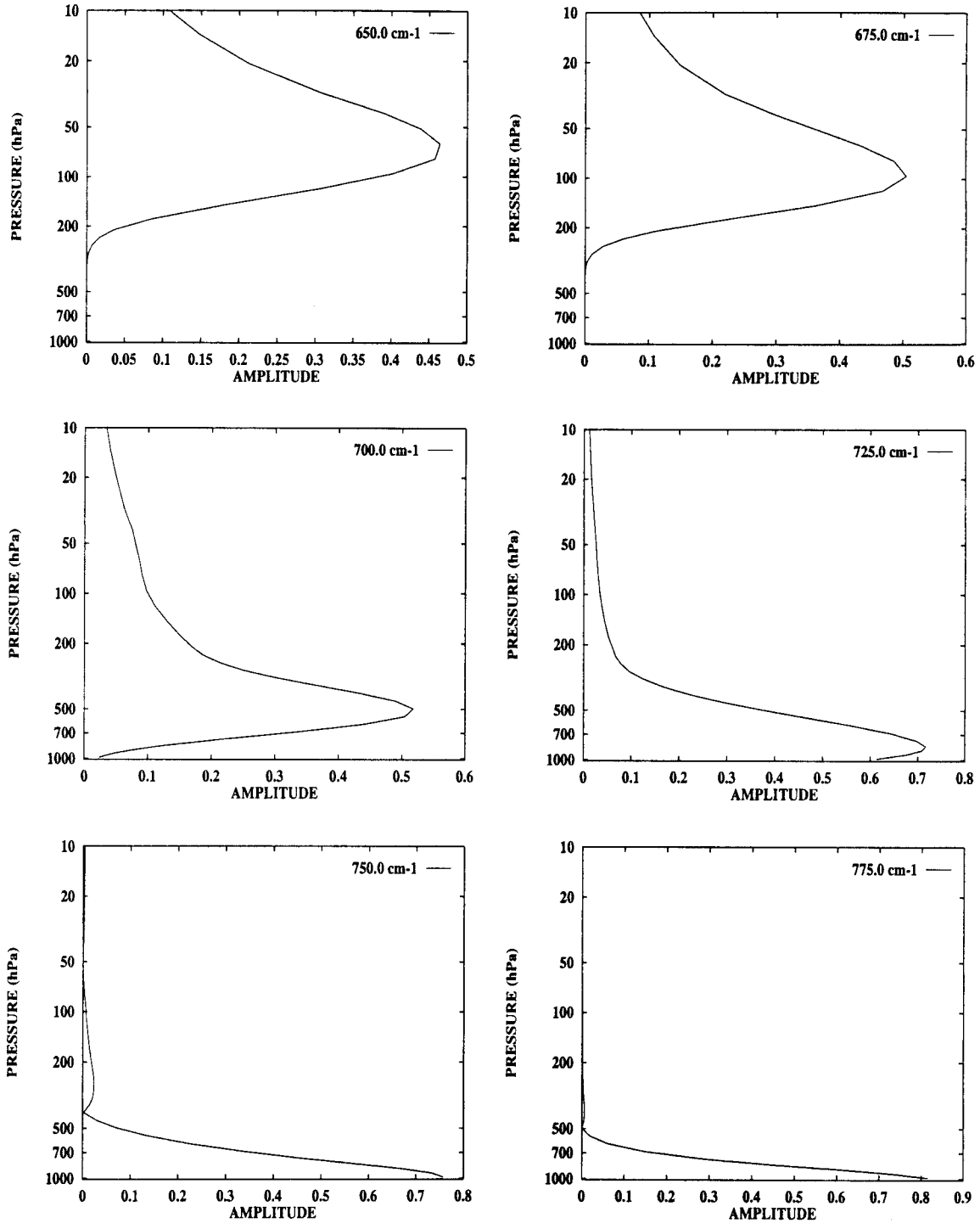
FIG. 6. Atmospheric temperature Jacobian profile for IASI and for six channels in the 15.5–12.5-μm spectral region.

number of channels is 270. However, it is important to note that layers 23–28 have a deficit in channels and that the sensitivity is higher in the lower atmosphere (Fig. 9).

The 4.7–4.0-μm spectral region is also important for the atmospheric temperature profile retrieval for two reasons. First, the lower-atmospheric Jacobians are narrower than in the 15.5–12.5-μm region, allowing for a better vertical resolution. Second, the channels are less affected by water vapor.

However, because of the larger noise in this spectral domain, the channel selection has to be performed dif-
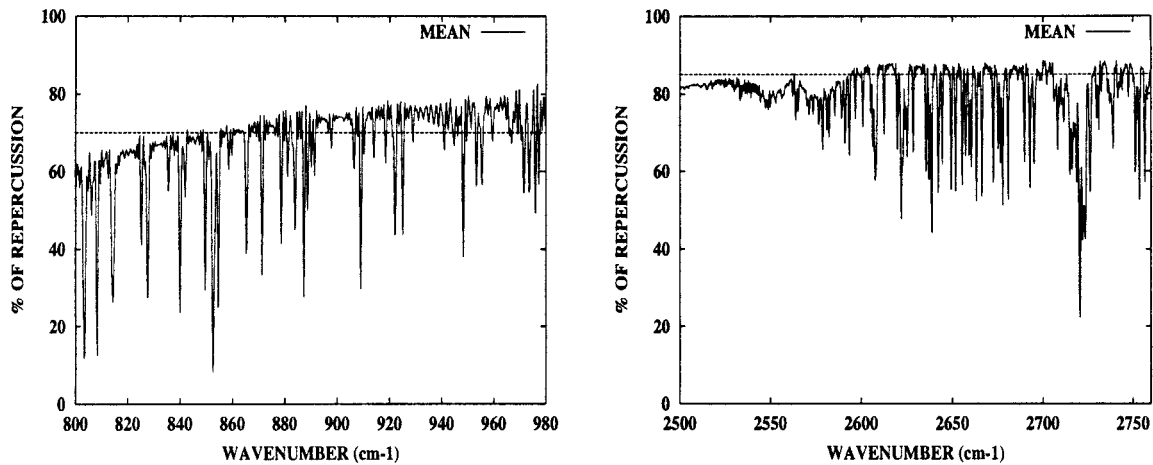
FIG. 7. Sensitivity in percent to a 1-K perturbation of surface temperature vs wavenumber in the two IASI windows (800–980 and 2500–2750 cm^{-1}).

ferently than in the 15.5–12.5-μm region. The IASI noise (see section 2b)—the standard deviation of the Gaussian noise—may be as large as a few degrees for channels sensing the higher layers (lower brightness temperatures). The redundancy of the information due to the number of channels does not compensate for this noise. The spectral range used consequently covers mainly the lower atmospheric layers. The Jacobian analysis selects channels in the 2140–2240-cm^{-1} spectral range (401 channels).

b. Network learning and testing

All the atmospheres used in the learning and the testing phases are described by 30 atmospheric temperatures (4A levels up to 7 hPa for 32-km height) and the corresponding 671 selected brightness temperatures computed by 4A. The neural network structure used for the regression is then 671–50–30: 671 units in the input layer (the 671 selected channels in the 15.5–12.5-μm and the 4.7–4.0-μm spectral regions), 50 units in the
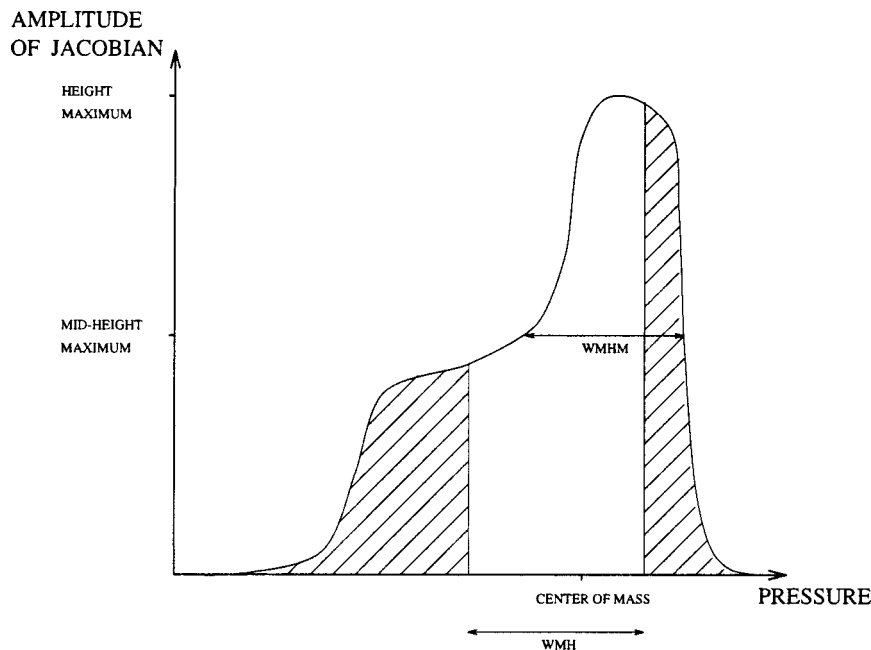

FIG. 8. Schematic representation of the quality criteria on channel Jacobian used for the channel selection on IASI spectra: WMHM is the width at midheight maximum (using height maximum) and WMH is the width at midheight (using area below the Jacobian).
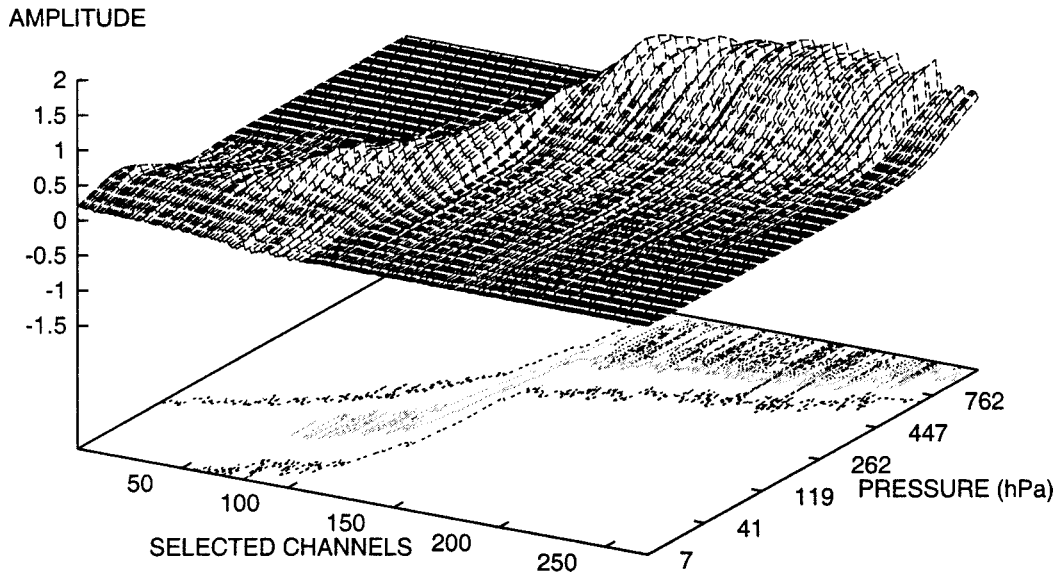
FIG. 9. Mean temperature Jacobian for TIGR3 for the 270 channels (ordered by maximum absorption altitude) selected from the 15.5–12.5-μm spectral region.

hidden layer, and 30 units in the output layer (the 30 lower atmospheric temperatures in 4A levels, the interpolation to 32 1-km levels being made afterward). The number of neurons in the hidden layer, 50, has been determined using the generalization test.

We have tested four configurations: the "all-airmasses" and the "tropical-airmass" configurations, with and without the four-pixel averaging scheme (noise divided by two; see section 2b).

1) ALL-AIRMASSES CONFIGURATION

We have merged TIGR1 and TIGR3 databases of section 4a, and the resulting 3155 atmospheres have been randomly subdivided into a learning base of 2700 atmospheres and a test base of 455 atmospheres.

The rms-fit errors (given for the 32 atmospheric 1-km layers) for the learning and the test sets are shown in Fig. 10a for the one-pixel configuration and Fig. 10b for the four-pixel configuration. We have overall good agreement among the computed and "observed" temperature profiles: rms errors are close to 1 K on average (less than 1.3 K except near 10 hPa). Also, we can see that we are facing some problems in two vertical regions:

1) In the upper layers of the atmosphere, IASI provides poor information above 20 hPa (see Fig. 5) because of the fact that the Jacobians of the channels sounding these layers are more vertically extensive than channels near surface and their amplitudes are smaller. So, the compensation phenomenon is more important in this vertical region. Some of our experiments have shown that the addition of the AMSU/A (also planned for flight on Aqua) information improves results in this vertical region (Aires 1999).

2) In the near-surface layers, the difference T40 $\neq$ Ts complicates the retrieval because of, in part, the compensation phenomenon (an underestimation of temperature in one layer is compensated by an overestimation in a nearby layer). Consideration of specific neural network compensation phenomena is given in Aires et al. (1999) and Aires (1999). It is possible that the simultaneous retrieval of Ts and T40, being more constrained, may solve this problem.

Thus, even though the TIGR database possesses atmospheric situations with highly variable temperature profiles, the rms errors obtained in Figs. 10a and 10b are close to the IASI objective (1 K of rms error for 1 km in vertical resolution).

The use of four-pixel averages uniformly decreases (by about 0.1 K) the rms error in the atmospheric layers. This relatively small improvement can be explained by the fact that IASI has many more channels than do current instruments. The redundancy between channels already suppresses part of the instrument noise. The retrieval of atmospheric temperature is more sensitive to fit error than the retrieval of surface temperature because IASI has fewer channels for atmospheric temperature than for surface temperature. Furthermore, the atmospheric temperature channels have a radiative transfer function–Jacobian that is vertically broader, which means that the information that they provide is more ambiguous than those of surface temperature channels. It is then normal that the retrieval of atmospheric temperature profile is more sensitive to fit error and less to instrument noise error. This result shows also that the solution regularization used to avoid noise effects, by the IP method, is sufficiently efficient that the reduction of noise by pixel-averaging has little impact on the qual-
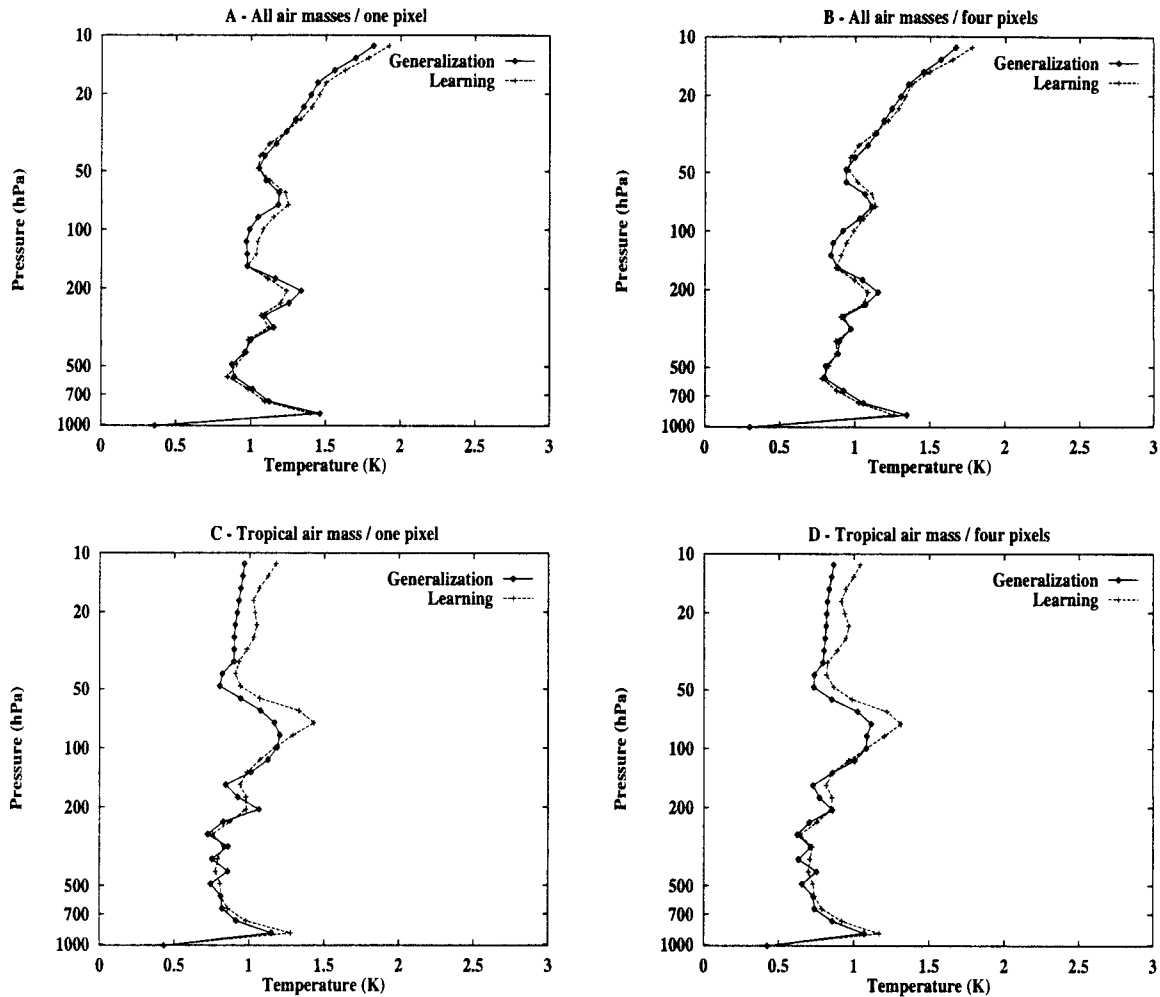
FIG. 10. Rms error profile for the atmospheric temperature retrieval in the learning set (continuous line) and in the generalization set (discontinuous line) for the (a) all-airmasses–one-pixel, (b) all-airmasses–four-pixel, (c) tropical-airmass–one-pixel, and (d) tropical-airmass–four-pixel configurations.

ity 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. Five randomly chosen examples of retrievals using the test set are shown in Fig. 11.

2) TROPICAL-AIRMASS CONFIGURATION

We have merged the Tropical-TIGR1 and the Tropical-TIGR3 databases of section 4a, and the resulting 1070 atmospheres have been randomly subdivided into a learning base of 1000 atmospheres and a test base of 70 atmospheres.

The rms errors (given for the 32 atmospheric 1-km layers) in the learning and the test set are given in the Fig. 10c for the one-pixel configuration and in the Fig. 10d for the four-pixel configuration. We see that the rms error profile is significantly improved at 1 K, so the specialization of the neural network to the tropical air

mass is important. As above, the rms error is also decreased by about 0.1 K with the four-pixel average configuration.

It is important to note that the specialization of the neural network for a specific air mass improves the retrievals but requires a training database with a larger number of atmospheres. In this case, the 1070 tropical atmospheres are not sufficient, so differences between the learning and the test databases are not negligible. Future work should address this very important problem of both extensive and comprehensive learning and test databases.

7. Conclusions and perspectives

A neural network approach uses maximum a priori information to limit the number of free parameters in the neural model so as to constrain the retrieval of surface and atmospheric temperatures to a "better-posed"
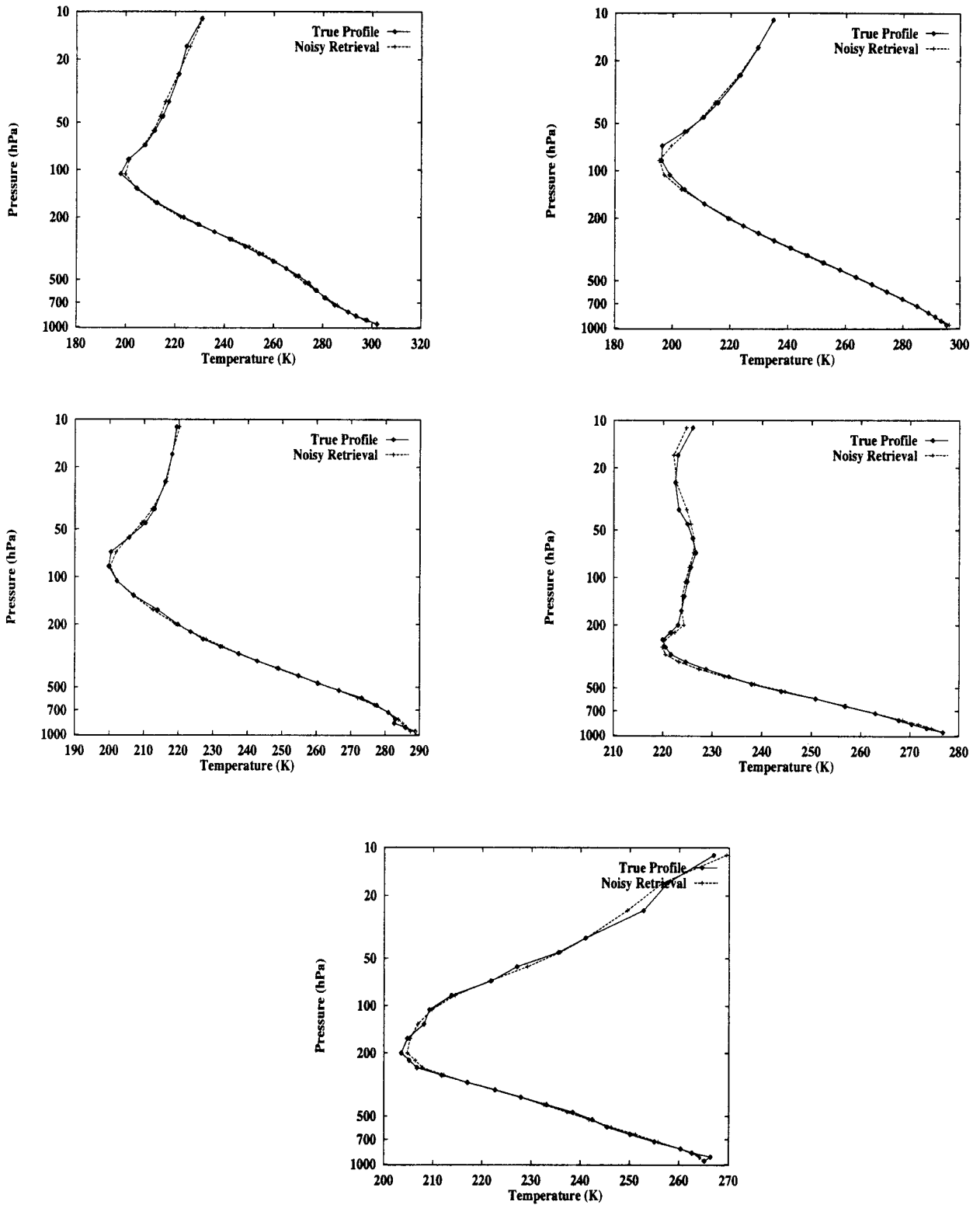
FIG. 11. Five atmospheric temperature profile retrieval examples in the all-airmasses–one-pixel configuration.

problem. The method is trained using the TIGR database, that is, a vast and complex set of atmospheric situations (from radiosonde measurements that are much more irregular than model output) with a wide range of radiosonde conditions, including rare events. This fact is important in judging the quality of results. The surface temperature for tropical situations yields an rms error of 0.4 K for instantaneous retrievals. Results for atmospheric temperature profile retrievals are given for four configurations (all-air masses or tropical-air mass, with and without the four-pixel average). Results are close to the specifications of the World Meteorological Organization (WMO) for the all-airmasses configurations, namely, 1-K rms error for the instantaneous temperature retrieval with 1-km vertical resolution. The specialization of the tropical-air mass significantly improves the results, which means that using a specialized neural network for a few different air masses is the appropriate strategy to adopt, but a larger dataset is then required to train these specialized models. It is important to note that the results obtained for the IASI retrievals entirely depend on the complexity of the dataset used to perform the statistics. Thus, this work has demonstrated the potential of the IASI instrument to achieve the WMO specifications for realistic conditions even for the complex situations included here. This new instrument is a clear advance over current instruments. The MLP inversion technique developed here for the processing of IASI observations is flexible enough to introduce a priori information into the retrieval scheme, is robust to noise, is accurate, and is very fast.

We plan to use independently a neural network for the two other air masses (temperate and polar) by increasing the TIGR database. Another idea is to use this methodology with more spectral channels so as to retrieve not only the surface temperature and the temperature profile, but also water vapor and ozone profiles. The simultaneous retrieval of these variables is expected to exploit the correlations between variables so as to constrain better the inversion process. Considerable improvements are expected by the use in parallel of AMSU/A observations [see Prigent et al. (2001) for multi-intrument information fusion by neural network]. Further improvement also may be expected by the introduction of a first-guess solution in the MLP inversion (Aires et al. 2001).

REFERENCES

Achard, V., 1991: Trois problèmes clés de l'analyse tridimensionelle de la structure thermodynamique de l'atmosphère par satellite: Measure du contenu en ozone, classification des masses d'air, modélisation hyper-rapide du transfert radiatif (Three key problems of the 3D analysis of the thermodynamical structure of the atmosphere by satellite: Ozone content measure, air-mass classification, and hyper-fast radiative transfer). Ph.D. dissertation, Université Pierre et Marie Curie, 168 pp. [Available from Laboratoire de Météorologie Dynamique, École Polytechnique, 91128 Palaiseau Cedex, France.]

Aires, F., 1999: Problèmes inverses et réseaux de neurones: Application à l'interféromètre haute résolution IASI et à l'analyse de séries temporelles (Inverse problems and neural networks: Applications to the high-resolution interferometer IASI and to time series analysis). Ph.D. dissertation, Université Paris IX/Dauphine, 220 pp. [Available from Laboratoire de Météorologie Dynamique, École Polytechnique, 91128 Palaiseau Cedex, France.]

——, R. Armante, A. Chédin, and N. A. Scott, 1998: Surface and atmospheric temperature retrieval with the high resolution interferometer IASI. Preprints, *Ninth Conf. on Satellite Meteorology and Oceanography,* Paris, France, Amer. Meteor. Soc., 181–186.

——, M. Schmitt, N. A. Scott, and A. Chédin, 1999: The weight smoothing regularization for Jacobian stabilization. *IEEE Trans. Neural Networks,* **10,** 1502–1510.

——, C. Prigent, W. B. Rossow, and M. Rothstein, 2001: 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,** 14 887–14 907.

Amato, U., and C. Serio, 1997: The impact of random noise on the performance of a new infrared atmospheric sounding interferometer (IASI) configuration. *Int. J. Remote Sens.,* **18,** 3135–3143.

Bishop, C., 1996: *Neural Networks for Pattern Recognition.* Clarendon Press, 482 pp.

Cayla, F., B. Tournier, and P. Hebert, 1995: Performance budgets of IASI options. Tech. Rep. IA-TN-0B-5476-CNE, 21 pp. [Available from Centre National d'Études Spatiales, 2 place Maurice Quentin, 75039 Paris Cedex 01, France.]

Chédin, A., N. A. Scott, C. Wahiche, and P. Moulinier, 1985: The improved initialization inversion method: A high resolution physical method for temperature retrievals from TIROS-N series. *J. Climate Appl. Meteor.,* **24,** 128–143.

Chéruy, F., J. Susskind, N. A. Scott, A. Chédin, and J. Joiner, 1993: GLA and LMD approaches to the processing of AIRS and IASI observations. *High Spectral Resolution Infrared Remote Sensing for Earth's Weather and Climate Studies.* NATO ASI Series I, Vol. 9, Springer-Verlag, 101–112.

——, N. A. Scott, R. Armante, B. Tournier, and A. Chédin, 1995: Contribution to the development of radiative transfer models for high spectral resolution observations in the infrared. *J. Quant. Spectros. Radiat. Transfer,* **53,** 597–611.

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

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

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

Escobar, J., 1993: 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 (Database for atmospheric parameters at global scale; neural network inversion of present and future remote sensing satellites). Ph.D. dissertation, Université Denis Diderot 190 pp. [Available from Laboratoire de Météorologie Dynamique, École Polytechnique, 91128 Palaiseau Cedex, France.]

——, A. Chédin, F. Chéruy, and N. A. Scott, 1993: Réseaux de

neurones multicouches pour la restitution de variables thermo-
dynamiques atmosphériques à l'aide de sondeurs verticaux sa-
tellitaires (Multilayer neural network for retrieval of atmospheric
thermodynamical variables using satellite remote sensing).
Compt. Rend. Acad. Sci. Paris, **317,** 911–918.

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

Hidalgo, H., and E. Gómez-Treviño, 1996: Application of construc-
tive learning algorithms to the inverse problem. *IEEE Trans.
Geosci. Remote Sens.,* **34,** 874–885.

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

Jain, A., and D. Zongker, 1997: Feature selection: Evaluation, ap-
plication, and small sample performance. *IEEE Trans. Pattern
Anal. Mach. Intelligence,* **19,** 153–158.

Jordan, M., and D. Rumelhart, 1992: Forward models: Supervised
learning with a distal teacher. *Cognit. Sci.,* **16,** 307–354.

Kindermann, J., and A. Linden, 1990: Inversion of neural networks
by gradient descent. *Parallel Comput.,* **4,** 277–286.

Prigent, C., F. Aires, W. B. Rossow, and E. Matthews, 2001: Joint
characterization of vegetation by satellite observations from vis-
ible to microwave wavelengths: A sensitivity analysis. *J. Geo-
phys. Res.,* **106,** 20 665–20 685.

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

Rabier, F., N. Fourrié, D. Chafaï, and P. Prunet, 2001: Channel se-
lection methods for infrared atmospheric sounding interferom-
eter radiances. *Quart J. Roy. Meteor. Soc.,* in press.

Rieu, H., J. Escobar-Munoz, N. A. Scott, and A. Chédin, 1996: SSM/
T forward modelling using neural networks. *J. Quant. Spectros.
Radiat. Transfer,* **56,** 821–833.

Rodgers, C. D., 1976: Retrieval of atmospheric temperature and com-
position from remote measurements of thermal radiation. *Rev.
Geophys. Space Phys.,* **14,** 609–624.

——, 1990: Characterization and error analysis of profiles retrieved
from remote sounding measurements. *J. Geophys. Res.,* **95,**
5587–5595.

Rossow, W. B., C. L. Brest, and L. C. Garder, 1989: Global, seasonal
surface variations from satellite radiance measurements. *J. Cli-
mate,* **2,** 214–247.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams, 1986: Learning
internal representations by error propagation. *Parallel Distrib-
uted Processing: Explorations in the Microstructure of Cogni-
tion: Foundations,* D. E. Rumelhart, J. L. McClelland, and PDP
Research Group, Eds., Vol. I, MIT Press, 318–362.

Scott, N. A., and A. Chédin, 1981: A fast line-by-line method for
atmospheric absorption computations: The automatized atmo-
spheric absorption atlas. *J. Appl. Meteor.,* **20,** 802–812.

Sontag, E., 1992: Feedback stabilization using two-hidden-layer nets.
IEEE Trans. Neural Networks, **3,** 981–990.

Tournier, B., R. Armante, and N. A. Scott, 1994: Stransac-93, 4A-
93: Développement et validation des nouvelles versions des
codes de transfert radiatif pour application au projet IASI (De-
velopment and validation of new versions of the radiative trans-
fer code for application to the IASI project). ARA/LMD Group
Tech. Rep. 201, 16 pp.

Twomey, S., 1977: *Introduction to the Mathematics of Inversion in
Remote Sensing and Indirect Measurements.* Elsevier Science,
243 pp.