Measure and exploitation of multisensor and multiwavelength synergy for remote sensing: 2. Application to the retrieval of atmospheric temperature and water vapor from MetOp

Filipe Aires,[1,2,3] Maxime Paul,[1,4] Catherine Prigent,[5] Björn Rommen,[6] and Marc Bouvet[6]

Received 30 June 2010; revised 22 October 2010; accepted 15 November 2010; published 20 January 2011.

[1] In the companion paper, classical information content (IC) analysis was used to measure the potential synergy between the microwave (MW) and infrared (IR) observations from Atmospheric Microwave Sounding Unit-A, Microwave Humidity Sounder, and Improved Atmospheric Sounding in the Infrared instruments, used to retrieve the atmospheric profiles of temperature and water vapor over ocean, under clear-sky conditions. Some limitations of IC were pointed out that questioned the reliability of this technique for synergy characterization. The goal of this second paper is to develop a methodology to measure realistic potential synergies and to construct retrieval methods able to exploit them. Three retrieval methods are considered: the *k* nearest neighbors, the linear regression, and the neural networks (NN). These statistical retrieval schemes are tested on an application involving IR and MW synergy. Only clear-sky, near-nadir radiances over ocean are considered. The IR/MW synergy is expected to be stronger in cloudy cases, but it will be shown that it can also be observed in clear situations. The inversion algorithms are calibrated and tested with simulated observations, without any loss of generality, using similar theoretical assumption (same radiative transfer model, observational noise, and a priori information) in order to truly compare the IC and the direct statistical retrieval approaches. Multivariate and nonlinear methods such as the NN approach show that there is a strong potential for synergy. Synergy measurement tools such as the method proposed in this study should be considered in the future for the definition of new missions: The instrument characteristics should be determined not independently, sensor by sensor, but taking into account all the instruments together as a whole observing system.

Citation: Aires, F., M. Paul, C. Prigent, B. Rommen, and M. Bouvet (2011), Measure and exploitation of multisensor and multiwavelength synergy for remote sensing: 2. Application to the retrieval of atmospheric temperature and water vapor from MetOp, *J. Geophys. Res.*, *116*, D02302, doi:10.1029/2010JD014702.

1. Introduction

[2] *Aires* [2011] tried to explain some synergy mechanisms, how they occur and how to use them. The first conclusions were twofold. First, there exist different types of potential synergy (additive, unmixing, indirect or denoising), and each one can have a strong impact on the retrievals. The experiments performed with a simple linear model illustrated well how to use all the a priori information in the retrieval problem. Second, classical information content (IC) analysis may not be adequate for measuring synergy. The application focused on the retrieval of the atmospheric temperature and WV profiles using three instruments (Atmospheric Microwave Sounding Unit-A (AMSU-A), Microwave Humidity Sounder (MHS), and Improved Atmospheric Sounding in the Infrared (IASI)) all on board the MetOp platform and for clear-sky, over ocean situations. We showed that IC is very sensitive to the hypothesis that are used: uncertainty estimates are dependent on the Jacobian, and they change when considering only temperature or water vapor or when both Jacobians are included in the estimation. In order to obtain a realistic uncertainty characterization, it is necessary to obtain truly state-dependent statistics and to take into account all the sensitivities (i.e., Jacobians) of the observations. The

[1]Laboratoire de Météorologie Dynamique, CNRS, IPSL, Université Pierre et Marie Curie, Paris, France.
[2]Also at Laboratoire de l'Etude du Rayonnement et de la Matière en Astrophysique, CNRS, Observatoire de Paris, France.
[3]Now at Estellus, Paris, France.
[4]Now at Laboratoire de l'Etude du Rayonnement et de la Matière en Astrophysique, CNRS, Observatoire de Paris, Paris, France.
[5]Laboratoire de l'Etude du Rayonnement et de la Matière en Astrophysique, CNRS, Observatoire de Paris, Paris, France.
[6]Wave Interaction and Propagation Section, European Space Agency, Noordwijk, Netherlands.

0148-0227/11/2010JD014702

simplicity of IC analysis is lost. No synergy between infrared and microwave observations was found but the reliability of IC to measure synergy was questioned.

[3] In this paper, we focus on the methodologies to optimize the combination of satellite information from different instruments. To demonstrate the potential of the method, we apply it to selected atmospheric parameters and wavelength bands under specific instrument geometry. The retrieval chain that is developed is designed for the MetOp-A satellite. This platform provides coincident observations in the IR, IASI, and in the microwaves, AMSU-A and MHS, with nadir geometries, for clear-sky and over ocean situations. We concentrate on the major atmospheric parameters, namely, temperature and water vapor profiles, for which the selected MetOp-A instruments are particularly sensitive. The developed methodology is very general and flexible and can be adapted to other applications, i.e., other variables or instruments.

[4] The preprocessing of IASI observations is necessary before any retrieval: this is presented in section 2. The retrieval methodologies (k-NN, linear regression, and neural networks) are introduced in section 3. Section 4 provides the results in terms of retrieval statistics and synergy measures. Finally, conclusions and perspectives are drawn in section 5.

2. Preprocessing of the IASI Observations

[5] In a retrieval scheme, it is particularly important to conduct a systematic search of the best possible predictors. The information used for the atmospheric profile estimation should be as directly related to it as possible, and should avoid any other sources of contaminating information (i.e., variability of the satellite observations related to a variable that is not retrieved). As a consequence, a careful data analysis should be performed first.

2.1. Dimension Reduction Approach

[6] 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 HIRS on TIROS-N Operational Vertical Sounding (TOVS). This is a major problem and classical retrieval algorithms are unable to deal with this huge amount of information. Iterative methods require a fast direct model with its Jacobians (i.e., the first derivative of the observation with respect to retrieved variables), such as the RTTOV model [*Saunders et al.*, 1999; *Matricardi et al.*, 2004]. Variational assimilation techniques also need a fast forward model with the Jacobians or the tangent linear operator. These approaches cannot use the full raw IASI information. As a consequence, dimension reduction techniques have to be applied to reduce the size of the data.

[7] Some of these approaches are physically based: the radiative transfer Jacobians indicate the channel sensitivity to geophysical constituents, their altitude and the vertical resolution. The physical channel selection procedure consists in sampling as uniformly as possible the vertical for each geophysical variable. An example of such a channel selection procedure is given by *Aires et al.* [2002a] where temperature and water vapor Jacobians are used to select 442 IASI channels among the original 8461. The limitation of this method is that the selection takes into account only the direct RT, not the inversion process: the selected channels might not be optimal for the inversion. Another approach consists in using the Jacobian information to construct a limited number of "metachannels". These metachannels gather few original channels based on their similarity and this integration of multiple original channels makes them more robust to noise for evident signal-to-noise reasons. This procedure is a feature selection, it implies a transformation of the original data, not a channel selection.

[8] Another family of compression technique has emerged from statistics and information theory. Concepts such as degrees of freedom [*Rodgers*, 2000] or entropy reduction are used to select appropriate channels or compress data. A few techniques can be cited: Data Reduction Matrices [*Menke*, 1984], minimum entropy [*Shannon*, 1949; *Huang and Purser*, 1996], Singular Value Decomposition [*Prunet et al.*, 1998], iterative approach [*Rodgers*, 2000]. Four channel selection techniques are compared by *Rabier et al.* [2002] for the IASI instrument. It should be noted that some compact representation techniques are lossless (e.g., ZL algorithm [*Ziv and Lempel*, 1979]). Lossless compression algorithms usually exploit statistical redundancy so that the data is represented more concisely without error. This type of compression is widely used in communication or media storage. Other compression techniques are destructive (e.g., channel selection) because the compression induces some level of degradation in the signal. This loss of information has to be acceptable for the application under consideration. This is an important point that needs to be controlled for Earth satellite observation applications.

[9] In the last decade, the technique that has found wide usage in the satellite community is the Empirical Orthogonal Function (EOF), i.e., another name for classical Principal Component Analysis (PCA). Although widely used for statistical analysis, this technique has proved efficient for raw compression [*Huang and Antonelli*, 2001; *Eriksson et al.*, 2002] and for the denoising of satellite observations [*Aires et al.*, 2002b]. EOF extracts from the original IR spectra the dominant components: these components are decorrelated to each other and explain a maximum of the satellite observation variance. EOFs are a linear transformation of the original IR spectra and can be considered as metachannels, each one being a weighted sum of the original channels. The EOF compact representation of the observed spectra can be used to transfer the data and reduce instrument noise, or it can be used directly in the retrieval process [*Aires et al.*, 2002c, 2002d]. *Smith and Taylor* [2004] use the EOF technique for the cloud clearing of IASI observations. In the work by *Liu et al.* [2009], the PCA is applied independently to the three IASI bands.

2.2. Compression

[10] Let $\mathcal{D} = \{y^e; e = 1, \ldots, E\}$ be a database of $E = 10000$ spectra, y, of dimension $M = 8461$. Let C_{yy} be the $M \times M$ covariance matrix of the $\mathcal{D}$ database. Let V be the $M \times M$ matrix with columns equal to the eigenvectors of C_{yy} and let L be the diagonal $M \times M$ matrix with the M associated eigenvalues in decreasing order (by definition $C_{yy} \cdot V = V \cdot L$).

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

$$\begin{cases} h = F \cdot y = F_{1\star} \cdot y_1 + \ldots + F_{M\star} \cdot y_M \\ y = F^{-1} \cdot h = F^t \cdot h = h_1 \cdot F_{\star 1} + \ldots + h_M \cdot F_{\star M} \end{cases} \quad (1)$$

where ^{t} is the transpose operator. The vectors $\{F_{i\star}; i = 1, \ldots, M\}$, i.e., rows of F, are called the filters and the normalized eigenvectors $\{F_{\star i}; i = 1, \ldots, M\}$ are called the PCA basis functions. Because these eigenvectors are an orthogonal basis set for representing the IASI spectra, y, we will refer to them as eigenspectra. By definition, the coordinates of the new data h are uncorrelated since:

$$< h \cdot h^t > = < F \cdot y \cdot y^t \cdot F^t > = < F \cdot C_{yy} \cdot F^t > = I_{M\times M}, \quad (2)$$

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

[12] Practically, the first step in a PCA approach is to compute the 8461 × 8461 covariance matrix $C_{yy} = < (y - < y >) \cdot (y - < y >)^t >$ of the database, where y is an IASI spectrum composed of the 8461 wavelengths. The PCA representation could be used on only specific region sensitive to the parameters to be retrieved (i.e., temperature and water vapor here). However, it is shown that PCA is able to deal with the full IASI spectra. The eigenvalue matrix L and the corresponding eigenvectors V of this covariance matrix C_{yy} are then computed using a Cholesky or a SVD decomposition.

[13] The solid line in Figure 1 represents the average (over the full IASI spectrum) compression error when increasing the number of PCA components. The more components used in the PCA representation (i.e., more eigenspectra used for the compression), the lower the compression error. When all 8461 components are retained, there is no compression error left. It can be seen that taking 5 components is not enough, but the error decreases rapidly with the number of components. With only $N = 50$ (the 50 first principal components), the RMS compression error of the IASI spectra averaged over the whole data set is close to 0.05 K, which is much lower than the average IASI noise which is close to 1 K. The spectral repartition of the error (not shown) is uniform over the IASI channels.

2.3. Denoising

[14] Noise is also a major concern since IASI has high levels of instrument noise in some spectral regions [see, e.g., *Aires et al.*, 2002b]. In this context, the full use of the IASI channels can help reduce instrument noise effect by exploiting information redundancy in channels. There is a possibility to suppress part of the noise during the compression process. Experiments show that statistically, higher-order components (1) explain less information (by definition of PCA), (2) are less physical because they result from mixing of the fraction/remaining parts of more important components, and (3) have lower signal to noise ratio since they convey less information and tend to represent noise. The idea is to find the patterns that have a physical meaning, these patterns define a "physical" subspace in the observation space. The noisy observations are projected in this subspace. The orthogonal subspace that would appear with additional components found with noisy data would likely be coding a part of the space described by the noise and should therefore be suppressed. In our denoising approach the lower-order principal components ($h_1, \ldots, h_N$) of a PCA decomposition are assumed to describe the real variability of the observations, or the signal, (here the IASI spectra) and the remaining principal components ($h_{N+1}, \ldots, h_M$) describe higher frequency variabilities in the IASI spectrum. These higher frequencies are more likely to be related to the white Gaussian noise of the instrument, or to the variability of minor atmospheric constituents. Only major constituents such as temperature, water vapor and ozone are of interest here and the higher-order components are considered to describe noise (instrumental noise plus irrelevant information).

[15] Let $\overline{F}$ be the $N \times M$ truncated matrix of F. The PCA decomposition uses this truncated matrix to project IASI spectra, y, of dimension $M = 8461$ into a space of lower dimension N (with $N \leq M$): $h = \overline{F} \cdot y$ and $\hat{y} = \overline{F}^{-1} \cdot h$ (same as equation (2) but with N instead of M and where $\overline{F}^{-1}$ is a generalized inverse matrix since $\overline{F}$ is not square). The compression error $|| y - \hat{y} ||$ is given by $|| h_{N+1} \cdot F_{\star N+1} + \ldots + h_M \cdot F_{\star M} ||$, where $|| \cdot ||$ is the Euclidean norm. PCA is optimum for the least squares errors criterion $\frac{1}{E} \sum_{e=1}^{E} || y^e - \hat{y}^e ||^2$ [*Joliffe*, 2002].

[16] In practice, a PCA is first performed on no-noise spectra, it provide the most important spectral features present in IASI observations so that the resulting eigenspectra contain only signal information and are not used to describe noise. The orthogonal components that would appear when noise is introduced in the data are not considered here because they would more likely be related to noise, by design. In the operational stage, observed spectra, y, are projected into the regular subspace of the first components, describing the real variability of IASI spectra (we will comment on how to choose N in the following). In the resulting compression h, the variability attributed to the instrumental noise is then partially suppressed. The compression h can be directly used in a retrieval scheme, or it can be uncompressed to obtain $\hat{y}$, the spectrum partially denoised.

[17] We could have used a Wiener filter or a Noise-Adjusted PCA [*Blackwell*, 2005] but the simple PCA technique used here has already shown its capacities to improve retrieval schemes [*Aires et al.*, 2002b].

[18] The average standard deviation of the instrument noise in IASI measurement is represented by the black line in Figure 2. This NEΔT (Noise Equivalent temperature difference) noise depends on both the frequency and the measured brightness temperature [*Aires et al.*, 2002a]. (Random vibrations of the "porch swing" mechanism used to vary the path lengths of Michelson interferometers like IASI introduce nonnegligible correlated errors in the radiance spectrum. Since we have no information on these correlated noises, no correlations are used in this study. If such information were given, these correlations could be introduced in the learning process.) Some spectral regions, in particular in the third band (2000–2760 cm^{-1}), have very important noise characteristics with levels reaching few Kelvins. As commented earlier, the PCA compression can reduce this instrument noise. The dotted line in Figure 1 represents the average (over the full IASI spectrum) denoising error $< \hat{y}_\eta - y >$ (compressed and then uncompressed noisy spectrum, $\hat{y}_\eta = \overline{F}^{-1} \cdot \overline{F} \cdot (y + \eta)$, minus no-noise spectrum, y) with respect to the number of PCA components used for the compression. After a decrease of the error with increasing PCA number due to a better compression, the denoising error increases. This increase of the denoising error for an increased number of components results from a more accurate representation of the noise. It is more optimal to use a good compromise of PCA components in order to represent well the signal (temperature and water vapor information) and not the instrument noise. The optimal

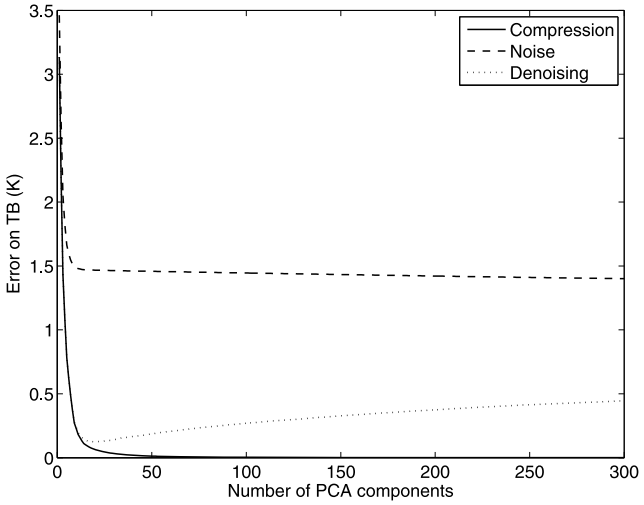

Figure 1. Compression errors: solid line shows the compression errors of no noise spectra $<\hat{y} - y>$, dashed line shows the compression errors of the noisy spectra $<\hat{y} - y>$, and dotted line shows the denoising errors $<\hat{y}_\eta - y>$ where η is the number of components used to retrieve the spectrum (please refer to the text).

number of components for denoising purpose is about 30: The minimum reached by the dotted curve in Figure 1 is about 20 but 30 components are kept to avoid the suppression of potentially informative components (the NN is able to disregard nonrelevant input information). This optimum number of components depends not only on the spectral characteristics of the IASI observations, but also on the noise level, and on the data set used to perform the PCA and the resulting statistics. Asymptotically (i.e., $N = 8461$), the compression error $< \hat{y} - y >$ converges to zero (perfect representation of the no-noise IASI spectra), but the denoising error converges to the instrument noise (perfect reconstruction of the noisy spectra). It should be noted that for retrieval purpose, the optimal number of components can be different: the retrieval scheme can exploit higher-order components.

[19] The dashed line in Figure 1 represents the compression error of the noisy spectra $< \hat{y}_\eta - y_\eta >$. This error decreases first sharply because of a better representation of the signal. Then, each PCA component represents, little by little, a small part of the instrument noise which explains the very small decrease.

[20] Figure 2 shows the spectral statistics of the denoising errors on the atmospheric database. Using only 5 components (blue line) is not enough: In the first and second spectral bands, the denoising error is still often larger than the instrumental noise. However, it is shown that the third band is already considerably denoised (0.5 K of RMS instead of more than 2 K). The use of 30 components for the compression/denoising has excellent statistics, with levels lower than 0.1 K for a lot of spectral regions. We see how the scheme is able to retrieve the signal part (i.e., no-noise spectrum) in a noisy observation. This is particularly true for high noise level spectral regions like 2495–2500 cm^{-1} where the scheme has used the information of flat spectrum to avoid the oscillations due to the instrument noise. The denoising error in the region between 2300 and 2400 cm^{-1} is still at about 0.5 K probably due to the contamination of other atmospheric constituents such as NO_2. With 200 components, the denoising error is bigger than with 30 components, confirming that the higher-order components are used by the PCA to code errors and not valuable signal.

[21] As shown in Figure 1, the best compromise between the compression error, requiring a large number of components, and a denoising error, requiring the limitation of

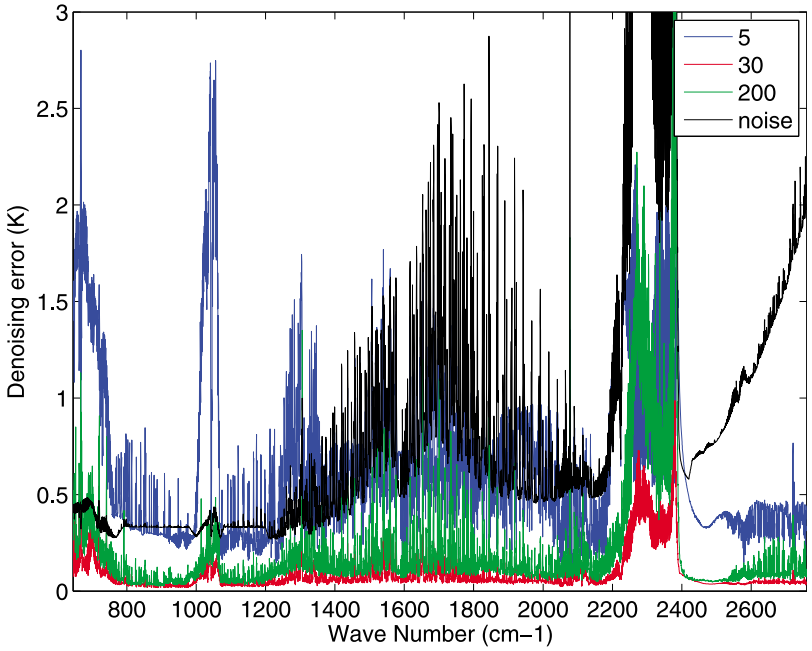

Figure 2. Spectral statistics of the denoising error when using 5, 30, and 200 PCA components to represent noisy spectra. The instrumental noise (in black) is represented for comparison purposes.

the number of components used so as to avoid representing the noise is about 30 components. This compromise is good, of course, only in a statistical sense. If a spectral region is of particular interest (because of a specific constituent absorption), the denoising of the entire spectrum is not necessarily the optimal solution. The particular spectral region may be neglected statistically and the compression/denoising scheme might not represent well this information. Control of errors for each spectral region is crucial if such spectral regions are of interest. Then, even if 30 components seem to be the perfect compromise for compression/denoising of the whole spectrum, it might be useful to keep some raw channels, without the application of PCA. This would be the case for example for trace gas retrievals.

2.4. PCA Components

[22] In order to validate our PCA representation of the IASI measurement spectra, the correlation between these components and the temperature (Figure 3, top) and water vapor (Figure 3, bottom) at the various atmospheric levels is shown. Water vapor is hereafter defined by the relative humidity in %. The sign of these correlations is not relevant, only the amplitude is relevant in a PCA analysis of the IC. The first components include information on both temperature and water vapor. It is well known that PCA mixes in its components the variability of different variables [*Aires et al.*, 2002c]. Correlations reach the 0.6 level. After the 10th component, the information essentially concerns water vapor. The components point more precisely at some vertical layers for water vapor than for temperature suggesting that temperature profiles are smoother (with less degrees of freedom) than water vapor. Some of the components (16, 17, 18) do not seem to be significantly related to either temperature or water vapor, they could be correlated to other surface of atmospheric variables such as ozone.

[23] In order to look more precisely into the correlation between PCA components and temperature or water vapor information, Figure 4 represents the correlation profiles. Again, the sign of correlation is not important. This type of "metric" is similar to weighting functions or Jacobians, but it involves statistics that are directly related to the inversion algorithm to be developed. The first component (crosses) is highly correlated to temperature with a correlation close to one in the troposphere and an inversion at 300 hPa, and a correlation up to 0.5 for water vapor again with an inversion at 300 hPa. Like most of the first components in a PCA, it describes a general shape for temperature and water vapor. The second component (continuous line) has not much information on temperature but a water vapor information between 800 and 300 hPa. Component 3 (dashed line) has an information on stratospheric temperature. The fourth component pics at 300 hPa for temperature but indicates a contrast between 700 and 300 hPa for water vapor. The fifth component has a similar profile for temperature and water vapor and indicates an anomaly between 50 and 300 hPa. Component 5 describes an anomaly in both temperature and water vapor between 200 and 50 hPa.

[24] An inversion scheme is constrained by the relationships between inputs and outputs, this has been investigated in previous Figures 3 and 4. The relationships between the inputs is also an important component of the information content analysis. The correlation matrix among the inputs of the retrieval schemes has been represented in Figure 5: inputs are ordered with first, the 15 AMSU-A microwave channels, then the 5 MHS microwave channels and the first 17 infrared PCA components from IASI. Very high correlation exists between AMSU-A and MHS observations. Since both of them provide information on the temperature and water vapor, this is not surprising. The IASI components for IASI measurements have no correlation, by design, the PCA components are uncorrelated. First IASI components are highly correlated with microwave observations because both are highly correlated to water vapor and temperature. The PCA compression of the inputs could have been done on both IASI and AMSU-A/MHS observations. However, the compression and denoising step is most important for the IASI instrument, the retrieval scheme is able to deal with the 20 channels from microwave data and with their associated instruments noise. Furthermore, the interpretation of the PCA is easier when using only the infrared measurements from IASI.

3. Retrieval Approaches

[25] Three retrieval methods are considered to merge the microwave and infrared observations and to test their synergy: the k nearest neighbors (k-NN), the linear regression (LIN), and the neural network (NN). They are selected for their expected different abilities to deal with synergy: the first one is not directly adapted to synergy, the second one should slightly benefit from synergy, and the third one is very well adapted to the merging of information.

3.1. General Strategy

[26] At present, the most commonly used technique that is able to deal with multispectral observation is the "assimilation". This approach is largely used in NWP centers. The basic principle of assimilation is to combine NWP model outputs with satellite observations; the weighting of these two sources of information is inversely proportional to their respective uncertainties. Theoretically, the retrieved atmospheric profile f is given by *Aires* [2011, equation (2)]. Since the involved variables are multivariate and linked on to the other, the covariance matrices (errors, geophysical variables, satellite observations) are truly important and will appear in any synergetic retrieval methodology. The major limitation of assimilation is that it mixes the information coming from the satellites together with a numerical forecast: What is measured is not the IC of a particular observation but its marginal impact on the whole forecasting system, the excess of information compared to the whole forecast model and its already assimilated data. With the improvement of the forecast models and the assimilation schemes that use a lot of the already existing observations, this marginal impact can be negligible even if the true IC is not. What is really needed to measure the absolute synergy of multiple observations is an absolute not a relative measure of the IC.

[27] Furthermore, it is convenient to obtain independent satellite retrievals so that numerical models can be properly validated. Independent retrieval techniques able to use the synergy of multispectral observations need to be defined. This retrieval scheme must include the whole necessary processing chain: data compression, denoising, fusion, a calibration procedure, a coherent radiative transfer model

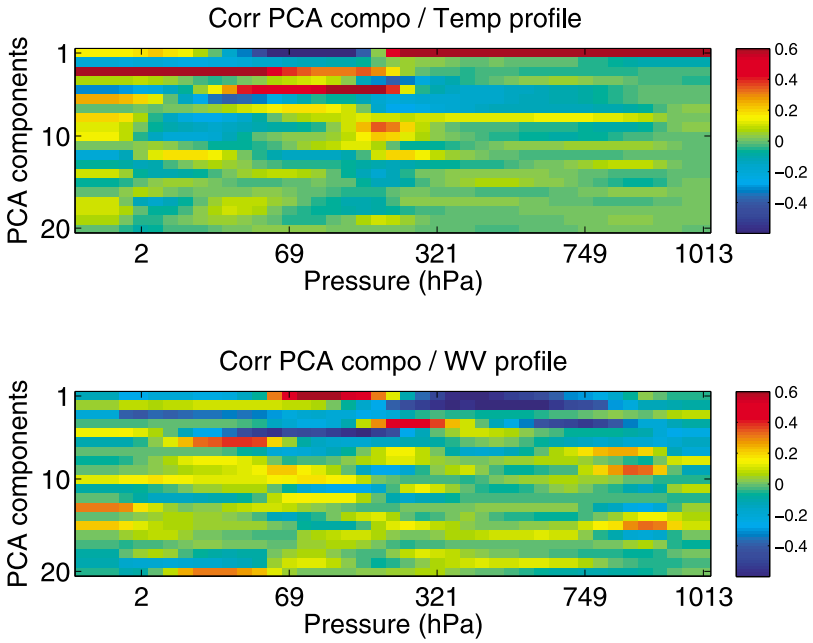

Figure 3. Correlation matrix between the extracted IASI PCA components and the (top) temperature and (bottom) water vapor atmospheric profiles.

across the spectral range, and of course, a multivariate and nonlinear inversion method flexible enough to be adapted to a large range of applications.

[28] Two basic strategies can be used: the retrieval of the different parameters can be hierarchical or simultaneous. In a hierarchical scheme, a major atmospheric parameter (e.g., the temperature) is first retrieved by using all available satellite observations. Then, this retrieved parameter is used together with the satellite measurements for the subsequent retrieval of another variable (e.g., water vapor). This approach uses the fact that some atmospheric parameters are dependent on others and that some specific retrieval algorithms need to follow this dependency structure. As an example, within the International Satellite Cloud Climatology Project (ISCCP) [*Rossow and Schiffer*, 1999], using IR observations from both geostationary and polar satellites, the cloud presence is first analyzed and then, for clear scenes, the surface skin temperature is estimated. An inconvenience of this scheme is the "cascade of errors" in the suite of retrieval algorithms; it is also difficult to obtain a valid characterization of the uncertainties on the retrieved products with such a strategy.

[29] The other approach consists in performing the multivariable retrieval at the same time in the algorithm. One advantage of this strategy is that the uncertainty characterization is easier [*Aires*, 2004; *Aires et al.*, 2004a, 2004b]. The solution is determined simultaneously for each parameter: this is preferable from an optimization point of view than a solution built piece by piece. The optimization process involves more variables and relationships and then might be more difficult. However, it is easier to obtain a good compromise when choosing a solution that satisfies all the satellite observations.

[30] This strategy calls for a (1) multivariate and (2) nonlinear approach. First, the multivariate aspect does not mean that we just put together the various parameters in the retrieval scheme; there is a necessity to use the potential synergy coming from the complex relationships that exist among the observations, among the geophysical variables, and between observations and variables. These relationships are described with covariance matrices; in the NN theory, they can also be described implicitly by an exhaustive data set [*Aires and Prigent*, 2007]. Second, the nonlinear aspect corresponds to the need for the retrieval algorithm to adapt to the situation. The relationship between the water vapor

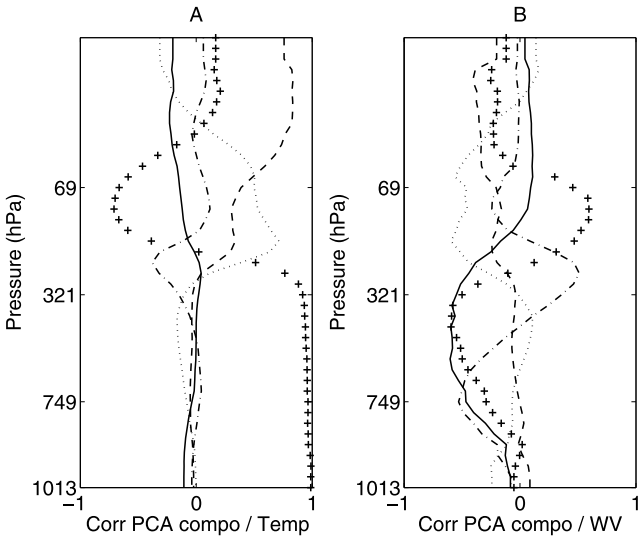

Figure 4. Correlation profiles between the IASI PCA components and (a) the temperature or (b) the water vapor profiles. Crosses are for PCA component 1, solid line is for component 2, dashed line is for component 3, dash-dotted line is for component 4, and dotted line is for component 5.

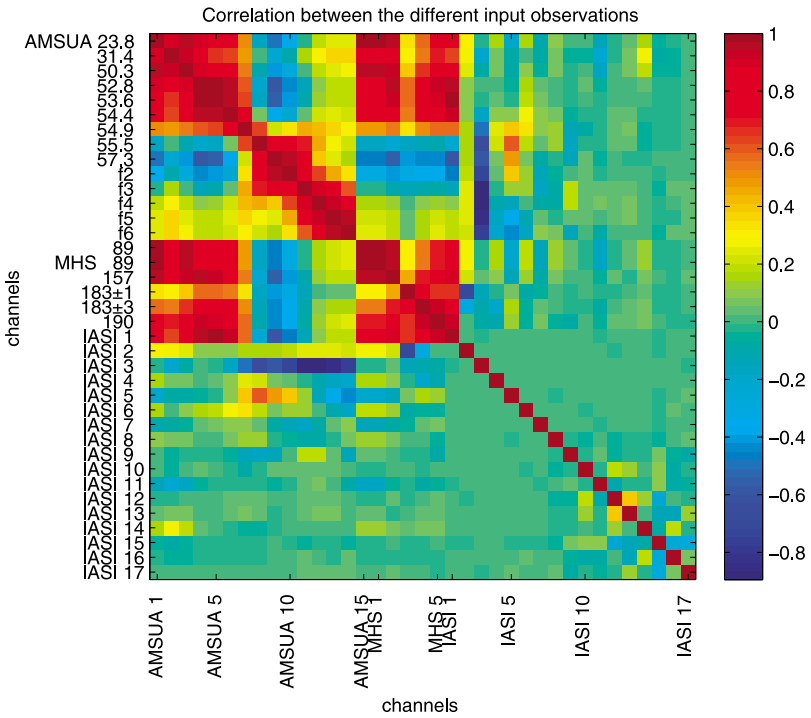

Figure 5. Correlation matrix between the inputs of the retrieval schemes: the 15 AMSU-A microwave channels, the 5 MHS microwave channels, and the infrared PCA components from IASI.

sensitive channels and the atmospheric water vapor will change based on the atmospheric situation (H_2O weighting functions can move vertically for dry or wet situations). Furthermore, we saw in the work by *Aires* [2011, section 2.7] that saturations and interactions can be part of a nonlinear model and that synergy from multiple observations can be extremely beneficial for the retrieval. The problem being complex, highly nonlinear, with a large number of degrees of freedom, and with a profusion of channels used in conjunction, the NN technique is an adequate candidate. However, we will consider two other retrieval techniques, the *k*-NN and linear regression approaches, for comparison purpose.

3.2. The *k* Nearest Neighbors Approach

[31] In the *k*-NN retrieval approach, a "reference" data set, $\mathcal{R}$, is built that includes a number E of situations described by a set of geophysical variables (i.e., the variables to be retrieved) and by a set of the associated satellite observations: $\mathcal{R} = \{(GEO^e, TB^e); e = 1, \ldots, E\}$. The Brightness Temperatures (TBs) can be real observations or radiative transfer simulations (the later is chosen in this paper). Once this data set is available, any new satellite observation TB_N is compared to the satellite observations in the reference data set $\mathcal{R}$. The k closer situations in $\mathcal{R}$ are used to define the inversion. Few approaches can be used: if only the closest situation is taken ($k = 1$), this scheme is a pattern recognition algorithm. Sometimes, the situations in the reference data set $\mathcal{R}$ are ordered in some way; in this case, it is easiest to retrieve the closest situation. The retrieval scheme is then called a LUT (look-up table) inversion.

[32] When $k > 1$, the solution GEO_N is defined as a weighted average of the k closest situations, the weights been defined, for example, as the inverse of the distance to the observation:

$$GEO_N = \frac{1}{\sum_{e=1}^{k} \frac{1}{d(TB_N,\ TB_e)}} \sum_{e=1}^{k} \frac{1}{d(TB_N, TB_e)} GEO^e$$

where $d(TB_N, TB_e)$ is the distance in the TB space. The higher k the smoother is the behavior of the inversion scheme. The parameter k controls the regularization of the inversion problem: when k increases, the bias of the model decreases, but its variance error increases (and the opposite when k gets smaller) following the bias variance dilemma [*Geman et al.*, 1992]. We will see in the following how to optimize the parameter k.

[33] The *k*-NN algorithm is a nonlinear model. Its behavior and retrieval statistic would converge to the neural network model if enough samples are available in $\mathcal{R}$: the entire solution space would be sampled with a level of precision only limited by the instrument noise [*Rydberg et al.*, 2009; *Jiménez et al.*, 2007; *Evans et al.*, 2005], but the computational time could become prohibitive of course. It is also a truly multivariate method. However, it should be noted that the method is entirely based on a distance in the satellite observation space. This distance gives the same weight to each of the TB inputs and no IC analysis on the GEO is used. If some nonpertinent channels are included in the TB space, they will not add any useful information and even worse, they can contaminate the actual pertinent information. The noninformative TBs pollute the information and therefore, the retrieval. In the ideal case, the distance d should only compare the appropriate channels, it

should even be able to weight by the true significance for the retrieval. This is difficult to do using a general distance but some idea can be explored. (Some special features can be introduced in the general framework of EOF analysis. The study of *Blackwell* [2005] on the radiance denoising is similar to the approach by *Aires et al.* [2002b], but Blackwell introduces the so-called "projected PCA" that takes into account the retrieval model in the construction phase of the EOF representation. The retrieval is performed using a linear regression, and the projected EOFs extract from AIRS spectra the components that are more related to the temperature profile. This allows for a higher level of compression for equivalent retrieval accuracy.) The rotation of EOFs is used in Factorial Analysis in order to extract more physically meaningful components. NN have been used as nonlinear compression tool as well [*Bishop*, 1996].

[34] Another comment is that this technique is very slow: the search for the closest situations in the reference data set $\mathcal{R}$ is very time consuming, each new observation TB_N being compared to all the *TB*s in the reference data set. It is not our purpose here to optimize the implementation of the retrieval algorithms. However, it should be pointed out that the approach to build the reference data set using a hierarchical clustering (section 3.2) of *Aires* [2011] could be used to limit the domain of search for each new observation TB_N: the search would be conducted only on the second-generation prototypes associated to the closest first-generation prototype.

3.3. Linear Regression

[35] Multiple linear regression is a very simple and classical technique. For more details, details to present this method, see any elementary statistical textbook. We just mention here that the LIN model is truly multivariate. Furthermore, contrarily to the *k*-NN approach, only the pertinent information is used from the inputs for the retrieval of a particular output. This means that meaningless information will not pollute the retrieval. By definition, and contrarily to *k*-NN, the LIN approach is not nonlinear and can suffer from outliers, saturations, interactions between inputs, and any nonlinear behavior. This technique is often used as a reference test for the NN method presented next.

3.4. Neural Networks

[36] NN techniques have proved very successful in developing computationally efficient algorithms for remote sensing applications. The Multi-Layered Perceptron (MLP) model [*Rumelhart et al.*, 1986] is selected here. It is a nonlinear mapping model: given an input *TB*, it provides an output *f* in a nonlinear way. In this paper, a NN model with only one hidden layer will be considered. The MLP model is defined by the number of input neurons (i.e., the size of the inputs, number of channels), the number of outputs (i.e., the size of the geophysical variables to retrieve $= 3 \times 43 = 86$) and the number of neurons in the hidden layers that control the complexity of the model. A study will be conducted to define the optimal number of neurons in the hidden layer. A balance needs to be found: Too many free parameters in the model can conduct to the overlearning (overparameterization) leading to degraded generalization properties. On the contrary, too few free parameters will yield underparameterization and bias error of the model.

[37] The NN is trained to reproduce the behavior described by a database of samples composed of inputs (i.e., the real observation *TB*s) and their associated outputs (i.e., the geophysical variables *f*), for $e = 1, \ldots, N$ with N the sample number in the training database. Provided that enough samples (TB^e, f^e) are available, any continuous relationship, as complex as it is, can be represented by a MLP [*Hornik et al.*, 1989; *Cybenko*, 1989]. The *TB* simulations presented by *Aires* [2011, section 3.4] are used as the learning data set. This learning data set includes some sources of errors, in particular, the radiative transfer errors. The learning algorithm used to train the NN is the classical Back-Propagation algorithm. This optimization technique has long proved its efficiency such problems. The quality criterion to maximize during the learning of the NN has to be carefully chosen. In particular, since we perform simultaneous retrieval of variables with completely different ranges of variability (few orders of magnitude, with very different probability density functions), the quality criterion needs to balance the weight of each of the retrieved variable. In this study, the outputs of the NN are the temperature and water vapor atmospheric profiles.

[38] The NN are trained on a learning data set of 9000 atmospheric situations, and the ability of the model to generalize to independent samples is monitored on a testing data set (500 situations). The learning is stopped when this generalization errors cease to decrease. In order to avoid the "learning" of the testing data set (i.e., many model experiments on some testing data set imply a bias of the selection of the better model toward the testing data set), a third database, i.e., the validation data set (composed of 500 situations), is used to estimate more realistic NN error estimates. There three error estimates are close enough, which demonstrate that there is no overtraining in the learning of the retrieval models. The results presented in section 4 are validation errors.

[39] The learning of a NN is sometimes sensitive to the initial conditions (i.e., the initial weights in the NN) so multiple training runs with different initializations are used to avoid this problem. Furthermore, the initialization is performed using the standard Nguyen-Widrow approach.

[40] The NN method is consistent with the Bayesian formulation. It can include all the a priori information used by the algorithm: the noise is taken into account, the distributions are not limited to Gaussians and the uncertainty on retrievals can also be estimated. The advantages of the NN over other methods are the fast processing time and the flexibility relative to analytical definition of a priori information.

[41] In addition, the NN is a remarkable model for information merging. Terms such as $x_1 * x_2$ are sometimes introduced in regression models $y = f(x_1, x_2)$ to allow for the interaction of two inputs. These interaction terms are nonlinear and can directly be represented by NN architectures and their use is optimized during the learning process. Saturation effects also play an important role when data is combined: an output can be sensitive to an input for a particular range and to another input for a different range. In order to represent this behavior, saturation effects need to be used in the regression model. This is well represented by the sigmoid functions in the NN architecture.

[42] Since the NN performances will be compared using different numbers of inputs (only the IR, only the MW or

both), it is important to address the relative stability of the training. A NN with more inputs will have more free parameters and more information to exploit; this can complexify the NN training, and the learning step can become much longer. An overparameterization can led to overtraining if the learning process is not regularized. Good testing and validation data sets, stopping criterion, and the multiple initialization of the weights for the learning tend to reduce this problem. One way to control for this effect would be to always keep the number of inputs fixed by adding, when necessary, inputs that represent noise. Then the network topologies would be the same in each case, and the network with fewer inputs would not gain an advantage due to its fewer free parameters. Experimental tests (not shown) suggest that the impact on the retrieval results is negligible, the addition of noise inputs do not perturb the retrieval, the NN is able to "learn" to disregard these inputs. This means that the learning process is well regularized, the potential overparameterization has no impact in part because we have enough samples.

4. Retrieval Results

[43] All results presented in this section are for the test database [see *Aires*, 2011, section 3] composed of 500 atmospheric situations. This test database is not included in the learning database used to calibrate the retrieval methods so it can be used to measure the generalization capacities of the various retrieval schemes.

4.1. Experimental Conditions

[44] It has been shown in section 2 that the PCA technique can be used to compress and denoise the observed IASI spectra. Such PCA components can also be used to perform a PCA retrieval, using the three inversion techniques of section 3: *k*-NN, LIN and NN approaches. It was found that about 30 components is the statistical optimum number of components for the denoising of the whole IASI spectra (section 2.3) but this number is not the optimum for the regression. The retrieval technique can find some information in higher-order components. Tests have been conducted with various numbers of PCA components as inputs of the retrieval scheme (not shown). The results of these experiments confirm that 100 components is the right number, in agreement with the literature on this subject [*Liu et al.*, 2009]. As a consequence, in this paper, 100 PCA components are used to represent the IASI spectra in the inputs of the retrieval schemes.

[45] During the learning and the testing of the retrieval methodologies, a simulated noise is used on IASI spectra. This noise is simulated using the instrument specification, i.e., following a Gaussian distribution with a standard deviation, without interchannel correlation. A NEΔT dependency on the observed scene is also taken into account [*Aires et al.*, 2002a]). The noise is randomly simulated every time that the samples are used in the learning or the testing process, in order to avoid the overfitting of the retrieval on particular noisy data. The introduction of this noise is very important: the retrieval scheme is able to "learn" which channels are the more reliable and which suffer from important instrumental noise.

[46] In Table 1, the number of neurons in the input, hidden and output layers is given for the various NN models. The total number of parameters is also provided for the NN and the LIN models (the *k*-NN method does not involve a parameterized model). As commented in section 3.4, the overparameterization (i.e., too many degrees of freedom in the model) can led to training difficulties such as the overtraining, for both LIN and NN methods. The number of samples in our learning data set (9000 samples) could be considered to be too limited for the size of our models The number of samples need to be limited because the radiative transfer simulations can be very time consuming, especially when they need to be performed in various configurations (such as the scanning angle). However, the introduction of noise in the samples during the learning stage, called "input perturbation", is equivalent to artificially increase the size of the learning data set because each sample is used multiple times with a different input noise. This is not the only advantage of this approach: during the learning, the level of certainty for each is the inputs is taught to the NN. This method avoids overtraining, and high gradients that often appear when too many parameters are in a NN are avoided. In fact, this simple input perturbation has been shown theoretically to be a Tikhonov regularization [*Bishop*, 1996; *Aires et al.*, 1999].

[47] The use of a second hidden layer could introduce more nonlinearity in the NN model. We tend to avoid these two hidden layers for multiple reasons: it does not make significant differences in many cases, what can be gained from the additional complexity representation of the two hidden layers can be lost because of overtraining/overparameterization problems, and the learning processes become longer so fewer configurations can be tested. This is an illustration of Occam's Razor principle [*Pearl*, 2010]. In this paper, few configurations were however tested with an additional hidden layer and no improvement in the retrieval statistics was observed.

4.2. The *k* Nearest Neighbors Method

[48] In this section, the *k*-NN retrieval method presented in section 3.2 is tested using the databases described by *Aires* [2011, section 3]. The impact of the number of neighbors in the quality of the retrieval algorithm is illustrated in Figure 6: The RMS error decreases first when the number of neighbors increases. A optimal number is found to be seven. The RMS error increases when too many neighbors are used in the *k*-NN algorithm: it has been seen in section 3.2 that the higher *k* the smoother the behavior of the inversion scheme. This means that small inversions cannot be retrieved.

[49] Figure 7 represents the root-mean-square (RMS) error statistics for the *k*-NN retrieval of the temperature profile. These RMS errors represent the difference statistics between the retrieved and real atmospheric profiles, the goal is to minimize these errors along the pressure axis, from the surface to the top of the atmosphere. These retrievals are performed using MHS only, AMSU-A only, MHS + AMSU-A, IASI only, IASI + AMSU-A + MHS and IASI + AMSU-A + MHS observations. The two last configurations are for the retrieval of temperature only and the simultaneous retrieval of temperature and water vapor. The results for these two configurations are almost identical. Indeed, for the *k*-NN algorithm, there is no difference between separate or joint retrieval of water vapor and temperature: selected samples will be the *k* nearest neighbors in measurement

Table 1. Number of Neurons in the Input, Hidden, and Output Layers of the NN[a]

Configuration	Number of Inputs	Number of Neurons in the Hidden Layer	Number of Outputs	Number of Parameters for NN	Number of Parameters for LIN
MHS	5	20	43	1,023	215
AMSU-A	15	30	43	1,813	645
MHS + AMSU-A	20	30	43	1,963	860
IASI	100	50	43	7,243	4,300
IASI + MHS + AMSU-A	120	50	43	11,253	5,160
IASI + MHS + AMSU-A[b]	120	70	86	14,576	10,320

[a]The total number of parameters is also provided for the NN and the LIN retrieval model, for comparison purpose.
[b]Simultaneous retrieval of temperature and water vapor atmospheric profiles.

space and the retrieved valuers are the averaged geophysical samples, without any interaction between the two.

[50] The AMSU-A microwave observation seems to be the best independent information for the retrieval of temperature, better than MHS observations (this is expected since it is a water vapor instrument) but also better than the high-resolution interferometer measurements from IASI. The combination of AMSU-A, MHS, and IASI data improves by about 1 K the retrieval in layers lower than 800 hPa and slightly degrades the AMSU-A retrieval at higher altitudes. As a consequence, the synergy cannot be said to be an improvement for the *k*-NN retrieval of temperature. This result might be different if a much larger data set was used.

[51] The RMS error statistics for the *k*-NN retrieval of the water vapor profile using MHS, AMSU-A, MHS + AMSU-A, IASI, IASI + AMSU-A + MHS and IASI + AMSU-A + MHS observations are shown in Figure 8. Again, the two last configurations are, first, for the retrieval of temperature only and, second, the simultaneous retrieval of temperature and water vapor. It can be seen that the merging of all the information is the best retrieval for most of the atmospheric layers. However, the benefit can be very marginal and the MHS-only retrieval can be better for some atmospheric layers.

[52] To explain this behavior, it is important to note that, in the *k*-NN retrieval method, the pattern recognition is performed in the space of the inputs, independently from the outputs. When using the simultaneous observations from AMSU-A, MHS, and IASI instruments, the pattern recognition step is not changed for temperature or water vapor retrievals. When all instruments are used for the retrieval of the temperature profile, the MHS channels (mostly sensitive to water vapor) have the same weight as the IASI or AMSU-A observations in the pattern recognition step. This means that the *k*-NN retrieval would work better without the MHS information that can perturb the temperature retrieval. In order to adapt the *k*-NN approach to the satellite information fusion, it would be necessary to adapt the pattern recognition approach to better take into account the output space, i.e., the space of the geophysical variables. It would also be necessary to adapt the distance in the input space since, in this application, raw brightness temperatures are combined with PCA components. This will be the subject of a forthcoming study.

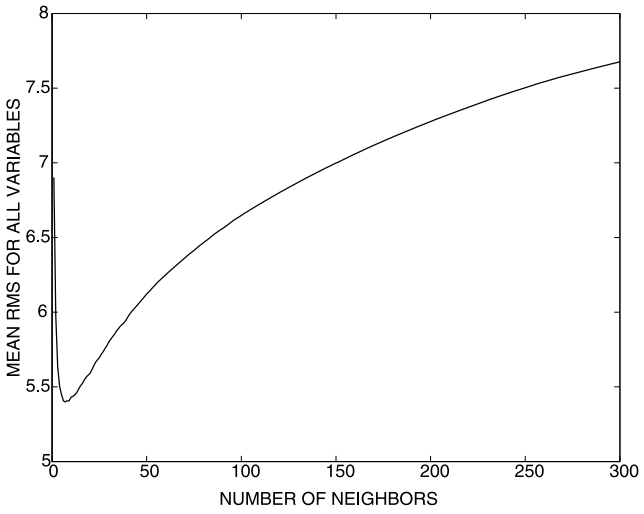

Figure 6. RMS error of the *k*-NN retrieval scheme when increasing the number of neighbors *k*.

4.3. Linear Regression

[53] The RMS error statistics for the linear regression retrieval of the temperature profile using MHS, AMSU-A, MHS + AMSU-A, IASI, IASI + AMSU-A + MHS and IASI + AMSU-A + MHS observations are presented in Figure 9. Again, the two last configurations are for the retrieval of

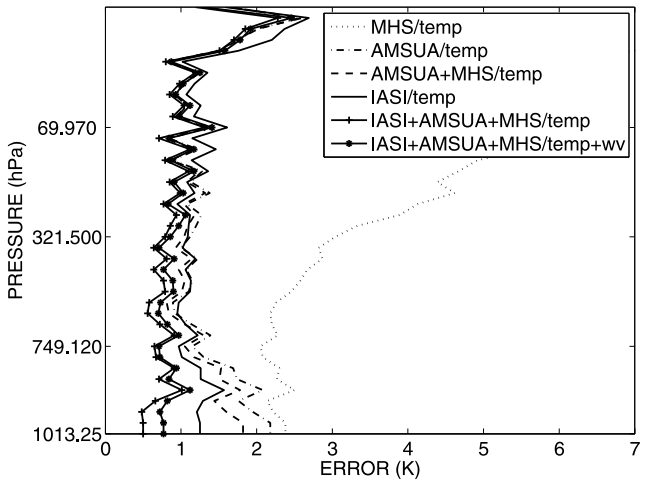

Figure 7. Root-mean-square error for the *k*-NN retrieval of temperature using MHS (dotted line), AMSU-A (dash-dotted line), MHS + AMSU-A (dashed line), IASI (solid line), IASI + AMSU-A + MHS (retrieval of temperature only, solid line with crosses), and IASI + AMSU-A + MHS (simultaneous retrieval of temperature and water vapor, solid line with asterisks) observations.

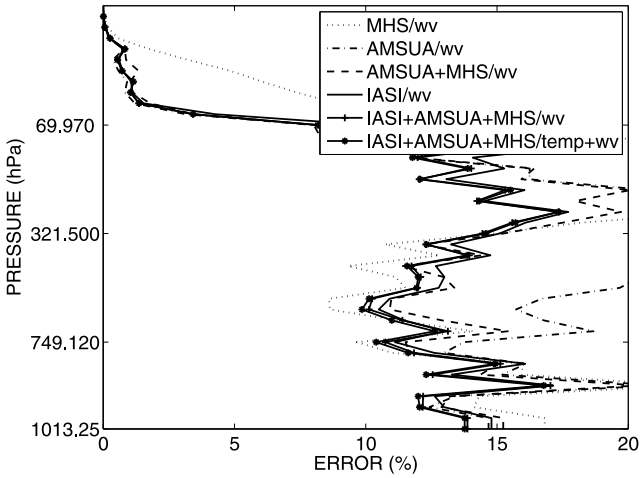

Figure 8. Root-mean-square error for the *k*-NN retrieval of water vapor using MHS (dotted line), AMSU-A (dash-dotted line), MHS + AMSU-A (dashed line), IASI (solid line), IASI + AMSU-A + MHS (retrieval of temperature only, solid line with crosses), and IASI + AMSU-A + MHS (simultaneous retrieval of temperature and water vapor, solid line with asterisks) observations.

temperature only and the simultaneous retrieval of temperature and water vapor. As expected, these two configurations have similar statistics, the linear retrievals of each geophysical variable (temperature or water vapor in any atmospheric layer) are independent, so there is no gain or loss in retrieving simultaneously temperature and water vapor. Not surprisingly, MHS is the less informative instrument for temperature, AMSU-A is next. IASI offers a significant improvement in retrieval statistics in the lower atmosphere (up to 700 hPa), AMSU-A being better for layers over 100 hPa. The input information being additive in a linear model, the more information available, the better the retrieval is. As a consequence, the combination of the three instruments provides a synergy effect, improving the retrieval by up to 0.5 K near the surface, and never degrading the retrieval of the best instrument.

[54] Figure 10 represents the RMS error statistics for the linear regression retrieval of the water vapor profile using MHS, AMSU-A, MHS + AMSU-A, IASI, IASI + AMSU-A + MHS observations. Again, the synergy impact is very positive, the RMS error is lower at every atmospheric layer than any individual instrument retrieval. The benefit can be large, especially near the surface where the RMS error can decrease from ~12.5% to ~10% (i.e., a ~20% decrease of the error).

4.4. Neural Networks

[55] Same as before, the RMS error statistics for the different configurations (Table 1) are given in Figure 11 for the NN retrieval. The two last configurations are for the retrieval of temperature only and the simultaneous retrieval of temperature and water vapor. MHS is again the less informative instrument for temperature, and AMSU-A is next. IASI offers some improvement in retrieval statistics in the lower part of the atmosphere (up to 750 hPa). Combining the three instruments improves the results considerably, especially near the surface where the RMS error decreases from 1.2 K (IASI retrieval) to about 0.5 K for the combined configuration. Like for the linear model, the more information available in the inputs, the best the retrieval. The NN is able to select in the inputs the information that is pertinent for the retrieval of temperature, even if water vapor information from MHS contaminates the inversion. The retrieval is not degraded when combining the information. The interesting feature is that the retrieval of temperature is degraded when water vapor is simultaneously retrieved. This will be discussed in section 4.6 about the indirect synergy.

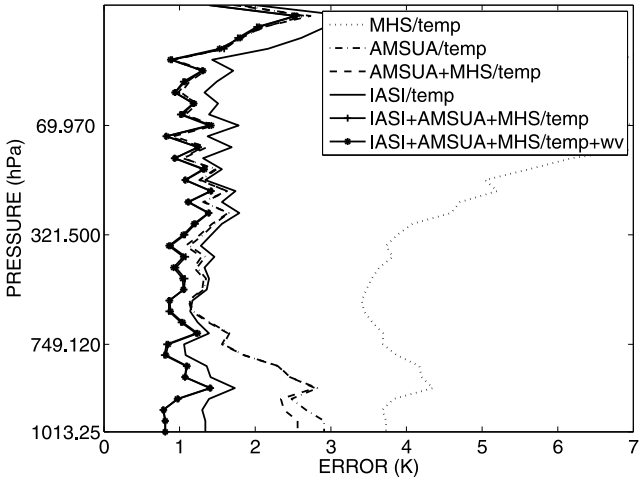

Figure 9. Root-mean-square error for the linear regression retrieval of temperature using MHS (dotted line), AMSU-A (dash-dotted line), MHS + AMSU-A (dashed line), IASI (solid line), IASI + AMSU-A + MHS (retrieval of temperature only, solid line with crosses), and IASI + AMSU-A + MHS (simultaneous retrieval of temperature and water vapor, solid line with asterisks) observations.

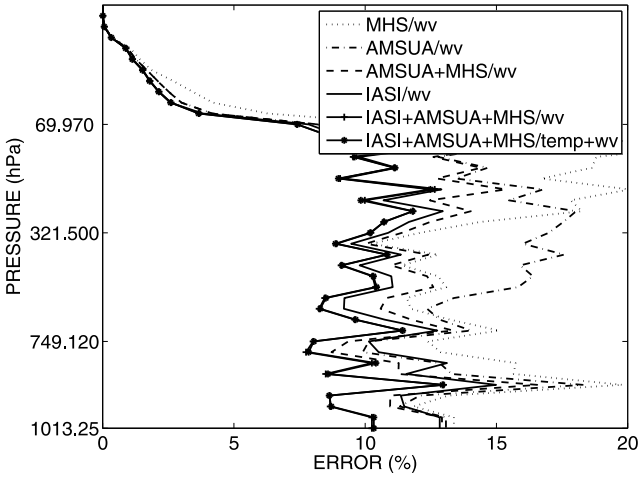

Figure 10. Root-mean-square error for the linear regression retrieval of water vapor using MHS (dotted line), AMSU-A (dash-dotted line), MHS + AMSU-A (dashed line), IASI (solid line), IASI + AMSU-A + MHS (retrieval of temperature only, solid line with crosses), and IASI + AMSU-A + MHS (simultaneous retrieval of temperature and water vapor, solid line with asterisks) observations.

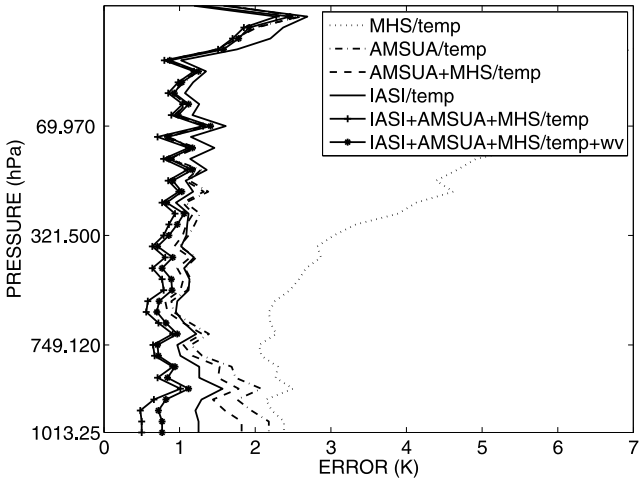

Figure 11. Root-mean-square error for the neural network retrieval of temperature using MHS (dotted line), AMSU-A (dash-dotted line), MHS + AMSU-A (dashed line), IASI (solid line), IASI + AMSU-A + MHS (retrieval of temperature only, solid line with crosses), and IASI + AMSU-A + MHS (simultaneous retrieval of temperature and water vapor, solid line with asterisks) observations.

[56] Figure 12 represents the RMS error statistics for the NN retrieval of the water vapor profile for different configurations. The two last configurations are for the retrieval of water vapor only and the simultaneous retrieval of temperature and water vapor. The MHS information provides more information than AMSU-A for the 750–300 hPa range, but AMSU-A is more efficient in the upper atmospheric layers. The infrared information from IASI is always more informative for water vapor than the microwave measurements. The synergy from the three instruments is always positive with an important decrease of the RMS error (from 10 to 7% RMS error in the lower layers).

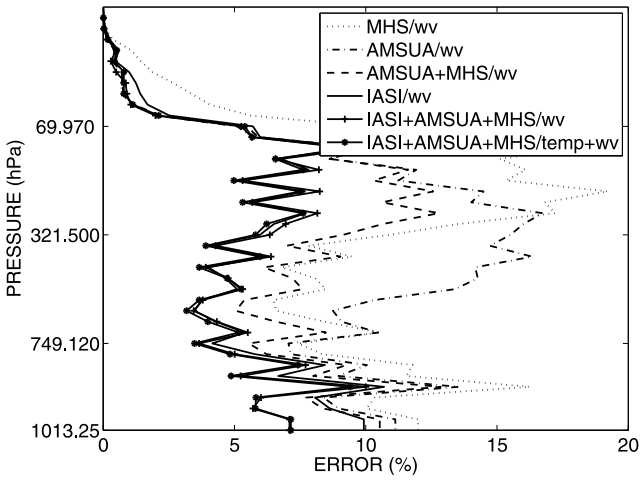

Figure 12. Root-mean-square error for the neural network retrieval of water vapor using MHS (dotted line), AMSU-A (dash-dotted line), MHS + AMSU-A (dashed line), IASI (solid line), IASI + AMSU-A + MHS (retrieval of temperature only, solid line with crosses), and IASI + AMSU-A + MHS (simultaneous retrieval of temperature and water vapor, solid line with asterisks) observations.

[57] Note that the IASI retrieval of both temperature or water vapor from the IASI components described in section 2 is not easy: the information included in the IASI components is mixed in the PCA representation because the higher-order components are more predominantly the (nonlinear) remaining parts of physical components [*Aires et al.*, 2002c]. As a consequence, the convergence of the learning step is more difficult (not shown, see *Aires et al.* [2004b]). When this IASI information is combined to the microwave information from AMSU-A and MHS, the information is more easily related to the geophysical variable, the loss function to minimize during the learning step is better constrained and the learning is much easier. In other words, combining the microwave information to the IASI components regularizes the problem. Regularization of the inverse problem is one of the benefits of combining the information from multiple instruments.

[58] It can be intriguing to obtain indirect synergy: for the type of NN architecture used in this experiment, i.e., a nonrecursive feed-forward network, each neuron in the output layer can be suppressed, the behavior of the other neurons would not change. However, due to the back propagation of errors during the learning of the NN, the outputs impact the hidden layer, and the connections from the input to the hidden layer. The hidden layer is often considered to be an efficient representation of the inputs. This efficient representation depends on the NN outputs and the errors measured during the learning. As a consequence, introducing pertinent outputs, with a particular correlation structure among them, will then have an impact on the general behavior of the NN, and some synergy can be expected.

4.5. Comparison of the Retrieval Methods

[59] The comparison of the three retrieval methods is illustrated in Figure 13 for the temperature and in Figure 14 for the water vapor. These results correspond to the case

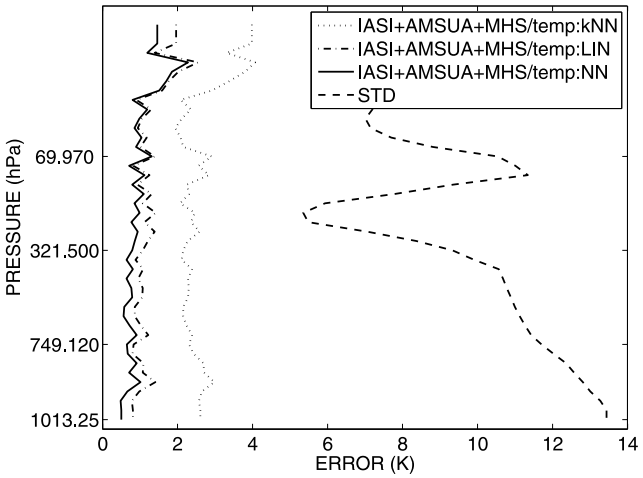

Figure 13. Root-mean-square error for the retrieval of temperature using IASI + AMSU-A + MHS observations for *k*-NN (dotted line), linear regression (dash-dotted line), and neural network (solid line) retrieval methods. The standard deviation of water vapor profiles is represented as a dashed line for comparison purposes.

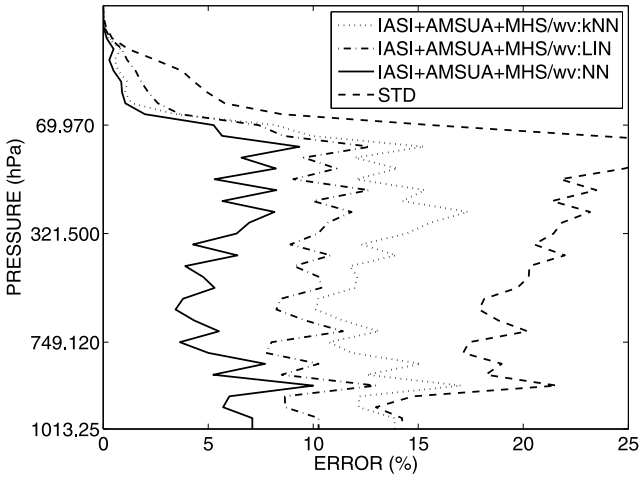

Figure 14. Root-mean-square error for the retrieval of water vapor using IASI + AMSU-A + MHS observations for *k*-NN (dotted line), linear regression (dash-dotted line), and neural network (solid line) retrieval methods. The standard deviation of water vapor profiles is represented as a dashed line for comparison purposes.

where the three instruments MHS, AMSU-A and IASI are combined. The standard deviation corresponding to the natural variability of the geophysical parameters to retrieve is also represented for comparison purpose. Figure 13 clearly shows that the retrieval of temperature is acceptable for the *k*-NN method, but that the LIN and NN methods performance is better. Furthermore, the NN is always more precise than the LIN regression.

[60] The situation is almost identical for the retrieval of water vapor (Figure 14). The *k*-NN is the worse retrieval with more than 10% RMS error, the LIN method yields about 10% RMS error. The NN outperforms the other two methods, with very interesting levels of accuracy: The RMS errors are close to 5% and are rather uniform from the surface to 100 hPa. Figure 14 clearly shows that the NN makes the difference with the other two methods when the relationship from the satellite observations to the geophysical parameter to retrieve (i.e., the water vapor) is complex and nonlinear; the impact is less important but still exist for simpler problems such as the retrieval of the temperature.

[61] In the work by *Divakarla et al.* [2006] the temperature and water vapor retrieved with the AIRS (Atmospheric InfraRed Sounder) instrument, which has similar characteristics to IASI, has been validated against radiosonde measurements and forecasts. The results obtained are a little bit higher than our theoretical estimates, around 1–1.5K for temperature and more than 20% for water vapor, but these differences are expected since it concerns real observations: (1) The independence of channel noise is probably not realistic; this degrades the results. (2) The coincidence of the satellite observations with the radiosondes also introduces errors.

4.6. Additive Synergy Measure

[62] The synergy factor is defined as the ratio between the best individual instrument and the multiple-instrument RMS retrievals. When this factor is equal to 100%, there is no synergy: using simultaneously all the instruments do not improve the results compared to using only the best instrument. When the factor is higher than 100%, the synergy is enhanced. A factor of 120%, for example, represents a decrease by 20% of the retrieval error statistics. The factor can also be lower than 100%: this appears when using simultaneously all the instruments degrades the retrieval statistics compared to the best individual instrument retrieval.

[63] The synergy factor is represented in Figure 15 for the retrieval of the temperature profile using the three methods (*k*-NN, LIN and NN). The *k*-NN algorithm is close to 100% with some gain in the lower atmosphere, up to 800 hPa, but with a degrading effect. As previously mentioned, the *k*-NN method would require further developments in order to benefit from the infrared/microwave synergy. The LIN method benefits from the synergy for all the atmospheric layers, in particular close to the surface. The impact can be important with an improvement of the retrieval statistics by more than 50% next to the surface and close to 25% in the middle troposphere. The NN inversion benefits significantly from the synergy: the retrieval errors can be reduced by a factor 2.5 at the surface. The synergy is enhanced for all atmospheric layers.

[64] Figure 16 shows similar statistics for the retrieval of water vapor. Again, the *k*-NN method is not optimal to merge information from the various captors: The impact of using simultaneously the three instruments (MHS, AMSU-A and IASI) is always negative, with an increase of errors up to 30%. The LIN method has a synergy factor close to 100% (meaning that no synergy is observed for the retrieval of water vapor) for higher atmospheric layers but the impact is quite positive for layers lower than 700 hPa. The NN method benefits from the synergy: The synergy factor reaches 140%, meaning that the errors are decreased by 40% when the three instruments are used together.

[65] In this section, the synergy obtained for the *k*-NN, LIN and NN methods have been compared and the results highlighted their potential for merging information. Note

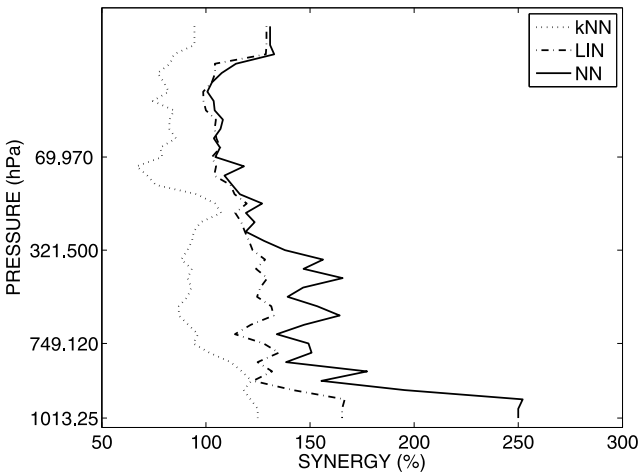

Figure 15. Synergy measure for the retrieval of temperature using IASI + AMSU-A + MHS observations for *k*-NN (dotted line), linear regression (dash-dotted line), and neural network (solid line) retrieval methods.

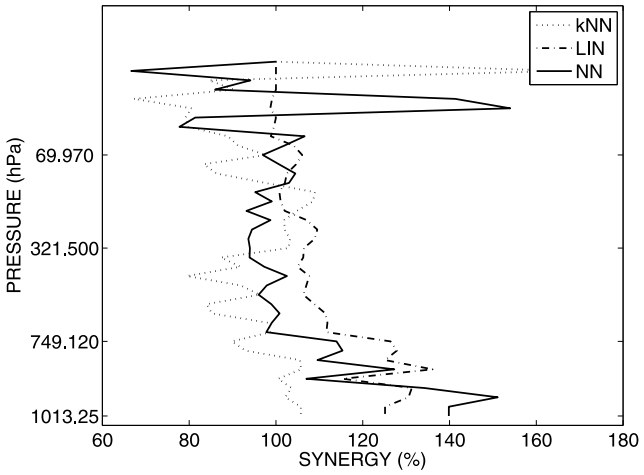

Figure 16. Synergy measure for the retrieval of water vapor using IASI + AMSU-A + MHS observations for *k*-NN (dotted line), linear regression (dash-dotted line), and neural network (solid line) retrieval methods.

that the choice of the retrieval method is not determined by this comparison only: the retrieval statistics described in section 4.5 are the definite quality criterion driving the choice for the inversion algorithm and Figures 13 and 14 have shown that the NN is always the best performing retrieval. The fact that synergy for the LIN can be higher at some altitudes than the synergy for the NN comes from the fact that the NN retrievals from the IR and MW are already very good and better than for the LIN model, the synergy is more difficult to obtain in this case.

5. Conclusion and Perspective

[66] This paper, together with the paper by *Aires* [2011], tried to explain some synergy mechanisms, how they occur and how to use them. The conclusions from this study are that (1) simple statistical retrieval tools can realistically measure the potential synergy of a set of satellite observations; (2) strong synergies exist between the microwave and infrared domains for the retrieval of atmospheric temperature and water vapor profiles, even for clear-sky conditions; and (3) the NN approach is able to exploit the synergy due to its truly multivariate nature and its nonlinear capacities.

[67] This study focused on the retrieval of the atmospheric temperature and water vapor profiles. A prototype retrieval chain has been developed and applied on three instruments (AMSU-A and MHS for the microwave, IASI for the infrared domain) all on board the MetOp platform. The method exploits optimally the synergy between the instruments. In addition, to maximize consistency between the retrieved variables and to help constrain the problem, the two profiles (temperature and water vapor) have been simultaneously estimated. This retrieval methodology is very general and can be applied to other sets of instruments, for different geometries, and to different atmospheric or surface parameters. A similar approach could be used to measure the 4 and 15 μm synergy for temperature retrieval or the long- and short-wave synergy for water vapor.

[68] This retrieval scheme will soon be extended to continental surfaces. It will also be tested on cloudy conditions. Since MW observations are less sensitive to clouds than IR measurements, even stronger synergy is expected for these cases.

[69] The final objective of this study is to develop a retrieval strategy to optimize the use of multiple satellite observations from the visible, the infrared, and the microwave for the estimation of atmospheric parameters. In particular, measurements from the GOME II instrument will be added, and the retrieval chain will estimate not only the temperature and water vapor but also the ozone atmospheric profiles.

[70] Ultimately, these types of tools should be considered in the future for the definition of new missions. The instrument characteristics should be determined not separately, independently for each sensor. Instead, all the instruments should be taken into account, together, to optimize globally the whole observing system.

[71] **Acknowledgment.** This project has been funded by ESA's General Studies Programme (GSP) under contract 21837/08/NL/HE, "Towards a synergetic approach for the retrieval of atmospheric geophysical parameters from optical/infrared and microwave measurements."

References

Aires, F. (2004), Neural network uncertainty assessment using Bayesian statistics with application to remote sensing: 1. Network weights, *J. Geophys. Res.*, *109*, D10303, doi:10.1029/2003JD004173.

Aires, F. (2011), Measure and exploitation of multisensor and multiwavelength synergy for remote sensing: 1. Theoretical considerations, *J. Geophys. Res.*, doi:10.1029/2010JD014701, in press.

Aires, F., and C. Prigent (2007), Sampling techniques in high-dimensional spaces for the development of satellite remote sensing database, *J. Geophys. Res.*, *112*, D20301, doi:10.1029/2007JD008391.

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

Aires, F., A. Chédin, N. Scott, and W. B. Rossow (2002a), A regularized neural network approach for retrieval of atmospheric and surface temperatures with the IASI instrument, *J. Appl. Meteorol.*, *41*(2), 144–159.

Aires, F., W. B. Rossow, N. A. Scott, and A. Chédin (2002b), Remote sensing from the infrared atmospheric sounding interferometer instrument: 1. Compression, denoising, and first-guess retrieval algorithms, *J. Geophys. Res.*, *107*(D22), 4619, doi:10.1029/2001JD000955.

Aires, F., W. B. Rossow, and A. Chédin (2002c), Rotation of EOFs by the independent component analysis: Towards a solution of the mixing problem in the decomposition of geophysical time series, *J. Atmos. Sci.*, *59*(1), 111–123.

Aires, F., W. B. Rossow, N. A. Scott, and A. Chédin (2002d), 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.

Aires, F., C. Prigent, and W. B. Rossow (2004a), Neural network uncertainty assessment using Bayesian statistics with application to remote sensing: 2. Output errors, *J. Geophys. Res.*, *109*, D10304, doi:10.1029/2003JD004174.

Aires, F., C. Prigent, and W. B. Rossow (2004b), Neural network uncertainty assessment using Bayesian statistics with application to remote sensing: 3. Network Jacobians, *J. Geophys. Res.*, *109*, D10305, doi:10.1029/2003JD004175.

Bishop, C. M. (1996), *Neural Networks for Pattern Recognition*, 482 pp., Clarendon, Oxford, U. K.

Blackwell, W. J. (2005), A neural-network technique for the retrieval of atmospheric temperature and moisture profiles from high spectral resolution sounding data, *IEEE Trans. Geosci. Remote Sens.*, *43*(11), 2535–2546.

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

Divakarla, M. G., C. D. Barnet, M. D. Goldberg, L. M. McMillin, E. Maddy, W. Wolf, L. Zhou, and X. Liu (2006), Validation of Atmospheric Infrared Sounder temperature and water vapor retrievals with matched radiosonde

measurements and forecasts, *J. Geophys. Res.*, *111*, D09S15, doi:10.1029/2005JD006116.
Eriksson, P., C. Jiménez, S. Bügker, and D. Murtagh (2002), A Hotelling transformation approach for rapid inversion of atmospheric spectra, *J. Quant. Spectrosc. Radiat. Transfer*, *73*(6), 529–543.
Evans, K. F., J. R. Wang, P. Racette, G. Heymsfield, and L. Li (2005), Ice cloud retrievals and analysis with data from the Compact Scanning Submillimeter Imaging Radiometer and the Cloud Radar System during CRYSTAL-FACE, *J. Appl. Meteorol.*, *44*, 839–859.
Geman, S., E. Bienenstock, and R. Doursat (1992), Neural network and the bias/variance dilemma, *Neural Comput.*, *1*(4), 1–58.
Hornik, K., M. Stinchcombe, and H. White (1989), Multilayer feedforward networks are universal approximators, *Neural Networks*, *2*, 359–366.
Huang, H. L., and P. Antonelli (2001), Application of principal component analysis to high-resolution infrared measurement compression and retrieval, *J. Appl. Meteorol.*, *40*, 365–388.
Huang, H.-L., and R. J. Purser (1996), Objective measures of the information density of satellite data, *Meteorol. Atmos. Phys.*, *60*, 105–117.
Jolliffe, I. T. (2002), *Principal Component Analysis*, 2nd ed., 487 pp., Springer, New York.
Jiménez, C., S. A. Buehler, B. Rydberg, P. Eriksson, and K. F. Evans (2007), Performance simulations for a submillimetre wave cloud ice satellite instrument, *Q. J. R. Meteorol. Soc.*, *133*(S2), 129–149.
Liu, X., D. K. Zhou, A. M. Larar, W. L. Smith, P. Schluessel, S. M. Newman, J. P. Taylor, and W. Wu (2009), Retrieval of atmospheric profiles and cloud properties from IASI spectra using super-channels, *Atmos. Chem. Phys.*, *9*, 9121–9142.
Matricardi, M., F. Chevallier, G. A. Kelly, and J.-N. Thépaut (2004), An improved general fast radiative transfer model for the assimilation of radiance observations, *Q. J. R. Meteorol. Soc.*, *130*, 153–173.
Menke, W. (1984), *Geophysical Data Analysis: Discrete Inverse Theory*, 160 pp., Academic, New York.
Pearl, J. (2010), *Causality: Models, Reasoning, and Inference*, 2nd ed., 478 pp., Cambridge Univ. Press, Cambridge, U. K.
Prunet, P., J.-N. Thépaut, and V. Cassé (1998), The information content of clear-sky IASI radiances and their potential for numerical weather prediction, *Q. J. R. Meteorol. Soc.*, *124*, 211–241.
Rabier, F., N. Fourrié, D. Chafau, and P. Prunet (2002), Channel selection methods for Infrared Atmospheric Sounding Interferometer radiances, *Q. J. R. Meteorol. Soc.*, *128*, 1011–1027.
Rodgers, C. D. (2000), *Inverse Methods for Atmospheric Sounding—Theory and Practice*, World Sci., London.
Rossow, W. B., and R. A. Schiffer (1999), Advances in understanding clouds from ISCCP, *Bull. Am. Meteorol. Soc.*, *80*, 2261–2288, doi:10.1175/1520-0477(1999).
Rumelhart, D. E., G. E. Hinton, and R. J. Williams (1986), Learning internal representations by error propagation, in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. I, *Foundations*, edited by D. E. Rumelhart, J. L. McClelland, and the PDP Research Group, pp. 318–362, MIT Press, Cambridge, Mass.
Rydberg, B., P. Eriksson, S. A. Buehler, and D. P. Murtagh (2009), Non-Gaussian Bayesian retrieval of tropical upper tropospheric cloud ice and water vapour from Odin-SMR measurements, *Atmos. Meas. Tech.*, *2*, 621–637.
Saunders, R. W., M. Matricardi, and P. Brunel (1999), An improved fast radiative transfer model for assimilation of satellite radiance observations, *Q. J. R. Meteorol. Soc.*, *125*, 1407–1425.
Shannon, C. E. (1949), Communication in the presence of noise, *Proc. ICE*, *37*, 10–21.
Smith, J. A., and J. P. Taylor (2004), Initial cloud detection using the EOF components of high-spectral-resolution infrared sounder data, *J. Appl. Meteorol.*, *43*, 196–210.
Ziv, J., and A. Lempel (1979), A universal algorithm for sequential data compression, *IEEE Trans. Inf. Theory*, *23*(3), 337–343.

F. Aires, Estellus, 93 bl Sébastopol, F-75002 Paris, France. (filipe.aires@estellus.fr)
M. Bouvet and B. Rommen, Wave Interaction and Propagation Section, European Space Agency, Keplerlaan 1, PB 299, NL-2200 AG Noordwijk, Netherlands. (marc.bouvet@esa.int; bjorn.rommen@esa.int)
M. Paul and C. Prigent, Laboratoire de l'Etude du Rayonnement et de la Matière en Astrophysique, CNRS, Observatoire de Paris, 61 av. de l'Observatoire, F-75014 Paris, France. (maxime.paul.2006@polytechnique.org; catherine.prigent@obspm.fr)