A Study on Retrieving Atmospheric Profiles from EOS/AIRS Observations*

GUAN Li (官 莉)¹, ALLEN Huang², and LI Jun (李 俊)²

¹Department of Electronic Engineering, Nanjing University of Information Science & Technology, Nanjing 210044, China
²University of Wisconsin, Madison, WI 53705, USA

(Received February 10, 2004; revised July 21, 2004)

ABSTRACT

The paper presents the algorithms for retrieving atmospheric temperature and moisture profiles and surface skin temperature from the high-spectral-resolution Atmospheric Infrared Sounder (AIRS) with a statistical technique based on principal component analysis. The synthetic regression coefficients for the statistical retrieval are obtained by using a fast radiative transfer model with atmospheric characteristics taken from a dataset of global radiosondes of atmospheric temperature and moisture profiles. Retrievals are evaluated by comparison with radiosonde observations and European Center of Medium-Range Weather Forecasts (ECMWF) analyses. AIRS retrievals of temperature and moisture are in general agreement with the distributions from ECMWF analysis fields and radiosonde observations, but AIRS depicts more detailed structure due to its high spectral resolution (hence, high vertical spatial resolution).

Key words: AIRS (Atmospheric Infrared Sounder), principal component regression, atmospheric profile retrieval

1. Introduction

The development of global climate and weather models requires accurate monitoring of atmospheric temperature and moisture profiles, as well as the contents of trace gases and aerosols. It is quite difficult to monitor continuously these parameters on a global scale.

Until recently, AIRS (Atmospheric Infrared Sounder) offers a new opportunity to improve global monitoring of temperature, moisture, and ozone distributions and changes therein. The high spectral resolution ($\nu/\Delta \nu \approx 1200$), and near-daily global coverage of AIRS enable it to observe the earth’s atmosphere and continuously monitor its changes. AIRS is a primary instrument on the Aqua polar orbiting satellite, which was launched on the Earth Observing System (EOS) Aqua platform on May 4, 2002 together with the Advanced Microwave Sounding Unit (AMSU-A) and Humidity Sounder of Brazil (HSB). Having a 650 cm$^{-1}$ (15 $\mu$m) to 2700 cm$^{-1}$ (3.7 $\mu$m) spectral range with 2378 spectral IR channels, spectral response functions (SRFs) with full widths at half maximum of $\sim \nu/1200$ (0.5-2.3 cm$^{-1}$) and noise levels on the order of 0.2 K (70% of AIRS channels have noise less than 0.2 K, 20% have noise less than 0.1 K), AIRS will be used to gather atmospheric profile soundings covering nearly the entire earth every day. The AIRS footprint is 13 km at nadir, as is the HSB footprint, with a 3×3 array of AIRS and HSB footprints falling into a single AMSU-A footprint (Susskind et al., 2003). Though some of these channels are not in desirable conditions because of high noise and other reasons, there are still 2047 channels available for this study. Figure 1 shows the simulated AIRS brightness temperature spectra under clear conditions for US standard atmosphere.

AIRS retrievals of atmospheric water vapor and temperature distributions are intended to advance understanding of the role played by energy and water cycle processes in determining the earth’s weather and climate. AIRS temperature and moisture products can be used together with other satellite measurements in numerical weather prediction models or in the regions where conventional meteorological observations are

*Supported by the National Natural Science Foundation of China under Grant No. 49775255.
sparse.

Although MODIS (MODe rate resolution Imaging Spectroradiometer) due to its increased high spatial resolution (1 km at nadir) measurements delineates horizontal gradients of moisture, temperature, and atmospheric total ozone better than its companion instruments such as AIRS (13 km at nadir), the MODIS broadband spectral resolution provides only modest information content regarding vertical profiles while AIRS contains more information about the atmospheric vertical distribution of temperature and moisture (Seemann et al., 2003).

2. Algorithm development

Suppose that scattering by the atmosphere can be neglected in the local thermal equilibrium. The true clear spectrum of infrared radiance emitted from the earth-atmosphere system is approximated by (Li et al., 2000)

\[ R = \varepsilon B_s \tau_s - \int_0^{p_s} B d\tau(0, p) + (1 - \varepsilon) \int B d\tau^* + R', \]

where \( R \) is the spectral radiance in the infrared region or the brightness temperature in the microwave region; \( B \) is the Planck radiance in the infrared region and is a function of pressure \( p \) and temperature; \( \tau \) is the atmospheric transmittance function and \( \tau^* = \tau_s^2 / \tau \); subscript “s” denotes surface; \( R' \) represents the contribution of reflected solar radiance in the infrared region; and \( \varepsilon \) is the surface emissivity.

Because the retrieval problem is ill posed, additional information is needed to constrain the solution. Often this is accomplished by means of a first-guess profile obtained from a climate mean, a regression technique, and/or numerical forecast products. A statistical regression model is generated for the first-guess retrieval from AIRS measurements under clear sky conditions. The fast forward model calculation of AIRS radiance is performed for each radiosonde case of the training dataset to provide a radiosonde-AIRS radiance pair for the statistical regression analysis. A regression equation is generated based on these theoretical calculations of radiance and matching radiosonde temperature, moisture, and ozone profiles. This regression equation can be applied to the real AIRS radiance to generate an excellent initial profile of the atmospheric state, as needed for the physical solution of the RTE (radiative transfer equation). In addition, the local satellite zenith angle of observation is also used as predictors.

Another way to generate the regression equation is by using the matchup file that contains the time and space collocated satellite radiance measurements and radiosonde profile. The advantage of the regression equation using the theoretical calculation over the real observation is that it avoids errors due to time and space differences between the satellite observation and radiosonde profile; however, using the real matchup data will overcome the impact of bias caused by the
imperfection of forward model calculation.

The retrieval procedure involves statistical regression and eigenvector regression. The radiative transfer calculation of the AIRS radiances is performed using a transmittance model called SARTA.

2.1 Statistical synthetic regression retrieval processing

A computationally efficient method for determining the distribution of atmospheric temperature, moisture, and ozone from satellite sounding measurements uses previously determined statistical relationships between observed or modeled radiances with corresponding atmospheric profiles. In the regression procedure, temperature and moisture are regressed together against radiances from CO$_2$, water vapor, and window channels. This method is often used to generate a first guess for a physical retrieval algorithm. The statistical regression algorithm for atmospheric temperature is described in detail by Smith et al. (1970), and is summarized below for cloud-free skies.

If the satellite-observed radiance or brightness temperature $R$ of each channel is known, then $R$ can be considered a nonlinear function of the atmospheric temperature profile, water vapor mixing ratio profile, surface skin temperature, etc. That is, $R = R(T,q,T_s,\Lambda) + \sigma$, where $\sigma$ is the instrument noise and other sources of error, or in general

$$Y = F(X) + \sigma,$$  \hspace{3cm} (2)

where the vector $X$ contains $L$ (101 levels of atmosphere) atmospheric temperatures, $L$ atmospheric water vapor mixing ratios (the water vapor is expressed as the logarithm of the mixing ratio in practical applications), one surface skin temperature, etc., and $Y$ contains $N$ (number of AIRS channels used) satellite observed radiance or brightness temperatures.

The AIRS grid has 101 levels and spans the range from 1100 to 0.005 hPa. The grid is designed to provide the vertical special resolution meeting the needs for accurate calculation of AIRS simulated radiances. The level numbering scheme used in the AIRS retrieval has level 1 closest to the satellite and level 101 closest to the surface.

The general inverse solution of Eq.(2) for the atmospheric profile can be written as

$$X(i,k) = A(i,n)Y(n,k),$$ \hspace{3cm} (3)

where $i$ is the number of parameters to be retrieved, $k$ is the number of profiles and surface data in the training sample, and $n$ is the number of channels and other predictors used in the regression procedure. The statistical regression algorithm seeks a ‘best fit’ operator matrix $A$ that is computed using the least square method with a large sample of atmospheric temperature and moisture soundings and collocated radiance observations. Minimizing the difference between synthetic observations and the regression model

$$\frac{\partial}{\partial A} |AY - X|^2 = 0,$$  \hspace{3cm} (4)

yields

$$A(i,n) = X(i,k)Y^T(k,n)[Y(n,k)Y^T(k,n)]^{-1},$$ \hspace{3cm} (5)

where $(Y^TY)$ is the covariance of the radiance observations and $(Y^TX)$ is the covariance of the radiance observations with the atmospheric profile.

Ideally, the radiance predictors $Y$ would be taken from actual AIRS measurements and used the time and space collocated radiosonde profiles $X$ to directly derive the regression coefficient $A$. However, radiosondes are only routinely 1100 and 2300 LST each day. It is, therefore, not possible to obtain many time and space collocated radiosondes and AIRS radiances that are globally distributed at a wide range of locations. Alternatively, synthetic regression coefficients can be generated from AIRS radiances calculated using a transmittance model with profile input from a global temperature and moisture radiosonde database. However, it involves the radiative transfer calculation and an accurate forward model in order to obtain a reliable regression relationship. Any uncertainties (e.g., a bias of the forward model) in the radiative calculations will influence the retrieval.

In AIRS regression procedure, the primary predictors $[Y]$ are radiances from 2047 channels. The predictors, or the parameters to be retrieved by regression, include the temperature profile, logarithm of the water vapor mixing ratio profile, and surface skin temperature. To perform the retrieval, Eq.(3) is applied to the
actual AIRS measurements, where $\mathbf{Y}$ is the observed radiances. The retrieved water vapor mixing ratio at each pressure level is checked for saturation and the mixing ratio is set equal to the saturation-mixing ratio if the relative humidity is greater than 100%.

2.2 Training data

In the AIRS retrieval algorithm, the global dataset of radiosonde observations in Sep. 2003 are used for training the regression. The original training data contain 8000 globally distributed clear-sky radiosonde profiles of temperature, moisture along with observations of surface temperature and pressure. (All profiles are extended from 1100 to 0.005 hPa and interpolated to 101 levels.) The radiative transfer calculation of the AIRS radiances is performed with the forward model for each profile from the training dataset to produce a temperature-moisture profile-AIRS radiance pair. The synthetic regression coefficients are generated using the simulated radiances and the matching atmospheric profile.

The estimated AIRS instrument noise was added to the calculated radiances before creating the coefficients. The noise was randomly generated with a Gaussian distribution, a standard deviation equal to the NEdT (noise equivalent delta temperature for a 250 K brightness-temperature scene) in AIRS channel characteristic file, and an average of zero.

In the simulation, 7000 profile data samples were used for training, while radiances simulated from the remaining 1000 profiles were used for testing the retrieval algorithm.

2.3 Forward model

In our study we use the SARTA V1.03 forward model. This package is “stand alone” implementation of the AIRS-RTA (Atmospheric Infrared Sounder Radiative Transfer Algorithm), one fast forward model for AIRS radiative transfer calculations (Strow et al., 2003). AIRS finally uses a physical algorithm for retrieval of atmospheric profiles, and consequently is dependent on an accurate, and fast, radiative transfer algorithm for computing clear-air radiance. A call to AIRS-RTA represents the most CPU-intensive part of the operational processing, hence the RTA must be fast. The high spectral resolution of AIRS coupled with its low noise should produce retrievals that are as good as, or better than the worldwide operational radiosonde network, if the forward model accuracy approaches the noise level of the instrument.

2.4 Application of principal component analysis

Principal components (PCs) and eigenvector regression is to provide the nearly full spectral information with less degradation, noise reduction, and tolerable loss of accuracy in temperature and water vapor retrieval. The technique is commonly used to reduce the dimensionality of a dataset with a large number of interdependent variables. This reduction is achieved by finding a set of $N_t$ orthogonal vectors in the input space of dimension, with $N_t < N_c$, which accounts for as much as possible of the data variance. Hence, the problem of dimensionality reduction is reduced to finding a linear transformation from the $N_c$-dimensional input space to an $N_t$-dimensional subspace spanned by $N_t$ orthogonal vectors defined above and hereinafter referred to as principal components (Huang et al., 2001).

In practice, one often uses the deviations of the radiances from the sample mean for $\mathbf{Y}$. Calculating regression coefficients

$$C = \frac{1}{N_s} [\Delta \mathbf{Y}_{tr} \Delta \mathbf{Y}_{tr}^T],$$

$$\Delta \mathbf{Y}_{tr} = \mathbf{Y}_{tr} - \bar{\mathbf{Y}}_{tr},$$

$U(n, n_e)$ is the first $n_e$ numbers of eigenvectors of matrix $C$.

$$R_{tr} = \Delta \mathbf{Y}_{tr} U,$$

$$A = \Delta \mathbf{X}_{tr} \Delta \mathbf{Y}_{tr}^T [\Delta \mathbf{Y}_{tr} \Delta \mathbf{Y}_{tr}^T]^{-1},$$

where $n$ is the number of channels, $N_s$ is the number of training samples and subscript “tr” denotes training data.

To perform the retrieval (independent)

$$R_{obs} = \Delta \mathbf{Y}_{obs} U,$$

$$\Delta \mathbf{Y}_{obs} = \mathbf{Y}_{obs} - \bar{\mathbf{Y}}_{tr},$$

$$\mathbf{X} = \hat{\mathbf{X}}_{tr} + A R_{obs}^T,$$
where subscript “obs” denotes observed or simulated radiances.

Figure 2 shows the RMS errors of averaged temperature and humidity from 1000 to 100 hPa using the different number of PCs for the 7000 sample profiles.

RMS errors are defined as

\[ E_{\text{rms}} = \sqrt{\frac{1}{N_s} \sum_{i=1}^{N_s} (X_{\text{RAOB}} - X_{\text{AIRS}})^2}, \]  

where \( X_{\text{RAOB}} \) and \( X_{\text{AIRS}} \) are the radiosonde observation and AIRS retrieved parameters respectively, and \( N_s \) is the total number of comparisons (\( N_s =7000 \)).

From Fig. 2 we can see that after first 30 PCs the \( E_{\text{rms}} \) is nearly a constant and according to the eigenvalue, the first 30 PCs can represent the 99.9988% information of all AIRS information, thus in our eigenvector regression we use the first 30 PCs.

3. Evaluation of products

The remaining 1000 profiles (independent) were used for testing the retrieval algorithm. We compared our retrieval profiles and a calculated the RMS errors with radiosonde observations in term of all 2047 good channels and only using first 30 PCs.

Figure 3 shows the RMS errors \( (E_{\text{rms}}) \) of the retrieved vertical profiles of temperature and moisture compared with the actual profiles for the 1000 independent regression retrievals. The most accurate AIRS temperature retrievals are between 800-300 hPa, where the \( E_{\text{rms}} \) is approximately 1 K. Near the surface, \( E_{\text{rms}} \) increases to 2 K. Moisture retrieval accuracy decreases with height from an \( E_{\text{rms}} \) maximum.

![Fig.2. Averaged RMS errors with different number of PCs.](image)

![Fig.3. The temperature (a) and humidity (b) \( E_{\text{rms}} \) of the retrieval for all test 1000 profiles. Solid lines are for using first 30 PCs and dashed lines are for using all 2047 good channels.](image)
Fig. 4. An example of a single-profile retrieval compared with radiosonde profile (Details see text).

of 2 g kg$^{-1}$ in the lower boundary layer. The $E_{\text{rms}}$ when the principal component analysis is used, is smaller than that if all channels are used because of the noise reduction.

Fig. 5. RMS errors (at each pressure level) of ECMWF analysis minus retrieval for temperature (a) and humidity (b) shown for clear pixels of granule 192 (2003-09-02). The numbers inside the panels refer to the RMS errors of skin temperature [K] and of total precipitable water [cm], respectively.

In Fig. 4 an example of a single-profile retrieval compared with radiosonde profile is presented. The
panels display (from left to right) the temperature profiles, the corresponding differences of RAOB minus the retrieved profile, humidity profiles and their differences. Retrieved profiles include using all channels (dash lines) and using first 30 PCs retrieval (solid lines).

Because of the fast-run speed and satisfying with the accuracy needed in real-time processing, we apply the eigenvector regression algorithm to a real time AIRS data retrieval experiment to granule 192 on 2003-09-02 day time. AIRS atmospheric and surface parameter retrievals require clear-sky measurements. Using cloud-mask algorithm we find nearly 4500 clear pixels in this granule. The retrieval results are being compared with the corresponding ECMWF analysis fields. Note that the analysis data have been spatially interpolated to AIRS measurement locations, and the difference in time is approximately one hour. The RMS errors of the difference between analysis and retrieval for all clear pixels (approximately 4500 in number) are given in Fig.5 for temperature and humidity profiles at each pressure level. Largest values occur at and near the surface, partly because of surface parameters such as surface skin temperature and surface emissivity which are not sufficiently and/or correctly represented in the training data set. Results obtained shows the acceptable retrieval performance in real-time processing. In this real-time experiment, the $E_{\text{rms}}$ of temperature retrievals is approximately 1 K and the $E_{\text{rms}}$ of moisture retrieval accuracy is less than 2 g kg$^{-1}$ except for the near surface.

4. Error sources

Retrieval accuracy, computation efficiency, and retrieval validation are important considerations when applying an algorithm to the real-time AIRS data processing. Several sources of errors must be addressed. Firstly, forward model errors can influence the retrievals. These may be due to atmospheric transmittance calculation error, due to inaccurate or insufficient representation of the atmospheric temperature and moisture profiles in the training dataset, and due to the surface uncertainties, such as surface elevation, emissivity, and skin temperature. Improvement of the forward model is important for deriving the AIRS products with a high level of accuracy. Secondly, the AIRS instrument detector noise and calibration error (observation error) can have an impact on the retrieval accuracy. Thirdly, cloud detection errors may also have a negative impact on the retrieval products. Regions with cloud contamination in the AIRS retrievals show inaccurate moisture and temperature. A challenge facing us is to maintain adequate global training data so that the synthetic regression algorithm produces accurate atmospheric retrievals.

5. Conclusions and future work

We have presented the algorithm and results of AIRS regression retrieval. Results obtained so far promise acceptable retrieval performance in real-time.

Future work to improve the algorithm will include radiance bias corrections. The forward model-calculated radiances have biases with respect to the AIRS measured radiance. The synthetic regression and physical retrieval methods use both measured and calculated radiances and thus require this bias to be minimized, trying to find a relationship between the radiance bias and measured radiances.

According to MODIS retrieval experience, to limit the retrievals to training data with physical relevance to the observed conditions, the training dataset were partitioned into several zones based on the 1000 cm$^{-1}$ brightness temperatures calculated from the profiles. Different zones would use different regression coefficients based on statistics. Thus retrieval is performed with a subset of regression coefficients in the same zone as the observed 1000 cm$^{-1}$ AIRS brightness temperature.

Due to the surface uncertainties, such as surface elevation, emissivity, and skin temperature uncertainties, time and space collocated surface temperature and moisture observations should be used as two additional predictors in the regression so that surface observations provide additional information to better constrain the statistical retrieval at near surface levels.

A more sophisticated nonlinear physical iterative solution of the radiative transfer equation will be
conducted using the results of regression retrieval as the initial profile to get the final temperature profile, moisture profile, and total atmospheric profiles.

REFERENCES


