

Identifying Cloud-Uncontaminated AIRS Spectra from Cloudy FOV Based on Cloud-Top Pressure and Weighting Functions

M. CARRIER AND X. ZOU

The Florida State University, Tallahassee, Florida

WILLIAM M. LAPENTA

NASA MSFC Global Hydrology and Climate Center, Huntsville, Alabama

(Manuscript received 27 March 2006, in final form 25 August 2006)

ABSTRACT

An effort is made to increase the number of Advanced Infrared Sounder (AIRS) cloud-uncontaminated infrared data for regional mesoscale data assimilation and short-term quantitative precipitation forecast (QPF) applications. The cloud-top pressure from Moderate Resolution Imaging Spectroradiometer (MODIS) is utilized in combination with weighting functions (WFs) to develop a channel-based cloudy-data-removal algorithm. This algorithm identifies “clear channels” for which the brightness temperature (BT) values are not cloud contaminated. A channel-dependent cutoff pressure (COP) level is first determined based on the structure of the WF of each channel. It is usually below the maximum WF level. If the cloud top (as identified by a MODIS cloud mask) is above (below) the COP level of a channel, this channel is then deemed cloudy (clear) and removed (retained). Using this algorithm, a sizable increase of cloud-uncontaminated AIRS data can be obtained. There are more usable domain points for those channels with higher COP levels. A case study is conducted. It is shown that instead of having less than 20% AIRS clear-sky observations, the algorithm finds 80% (58%) of the AIRS pixels on which there are channels whose COP levels are at or above 300 hPa (500 hPa) and the BT data in these channels at these pixels are cloud uncontaminated. Such a significant increase of the usable AIRS cloud-uncontaminated data points is especially useful for regional mesoscale data assimilation and short-term QPF applications.

1. Introduction

Infrared radiances are a very useful source of information for atmospheric data assimilation. Data obtained from infrared satellites, especially hyperspectral sounders such as the Advanced Infrared Sounder (AIRS), provide valuable information regarding the atmospheric thermodynamic state at superior vertical resolution (Pagano et al. 2002). Much work has been done to directly utilize radiance data in the current numerical weather prediction (NWP) environment via variational data assimilation techniques. Among the many obstacles to successfully assimilating radiance data into an atmospheric model is the presence of clouds, whose effect is difficult to properly account for

in the so-called forward radiative transfer model that is required in the data assimilation and prediction model (Strow et al. 2003; Derber and Wu 1998). In fact, at AIRS finite spatial footprint resolution (~ 14 km at nadir), less than 5% of observations can be absolutely unaffected by clouds. In other words, about 95% of the AIRS infrared footprints are possibly contaminated by clouds.

The presence of clouds in a satellite’s field of view (FOV) may contaminate the radiance data such that valuable atmospheric information is not easily attainable. Several techniques have been used at various research and operational centers to handle this problem. One method involves a conservative approach of locating only clear FOV from which IR channels can be used. This method uses threshold values for a cloud cost function and a longwave window channel’s observed background differences to determine which pixels (i.e., domain points) are cloudy or clear (Collard 2004). Another method involves altering the cloud-

Corresponding author address: Matthew Carrier, Department of Meteorology, The Florida State University, Tallahassee, FL 32306-4520
E-mail: mcarrier@met.fsu.edu

contaminated radiance values such that the new values more closely approximate clear-sky radiances for the given pixel; this results in the so-called cloud-cleared radiances (Derber and Wu 1998). One other method attempts to explicitly account for the clouds by using a cloudy radiative transfer model (RTM) designed to handle cloudy atmospheric profiles (Greenwald et al. 2002; Liou et al. 2005; Chevallier et al. 2004). This method, albeit promising, is still some time away from being applied in an operational setting.

There is an ongoing effort at The Florida State University, in cooperation with the National Aeronautics and Space Administration's (NASA's) Global Hydrology and Climate Center (GHCC), which aims to utilize radiance observations obtained from the AIRS instrument on board NASA's *Aqua* satellite to improve short-term, regional quantitative precipitation forecasts (QPFs). Unfortunately, using only clear-sky pixels from the AIRS data is prohibitive, as the number of clear-sky data points is extremely low. It is desirable to identify clear channels (e.g., channels not affected by clouds) from cloudy FOV rather than simply using channels at only clear FOV. Previous researchers have examined this problem in an attempt to utilize radiances from cloudy FOV. Most notably, McNally and Watts (2003) developed an algorithm now used at the European Centre for Medium-Range Weather Forecasts (ECMWF) for identifying clear channels from cloudy FOV. Cloudy data are flagged based on the departure of AIRS observations from simulated clear-sky radiances. The departures of simulated clear-sky radiances across the AIRS spectrum at each domain point are ranked by the magnitudes of the departures (i.e., cloud sensitivity) determined dynamically at each sounding location and band split by primary absorbing gas (CO_2 band, H_2O band, etc.). These ranked, band-split departures are filtered using a low-pass (boxcar) filter to remove the high-frequency noise caused by NWP background errors, RTM errors, and observational noise in order to isolate the cloud signal. Once this is done, the algorithm then searches for the channels for which the departure itself is greater than a threshold value (>0.5) and the departure gradient is greater than a threshold value (>0.2). All channels for which the radiance departures and departure gradients are in excess of *both* threshold values are flagged cloudy, and the others are flagged clear.

While shown to work well for most AIRS data, the algorithm had difficulty with near-surface channels as the error in the surface skin temperature is incorrectly flagged as a cloud signal. It is determined that this method, while well established and accurate (Lavanant et al. 2004), is not necessarily optimal for our purposes.

The reason for this is twofold: first, there is a great need to include as many near-surface channels as possible for AIRS data assimilation in regional precipitation forecast applications. More importantly, the ECMWF method requires the use of an RTM capable of handling the effects of scattering due to clouds. The RTM selected for this work simulates the radiance of a non-scattering atmosphere; therefore it is necessary to use the alternate approach of clear-channel identification outlined in this paper.

To increase the usefulness of the AIRS data for mesoscale model forecasts, a simple and inexpensive algorithm is developed that would allow a larger number of AIRS data (not cloud contaminated) to be used. This is done by identifying and utilizing all data obtained from cloudy pixels at those channels whose peak emissions lie well above the cloud tops, thereby increasing the number of usable (i.e., cloud free) pixels for radiances at specific AIRS channels. The algorithm described herein does not itself detect cloud signals; rather, it uses a Moderate Resolution Imaging Spectroradiometer (MODIS) cloud mask to determine the physical location of clouds within the domain. The algorithm determines, for each channel at each pixel, the vertical atmospheric level at and above which the existence of an opaque cloud covering the FOV will contaminate the radiances.

In the following discussion, the AIRS instrument is briefly introduced along with an overview of the selected RTM and test case (section 2). Section 3 describes the MODIS cloud mask and its application in identifying cloudy pixels in the AIRS data swaths. Section 4 reviews the first step of a limited cloud-contaminated data removal (LCCDR) algorithm that removes cloudy radiances based on the cloud-top pressure provided by the MODIS cloud mask and the channel-dependent cutoff pressure (COP) level. Section 4 will also discuss the second step of the LCCDR algorithm, consisting of an outlier identification method employed to detect and remove residual cloud-contaminated data from the first step of the LCCDR algorithm. The results of the LCCDR algorithm are presented in section 5, followed by conclusions and a preview of future work in section 6.

2. The AIRS instrument, RTM, and test case

a. The AIRS instrument

AIRS, one of the many instruments carried on board NASA's *Aqua* satellite, is a hyperspectral infrared sounder capable of collecting 2378 thermal infrared radiance observations across a spectrum from 3.7 to 15.4 μm . The cross-track swath dimension is 1650 km. The

spatial resolution for AIRS is 13.5 km at nadir (Aumann et al. 2003).

AIRS level-1B radiance data are available in hierarchical data format (HDF) format from the Goddard Earth Sciences Data and Information Services Center (GES DISC). AIRS data from each day are structured in 6-min swath packages known as granules. Since the *Aqua* satellite is sun synchronous, the data are available twice daily at the same local times each day, globally. The level-1B radiance data used for this effort consist of calibrated radiances assigned to each wavenumber in the spectrum; these data are not cloud cleared (which is available in level-2 data). The radiances are converted into brightness temperatures (BT) here with the inverse of the Planck function. Noisy and/or popping channels, as specified by several onboard calibration tests (space view test, onboard calibrator cool down test, etc.), are removed using the quality control package available with the AIRS radiance packages. For further information regarding the AIRS instrument, see Pagano et al. (2002).

b. The Radiative Transfer Model

The RTM selected for this project is the Stand-Alone AIRS Radiative Transfer Algorithm (SARTA; Strow et al. 2003). SARTA calculates radiance values for 2 pixels each second, with errors near that of the AIRS instrument itself (about 0.2 K at 250 K) (Strow et al. 2003). SARTA is designed to calculate a simulated AIRS radiance spectrum as the convolution of the monochromatic radiance R_ν with the AIRS spectral response function (SRF) for a specified channel i , such that

$$R_\nu^{\text{AIRS}} = \int R_\nu \text{SRF}_\nu d\nu, \quad (1)$$

where the SRF values are known for each channel. The value of R_ν , the monochromatic radiance leaving the top of a nonscattering clear atmosphere, is calculated by SARTA. The radiative transfer algorithm used by SARTA is specified to consider only four source terms: surface emission, atmospheric emission, downwelling atmospheric emission reflected by the surface, and reflected solar radiation. To make SARTA computationally fast, Strow et al. (2003) designed the RTM to compute simulated AIRS radiances generated from convolved transmittances (rather than convolved monochromatic radiances). The final radiative transfer algorithm takes this final form:

$$R_\nu = \varepsilon_\nu B_\nu(T_s) \mathfrak{T}_s^{\text{eff}} + \sum_{l=1}^{l_s} B_\nu(T_l) (\mathfrak{T}_{l-1}^{\text{eff}} - \mathfrak{T}_l^{\text{eff}}). \quad (2)$$

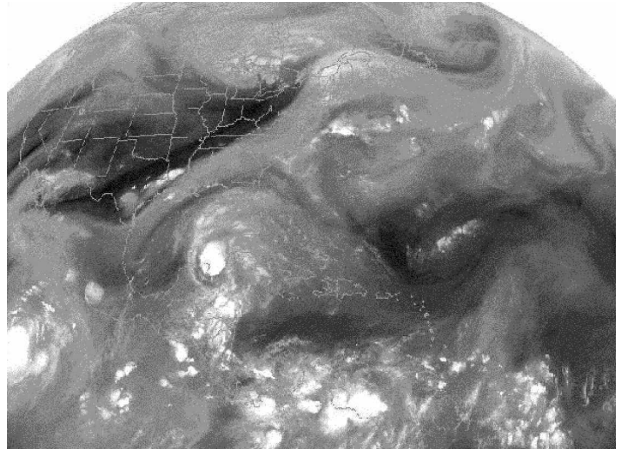


FIG. 1. Geostationary Operational Environmental Satellite-12 (GOES-12) midlevel water vapor image from 1200 UTC 11 Jul 2003. Image courtesy of National Climatic Data Center (NCDC) Satellite Browse Archive (see online at <http://cdo.ncdc.noaa.gov/GEOSBrowse/goesbrowser>).

In this equation, B_ν is the Planck function, ε_ν is the emissivity at frequency ν , $R_{\text{refl.th}}^{\text{eff}}$ is the reflected downwelling radiation, H_ν is the solar irradiance incident at the top of the atmosphere, $\mathfrak{T}_i^{\text{eff}}(\theta)$ is the layer-to-space transmittance, ρ_{solar} is the solar reflectance by the surface, and θ_{sun} is the solar zenith angle. Details on how each term in (2) is calculated can be found in Strow et al. (2003).

c. Selected test case

The test case selected for this work involves a strong midlevel moisture gradient located over the southeastern United States between 0800 UTC 11 July and 0800 UTC 12 July 2003 (Fig. 1). The fifth-generation Pennsylvania State University–National Center for Atmospheric Research Mesoscale Model (MM5, version 3) is used in this work to provide data as input to the SARTA model (Dudhia 1993). The 36-h MM5 forecast, initialized at 0000 11 July 2003, has a domain centered over St. Louis, Missouri, with a size of $150 \times 150 \times 35$ and a horizontal resolution of 20 km (Fig. 2, light gray area). The Grell cumulus convective scheme (Grell et al. 1995) and Blackadar planetary boundary layer (Blackadar 1979) are used in this forecast. The size of the MM5 forecast domain is large enough to cover the southeastern United States (the region of interest) while also overlapping with a majority of AIRS pixels from four separate scan swaths (or granules), which include granule 078 (0747–0753 UTC 11 July 2003), granule 188 (1847–1853 UTC 11 July 2003), granule 189 (1853–1859 UTC 11 July 2003), and granule 069 (0653–0659 UTC 12 July 2003).

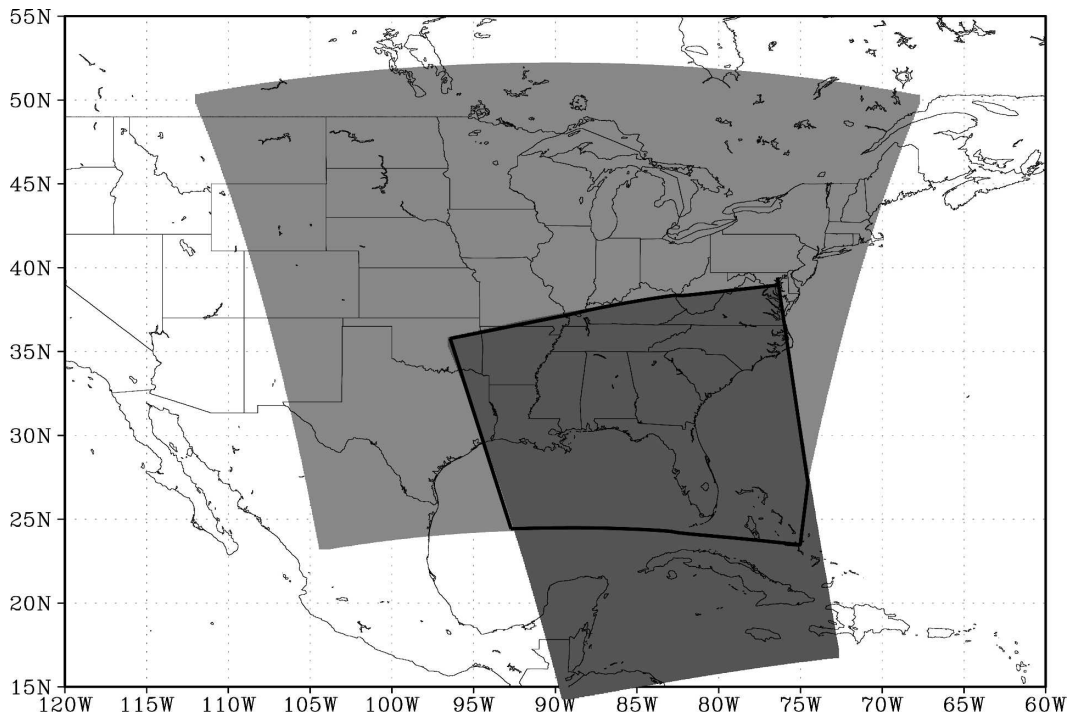


FIG. 2. MM5 forecast domain (light gray) and AIRS data domain (dark gray) used in this study that covers a 6-min scan swath (granule 188, 1847–1853 UTC 11 Jul 2003). Black box region indicates the MM5/AIRS overlapping domain.

Figure 3 displays AIRS BT data for AIRS observations from granule 188 for spectral channel $6.6 \mu\text{m}$, a water vapor channel. The AIRS instrument picks up the signature of the midlevel moisture gradient, as depicted by the BT gradient across the southeastern United States.

3. The MODIS cloud mask

To identify cloudy data in the AIRS FOV, a cloud mask is needed that quantifies the cloud property of each of the 12 150 AIRS pixels in a given granule. To accomplish this, researchers at NASA's GHCC have adapted the *Aqua* satellite's MODIS cloud mask products for cloud identification and cloud height to be used for AIRS data swaths (Haines et al. 2004; MODIS Cloud Mask Team 2002). MODIS has 36 spectral bands (21 within $0.4\text{--}3.0 \mu\text{m}$, 15 within $3.0\text{--}14.5 \mu\text{m}$) with spatial resolutions of 250 m (bands 1–2), 500 m (bands 3–7), and 1000 m (bands 8–36; Barnes et al. 1998).

The MODIS cloud mask (produced with the Earth Observing System Science Team's institutional algorithm) is obtained from the Distributed Active Archive Center (DAAC) for each AIRS time period used in this study. The MODIS cloud mask (known as MOD35) is available at 1-km resolution over the AIRS swath for

both day and night passes, since both instruments are carried on board the *Aqua* satellite. Although the quality of the MODIS cloud mask varies between night and day and land and ocean over the southeastern United States (Haines et al. 2004), its performance is generally quite good with greater than 80% of the cloud conditions being properly detected. The MODIS cloud data are interpolated to the AIRS swath and used to determine the percent cloud cover for each AIRS FOV (on average about 225 MODIS points are used for each AIRS footprint). Additionally, the MODIS cloud-top pressure (known as MOD06) levels are used to obtain a mean cloud-top pressure for each AIRS pixel using the cloud mask information. This approach provides three situational parameters for the selection of clear-sky AIRS pixels for each footprint: 1) a cloud/no cloud determination, 2) a varying threshold (0%–100%) that quantifies the percent of each AIRS pixel covered in clouds, and 3) a cloud-top pressure field that estimates the average height of the cloud covering each AIRS pixel. The first and simplest parameter leaves very few (can be as little as 5%) AIRS pixels from which to perform an analysis. The second parameter is not desirable for this study, as the intention is to use clear-channel data. The third parameter is used in the algorithm described in this paper to identify clear-channel

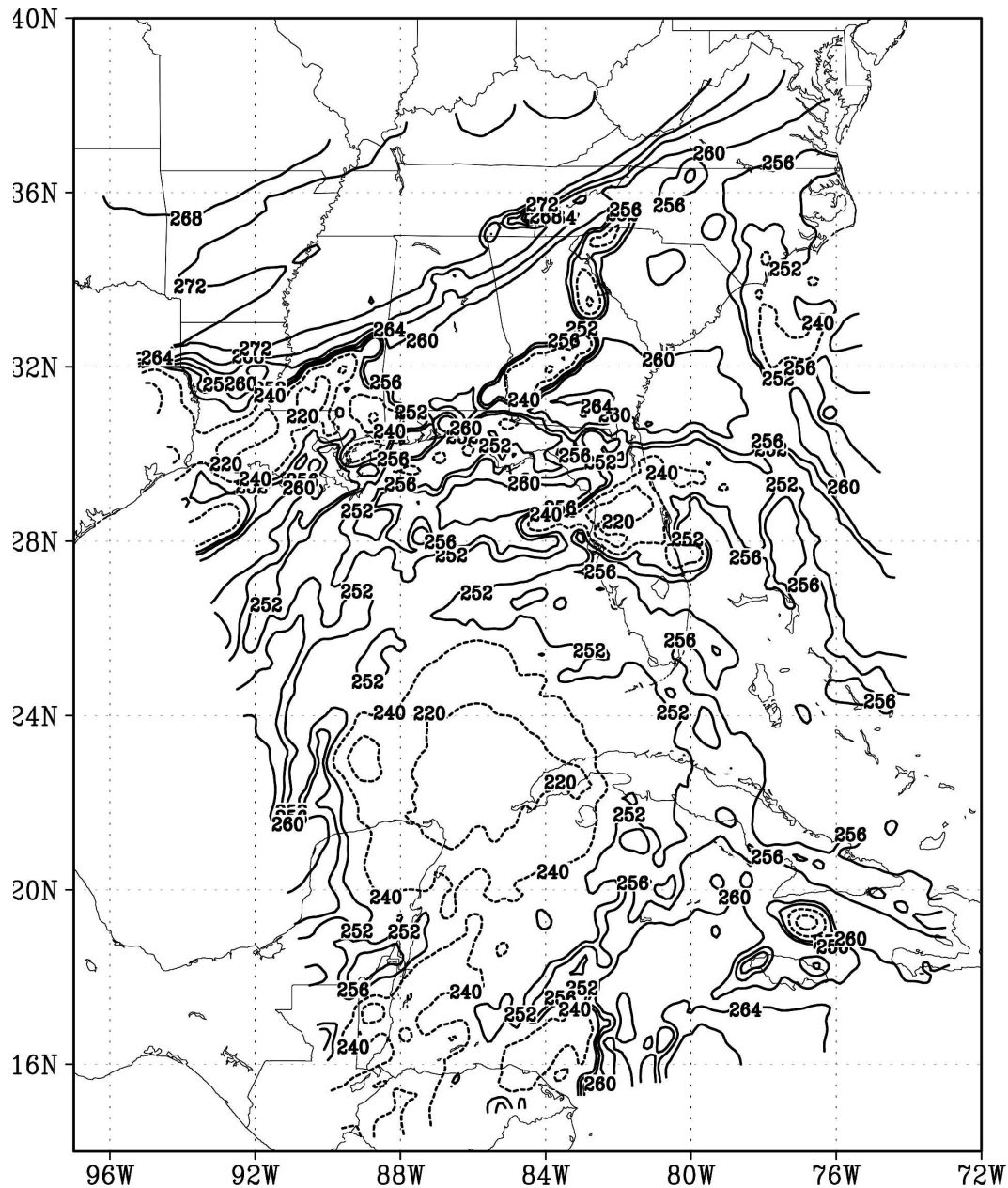


FIG. 3. AIRS observed brightness temperature (K) at the $6.6\text{-}\mu\text{m}$ water vapor channel for granule 188 (1847–1853 UTC 11 Jul 2003).

data. This is done by adding those channels whose COP levels are above the cloud top at each pixel, and thus whose radiance values are minimally affected by clouds. This is possible because emission at each spectral channel comes from varying levels in the atmosphere. For example, if a particular channel contains no emission from the lower atmosphere, then radiance measurement at this channel would not be contaminated by low-level stratus clouds. Therefore it would be usable as an additional clear-channel AIRS measure-

ment, which would otherwise be deemed cloudy and discarded.

4. Limited cloud-contaminated data removal algorithm

a. Theoretical aspects

Utilizing MODIS cloud-top pressure fields to remove cloud-contaminated channels at each pixel is not as straightforward as using any channel whose peak

weighting function height is above the cloud top. The radiance at each channel is a combination of emission from several vertical levels near the peak weighting function height. The relative contribution of each layer can be quantified by examining the vertical profile of the weighting function (WF), which is the pressure differential of the layer-to-space transmittance at each vertical level l , and spectral channel v , such that

$$W_l^v = \frac{\partial \tau_l^v}{\partial p}. \quad (3)$$

In essence, the WF quantifies the change in atmospheric transmission at each level; if the transmission changes significantly from one level to the next, the WF value would be large, as there would be a spike in atmospheric emission at that level. For this work, the WFs have been normalized and are dimensionless.

For some channels, the WFs are broad; that is to say, some channels have radiances that are the product of nearly equal amounts of atmospheric emission from a deeper atmospheric layer. On the other hand, some channels have WFs that are rather narrow, where the radiance is mainly the product of emission from a shallower atmospheric layer. Therefore, while a channel's peak WF may lie above a cloud top, the channel may still exhibit cloud contamination if the cloud top is sufficiently close to the peak emission level.

To account for this, a channel-dependent data removal algorithm is derived that removes cloud-contaminated AIRS channels on a pixel-by-pixel basis using the cloud-top information from MODIS and the WF structure at each channel and AIRS pixel.

b. Limited cloud-contaminated data removal algorithm

1) LCCDR CUTOFF PRESSURE TEST

The LCCDR algorithm consists of two steps: a COP test and a biweight test. To minimally remove cloud-contaminated AIRS observational data, the first step of the proposed LCCDR algorithm takes two factors into account: the structure of the WF and the cloud height at each AIRS pixel. To do this, the algorithm first constructs WFs using the transmittances at each channel and pixel obtained from the RTM. The WFs are calculated at each pixel within the domain at 97 vertical levels, which extend from 1000.0 to 0.0050 hPa. The algorithm then examines the structure of each WF in order to set a COP level, which is defined as the uppermost atmospheric level at or below which a cloud can exist and not cause cloud contamination of AIRS radiance at a selected channel and the corresponding pixel. In other words, if a cloud top is above the COP

level, radiance at that channel would be considered cloud contaminated.

The COP level is set for each channel and pixel based on the structure of the corresponding WF profile. The COP level is determined by a preselected ratio of the area under the WF profile curve above and below the COP level. Through an extensive study, utilizing most of the 2378 AIRS channels, a ratio of 1/4 has been determined to be the best for estimating the COP level,

$$\frac{\int_{p_{\text{sfc}}}^{p_{\text{COP}}} W_l^v d(\ln p)}{\int_{p_{\text{COP}}}^0 W_l^v d(\ln p)} \approx \frac{1}{4}. \quad (4)$$

Generally speaking, this ratio is set such that the combined emission from each level above the COP level is 4 times that which comes from the levels below. This ratio can be reset by the user depending on how sensitive to cloud effects one expects the algorithm to be.

The LCCDR algorithm constructs the WFs at each pixel and channel and then evaluates the ratio described by Eq. (4), first setting the COP level to be nearest to the surface. The ratio is calculated iteratively, setting the COP level to successively higher vertical levels until (4) is satisfied. Once (4) is satisfied, the corresponding COP level is set for that channel and pixel.

Channels with identical peak WF heights may not have identical COP levels. This is due to the difference in structure of the WF between channels. Some channels have "broad" WFs, whereas others have "narrow" WFs. This would impact the evaluation of Eq. (4) and therefore produce different COP levels. For instance, for two channels (one with a narrow WF and the other with a broad WF) and with identical peak WF heights, the COP level for the broad channel will be lower in the atmosphere than for the narrow channel. This feature is illustrated in Fig. 4.

If the COP level is found to exist at the surface for the channel in question, the algorithm eliminates this channel, regardless of the cloud height. Otherwise, each channel's COP level is compared to the MODIS cloud height for the AIRS pixel under investigation. If the MODIS data indicate that the cloud height is above the COP level for a particular channel, the algorithm excludes this channel for that particular pixel; if not, then this channel is allowed to pass. In addition to this, if a pixel is found to be cloud free as defined by the MODIS cloud mask, then all the channels (including near-surface channels) are passed for that pixel. This ensures that, at the very least, all clear FOV near-surface channels will pass the COP test.

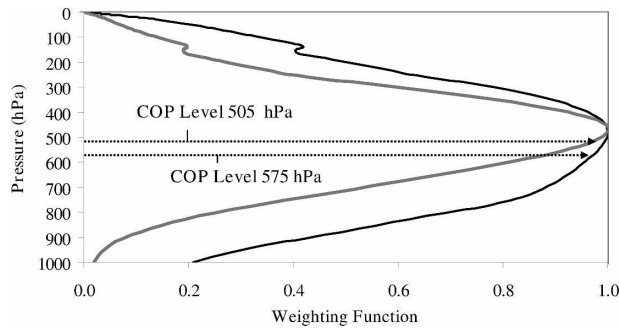


FIG. 4. The COP levels and the WF profiles for two channels with similar peak WF heights but different shapes: narrower channel 215 (14.06 μm , gray) and broader channel 264 (13.79 μm , black).

Figure 5 compares the average COP levels to the average peak WF levels for the MM5/AIRS overlapping domain for AIRS channels 1–1864 (Fig. 5a), as well as example spatial distributions for two AIRS channels, channel 201 (Fig. 5b) and channel 1583 (Fig. 5c). Figure 5a is sorted by decreasing peak WF level, with standard errors displayed at every 75 channels. The COP level is consistently below the peak WF level, ensuring that using the COP level (instead of the peak WF) will result in a more robust procedure to remove cloudy data. In addition to this, the COP level responds to changes in the WF. For a carbon dioxide channel (channel 201; Fig. 5b), the COP level varies from 340 to 420 hPa and responds in general to the modeled scan angle of the AIRS instrument (like the WF), producing a vertically lower COP level (420 hPa) at nadir than that on the edges of the scan swath (330 hPa). For a water vapor channel (channel 1583; Fig. 5c), the COP level responds to changes in profile moisture, as the COP level is vertically lower (720 hPa) for relatively dry atmospheric profiles (region across western Tennessee and Arkansas, behind the moisture gradient) than for relatively moist profiles (420 hPa in region ahead of the moisture gradient over Georgia and Florida). Figure 5d shows the difference between the COP levels and peak WF pressure levels for channel 201. As can be seen here, the difference is fairly uniform (between 50 and 100 hPa) across the domain—there is some hint of the moisture gradient as indicated by the elevated WF – COP difference across eastern Tennessee, through north Alabama and central Mississippi. Figure 5e shows this same difference, but for channel 1583. Here the COP level is vertically closer to the peak WF level where the atmosphere is drier (behind the moisture gradient, over western Tennessee and Arkansas) than it is for a relatively more moist atmosphere (ahead of the moisture gradient).

2) RESIDUAL OUTLIER IDENTIFICATION, THE LCCDR BIWEIGHT TEST

The COP test of the LCCDR algorithm effectively removes much of the cloud contamination; however, some residual cloud contamination may still be present. For example, the MODIS cloud screening has difficulty with the detection of thin cirrus because of the fact that the cloud and clear-sky temperatures are nearly the same in the presence of thin cirrus for those MODIS test channels that were used for determining cloud-top pressure. [See online at http://modis-atmos.gsfc.nasa.gov/MOD06_L2/qa.html.] However, in this instance where the cloud mask does not detect thin cirrus (or any other cloud type for that matter), the impact of cloud contamination is small for those channels that are not greatly affected by thin cirrus clouds. For those channels that are *greatly* affected by thin cirrus, the model and observation differences will be large and the biweight test of the LCCDR algorithm, to be described below, would remove that data point. This second step of the LCCDR algorithm identifies and removes any remaining outliers from the first step prior to any attempt to assimilate the AIRS data.

The biweight method for estimating the mean and standard deviation is used in this project (Lanzante 1996) to identify and remove the outliers. This method has been shown to be extremely useful in GPS data assimilation studies (Zou and Zeng 2006), and its application to AIRS BT is straightforward.

The statistical measure of relative error (and not the BTs themselves),

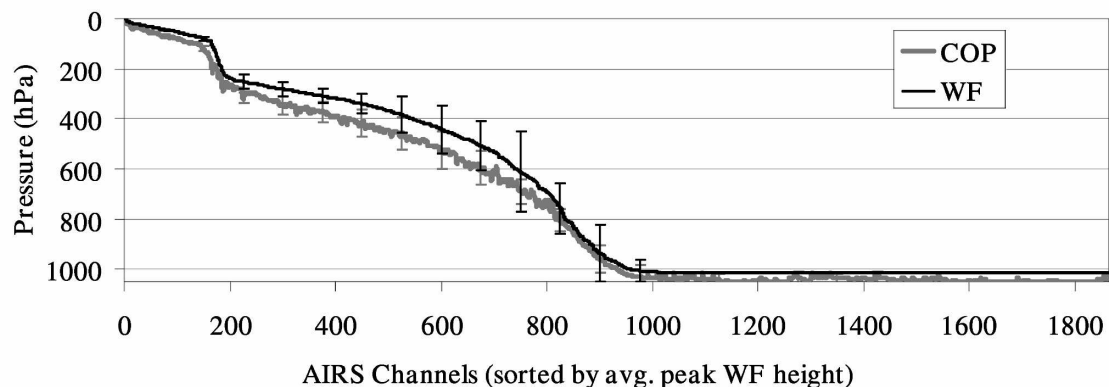
$$X_i^j = \frac{R_i^{j,\text{obs}} - R_i^{j,\text{rtm}}}{R_i^{j,\text{rtm}}}, \quad (5)$$

is used to determine if a data point can be flagged as an outlier, where $R_i^{j,\text{obs}}$ and $R_i^{j,\text{rtm}}$ are AIRS observed and SARTA-modeled radiance of the j th channel at the i th observation location. First, the median (M^j) and the median absolute deviation (MAD^j , defined as the median of the absolute values of the deviations of the dataset values from the median) of X_i^j ($i = 1, \dots, n$) are calculated. From M^j and MAD^j , a weighting function (w_i^j) corresponding to each observation (X_i^j) is computed as follows:

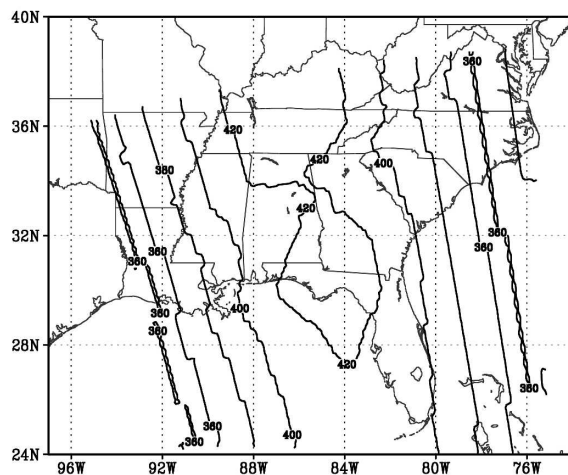
$$w_i^j = \frac{X_i^j - M^j}{c\text{MAD}^j}, \quad (6)$$

where c is a “censor” parameter value such that all data beyond a certain critical distance from c are given zero weight. The censor value used for this work is 7.5, which was used by both Lanzante (1996) and Zou and Zeng

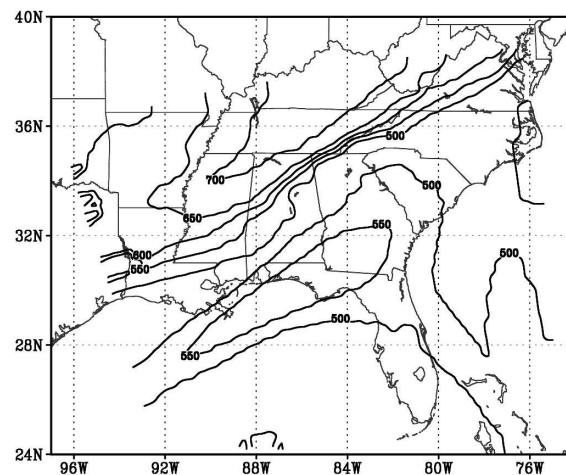
a.



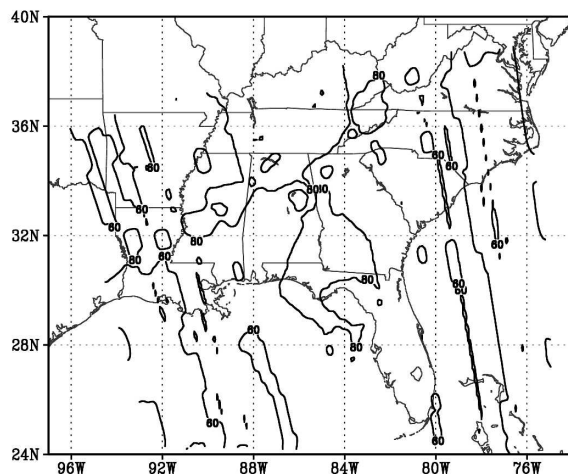
b.



c.



d.



e.

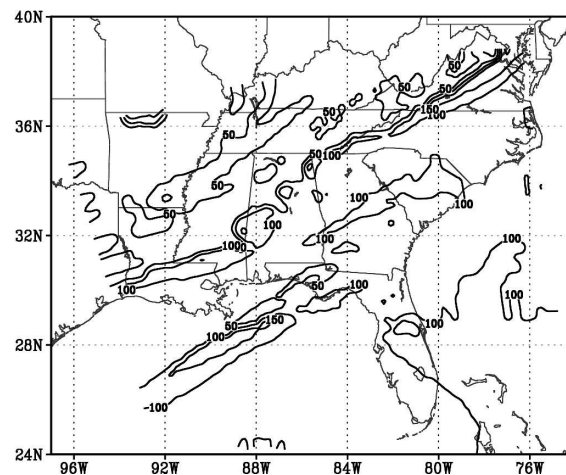


FIG. 5. (a) Average peak RTM WF and LCCDR COP pressure levels (sorted from highest to lowest peak WF level) for AIRS channels 1–1864. Error bars displayed for every 75 channels. The spatial distribution of LCCDR COP levels across the MM5/AIRS overlapping domain for the (b) 14.14- μm channel (AIRS channel 201) and (c) 7.13- μm channel (AIRS channel 1583). Spatial distribution of the difference between the COP and the peak WF pressure levels across the MM5/AIRS overlapping domain for the (d) 14.14- μm channel (AIRS channel 201) and (e) 7.13- μm channel (AIRS channel 1583).

(2006). In addition to this, w_i^j is set to 1.0 for any $|w_i^j| > 1.0$ to compute the censoring. The biweight mean \bar{X}_{bw}^j of the relative error of radiance at each channel j is defined as

$$\bar{X}_{bw}^j = M + \frac{\sum_{i=1}^n (X_i^j - M)[1 - (w_i^j)^2]^2}{\sum_{i=1}^n [1 - (w_i^j)^2]^2}, \quad (7)$$

and the biweight standard deviation (BSD), SD_{bw}^j , is defined by

$$SD_{bw}^j = \frac{\sqrt{n \sum_{i=1}^n (X_i^j - M)^2 [1 - (w_i^j)^2]^4}}{\sum_{i=1}^n [1 - (w_i^j)^2][1 - 5(w_i^j)^2]}. \quad (8)$$

It is clear from (7) and (8) that the data toward the center of the distribution are weighted more heavily than data near the tails of the distribution. This allows the biweight mean and BSD to be more resistant to outliers than the traditional mean and standard deviation, which applies equal weighting throughout the distribution. The biweight mean and BSD are then used to determine the so-called Z score of any particular observation using

$$Z_i^j = \frac{X_i^j - \bar{X}_{bw}^j}{SD_{bw}^j}. \quad (9)$$

AIRS BTs are removed if their corresponding Z score is greater than 2. As will be shown shortly, this second step of the LCCDR algorithm effectively eliminates any remaining outliers that may have passed through the LCCDR COP test.

The LCCDR algorithm is developed to serve the purpose of assimilating AIRS data using three-dimensional or four-dimensional variational data assimilation (3DVAR or 4DVAR) approaches. It is therefore best to remove those data that differ too much from the model fields and render a non-Gaussian error distribution of the sum of the observation and model errors. In doing so, the biweight portion of the proposed LCCDR algorithm might remove some data that are correct and might provide useful information to improve the model, since data that deviate greatly from the model can be caused by a large model error rather than a large observation error. In other words, consideration of the abilities of the 3DVAR/4DVAR schemes to assimilate AIRS data suggests that the need to develop a robust

scheme comes at the expense of some good observed data.

5. Results obtained from the LCCDR algorithm

a. LCCDR clear-channel identification

For the selected case (see section 2c) involving a strong midlevel moisture gradient located over the southeastern United States between 0800 UTC 11 July and 0800 UTC 12 July 2003, a total number of 3800 observational points are present in the AIRS/MM5 overlapping domain (see Fig. 6) at 1800 UTC 11 July 2003. If all cloudy data points are excluded, only 446 clear-sky data points (12%) are left within this domain at this time (black pixels in Fig. 6). However, there are a total of 1882 data points (50%) where there is no cloud above 800 hPa (gray pixels in Fig. 6). Therefore, there would be 1882 data points with channels whose COP levels lie above 800 hPa that are usable (black and gray pixels in Fig. 6). Figure 7 displays the percentage of the total number of AIRS pixels (y axis) that are free of cloud above the heights indicated by different pressure levels on the x axis. For instance, for all channels whose COP level is above 400 hPa, the total number of usable pixels could be as high as 2640, or 69% of the total available data.

Two AIRS spectral channels, a carbon dioxide channel at $14.14 \mu\text{m}$ (channel 201) and a water vapor channel at $7.13 \mu\text{m}$ (channel 1583), are used for an illustration of what types of data points are removed by the two sequential steps in the LCCDR algorithm. Figures 8 and 9 show the AIRS-observed (top panels) and model-simulated (bottom panels) BTs at these two channels (14.14 and $7.13 \mu\text{m}$) and the MODIS cloud-top pressures at all the pixels within the MM5/AIRS overlapping domain at 1800 UTC 11 July 2003. Points deemed cloudy by the first step of the LCCDR algorithm appear in blue; outliers identified by the second step of the LCCDR algorithm are shown in red. We observe that at places (geographic locations) where cloud-top pressure is below 400 hPa (cloud top vertically higher), the mean value of the observed BTs decreases consistently with the decrease of cloud-top pressure (top panels in Figs. 8 and 9). This is not seen in model simulations (bottom panels in Figs. 8 and 9), as cloud effects are not taken into consideration in the SARTA model. It is also noticed that the model-simulated BTs have a cold bias for those pixels with MODIS cloud-top pressures above the corresponding LCCDR COP levels; this is because the bias removal is done only after the LCCDR COP test has removed these cloudy data. These results further confirm that the COP levels determined by the LCCDR algorithm

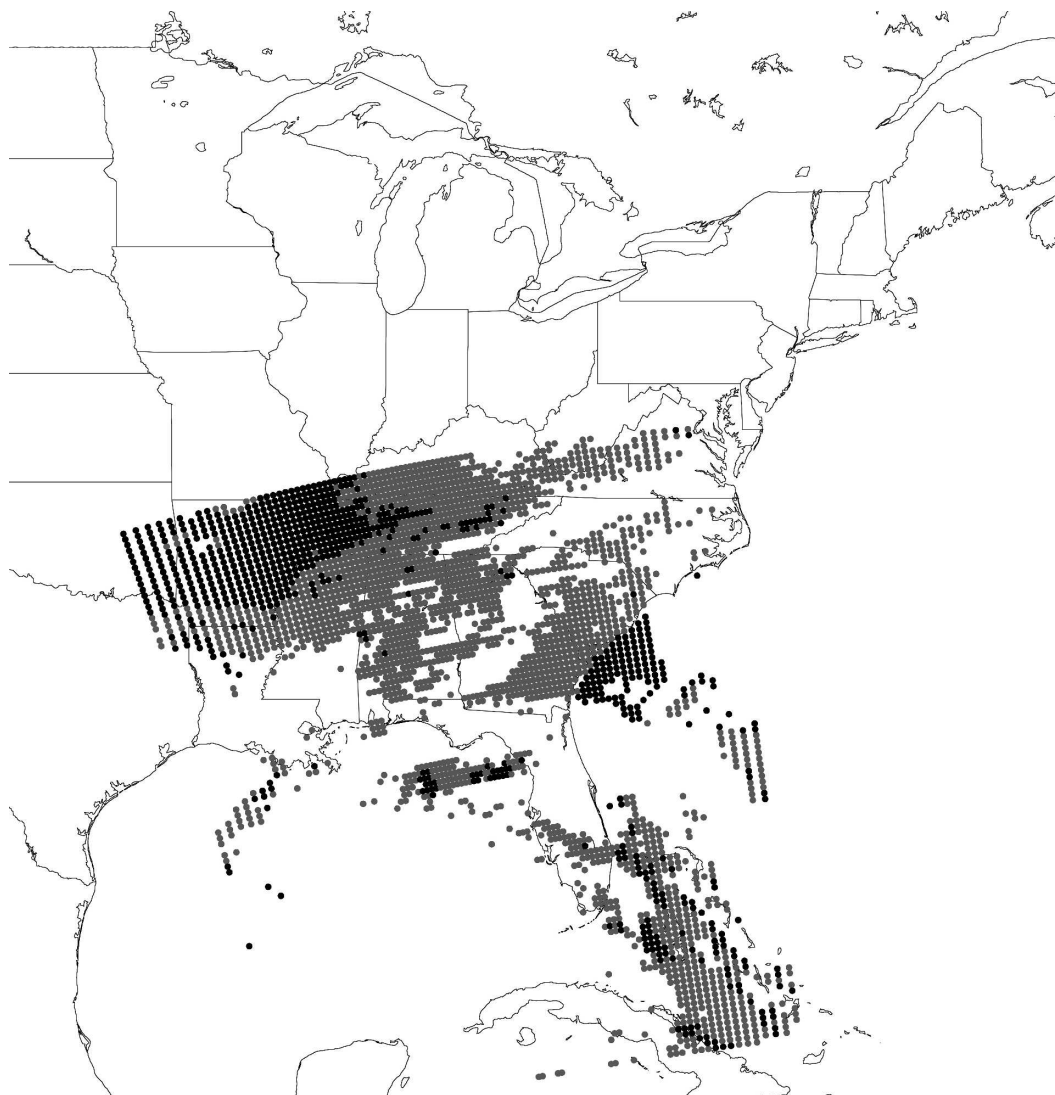


FIG. 6. MM5/AIRS overlapping pixels that contain no clouds (black pixels) and at which clouds do not exist above 800 hPa (gray pixels) near 1800 UTC 11 Jul 2003.

work well. Some data points identified as outliers by the biweight method (indicated in Figs. 8 and 9 as open dots) are intermixed with data points deemed usable by the LCCDR algorithm (black dots) for both observed

and simulated values of BT. These are points at which the observed and simulated BTs differ greatly. Characteristics of these outliers are therefore best illustrated in the following figures.

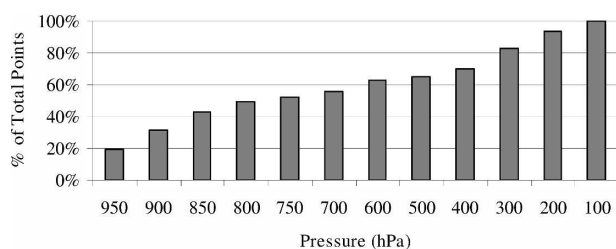


FIG. 7. The percentage of the total number of AIRS pixels (y axis) that are free of cloud above the heights of different pressure indicated on the x axis at 1800 UTC 11 Jul 2003.

Figures 10 and 11 show the comparison of AIRS observed and model-simulated (MM5/SARTA) BTs for all the points shown in Figs. 8 and 9, respectively. MM5/SARTA-simulated BTs are calculated using MM5 18-h forecast data from the test case described previously. Four distinct data types can be seen in Figs. 10 and 11: 1) data flagged as cloud contaminated by the LCCDR COP test and removed (gray diamonds), 2) data identified by the LCCDR biweight test as residual cloud contamination and removed (open dots), 3) data that have passed through both quality control checks and are considered to be clear-channel data by the LCCDR

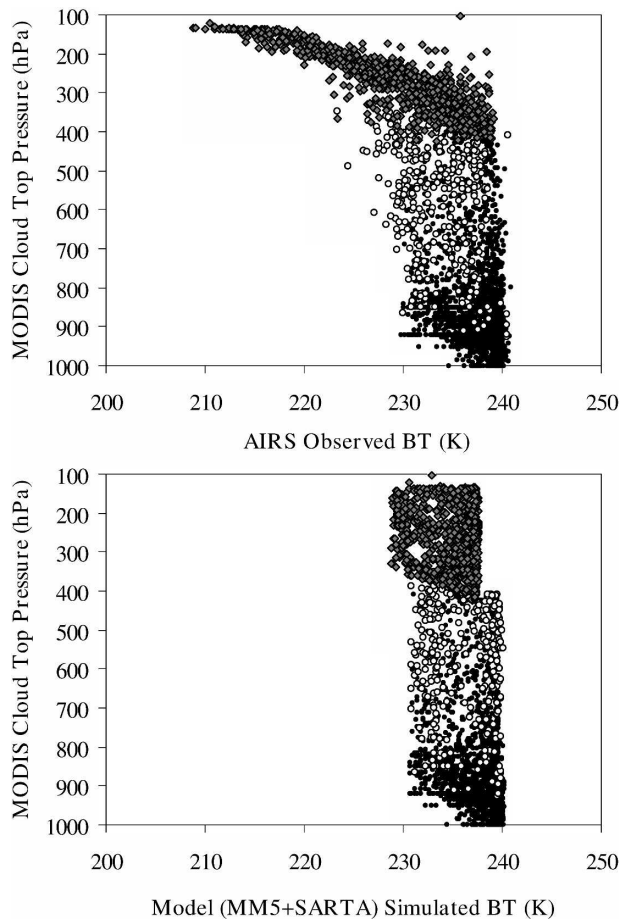


FIG. 8. MODIS cloud-top pressure (y axis) and (top) AIRS observed and (bottom) model-simulated BTs at the $14.14\text{-}\mu\text{m}$ carbon dioxide channel (channel 201; x axis) at all pixels within the MM5/AIRS overlapping domain at 1800 UTC 11 Jul 2003. Gray diamonds illustrate those points excluded by the LCCDR COP test, open dots indicate points excluded by the LCCDR biweight test, and black dots indicate those data points that successfully passed the LCCDR algorithm test.

algorithm (black dots), and 4) those data that are at clear-sky points (gray dots).

For the $14.14\text{-}\mu\text{m}$ channel, there are a great number of scattered outliers for which the observed BTs are colder than MM5/SARTA simulations (Fig. 10). The LCCDR clear-channel data (black and gray) show good agreement between MM5/SARTA and AIRS (correlation above 0.98) while retaining 61% of the total available domain points. For the $7.13\text{-}\mu\text{m}$ water vapor channel, the observed BTs are lower than simulated values at low BTs (Fig. 11). Overall, the correlation is high (above 0.96) while nearly 59% of the total available domain points are retained. As can be seen in these two figures, the LCCDR COP test identifies much of the cloud contamination, as those data that are severely cloud contaminated (indicated by those data

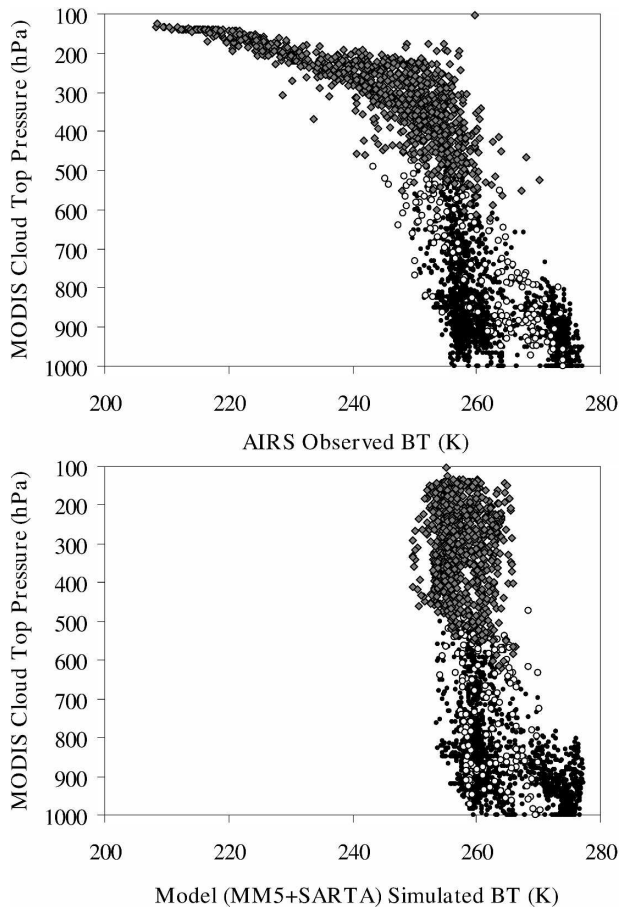


FIG. 9. Same as Fig. 8, but for the $7.13\text{-}\mu\text{m}$ water vapor channel (channel 1583).

that exhibit much colder AIRS BT values than MM5/SARTA BT values because of cloud effects) are removed from consideration by the LCCDR COP test. The data field is further thinned by the LCCDR biweight test, as some additional data that deviate from model simulations are flagged as outliers (either due to residual cloud contamination or some other as-yet-unidentified reasons, such as model deficiencies). The clear-channel data exhibit good agreement between AIRS observations and MM5/SARTA simulations.

Figure 12 shows the probability density function (PDF) of the differences between model-simulated and AIRS observed BTs for the two sample channels (201 and 1583) over the four AIRS granules. The clear-sky PDFs match very well with the LCCDR clear-channel PDFs, suggesting that the clear-channel data selected by the LCCDR algorithm possess the same characteristics as clear-sky data.

The results shown above include data from AIRS granule 188, which contains data between 1847 and 1853 UTC 11 July 2003. However, within the 36-h

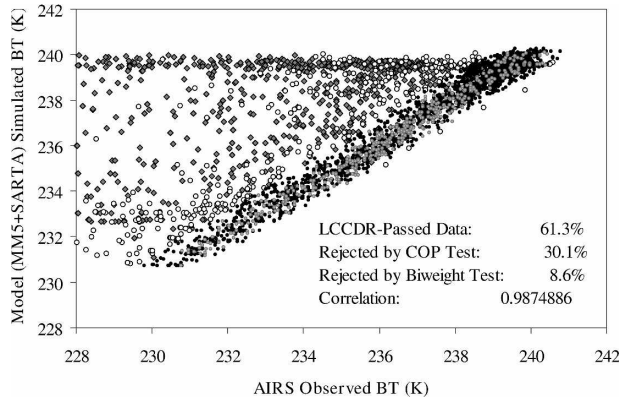


FIG. 10. AIRS observed (x axis) vs model-simulated (MM5 and SARTA) BTs at 1800 UTC 11 Jul 2003 for the $14.14\text{-}\mu\text{m}$ carbon dioxide channel (channel 201). Model data are generated from 18-h MM5 forecast data. Gray diamonds illustrate those points excluded by the LCCDR COP test, open dots indicate points excluded by the LCCDR biweight test, black and gray dots indicate those data points that successfully passed the LCCDR quality control method, and gray dots indicate clear-sky points (no cloud at any vertical levels).

MM5 forecast window for this case, there are another three different AIRS granules, which were also embedded within the MM5 domain (see section 2c). We have applied the LCCDR algorithm to all these data. Figure 13 shows the total number of clear-sky data points (gray line) and the total number of data points with clear channels (black line) for the first 1864 AIRS channels arranged by decreasing wavelength (Fig. 13, top panel) and increasing COP level (Fig. 13, bottom panel). As can be seen here, the total number of usable points increases except for window channels (Fig. 13, top panel). Channels that have a larger data increase are those that have higher COP levels.¹ For example, 80% (58%) of data points have usable AIRS channels whose COP levels are at or above 300 hPa (500 hPa). As the COP level lowers to the surface, all BTs must be removed at cloudy points, leaving only 20% of data that have clear FOV.

While many less data are removed, the root-mean-square error (RMSE) and mean absolute error (MAE) are both significantly reduced and the correlation coefficients are greatly increased after the LCCDR algorithm is applied. Figures 14a,b,c show RMSE, MAE, and correlation coefficients of model-simulated and AIRS observed BTs for the AIRS channels from 1 to 1864 at 1800 UTC 11 July 2003 for all AIRS data (thick black), clear-channel data (gray), and clear-sky data (thin black). The MAEs, RMSEs, and correlation co-

¹ The COP level is the level below which cloud can exist without causing cloud contamination.

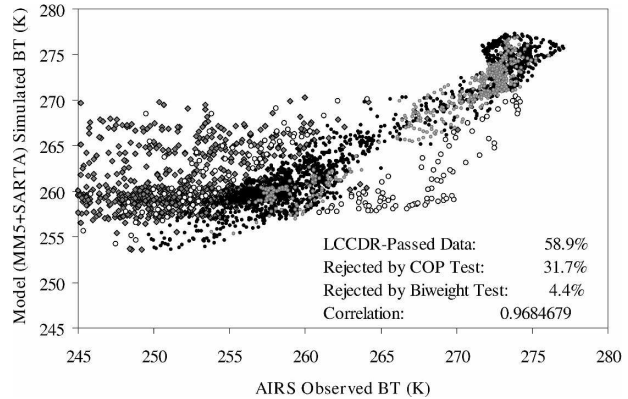


FIG. 11. Same as Fig. 10, but for the $7.13\text{-}\mu\text{m}$ water vapor channel (channel 1583).

efficients for all channels are shown in Table 1. For these statistics, channels at wavelengths at and smaller than $5\text{ }\mu\text{m}$ (channels past 1864) have been excluded because of difficulties with sun glint in the shortwave part of the AIRS spectrum. The clear-channel MAEs, the RMSEs, and correlation coefficients are nearly identical to those for clear-sky data. In fact, MAEs and RMSEs (correlation coefficients) of clear-channel data are slightly smaller (larger) than those of clear-sky data for water vapor channels. These results suggest that a good sample of clear channels within cloudy FOV are selected by the LCCDR algorithm.

b. Sensitivity of LCCDR to errors in MODIS cloud-top pressures and MM5 forecast

The LCCDR algorithm is dependent upon not only the accuracy of the MODIS cloud-top pressure levels, but also the reliability of the MM5 forecast data, which provide input to SARTA for, among other variables, WFs at each AIRS channel. Errors in the MODIS cloud-top pressures, or SARTA WFs—which impact the calculation of the LCCDR COP level—would seem to have the potential to reduce the effectiveness of the algorithm to identify clear channels. However, the LCCDR algorithm has been designed to account for reasonable amounts of error in both the MODIS cloud-top levels and errors in the COP level (caused by MM5 forecast deficiencies). If error exists in the MODIS cloud-top pressures, or an error in the COP level is such that the LCCDR COP test portion of the algorithm allows some “cloudy” channels to pass, the biweight portion of the algorithm screens these data and removes residual cloud-contaminated channels (based on their high Z-score values).

To illustrate this feature, the ability of the algorithm to handle error in these critical fields is examined. First,

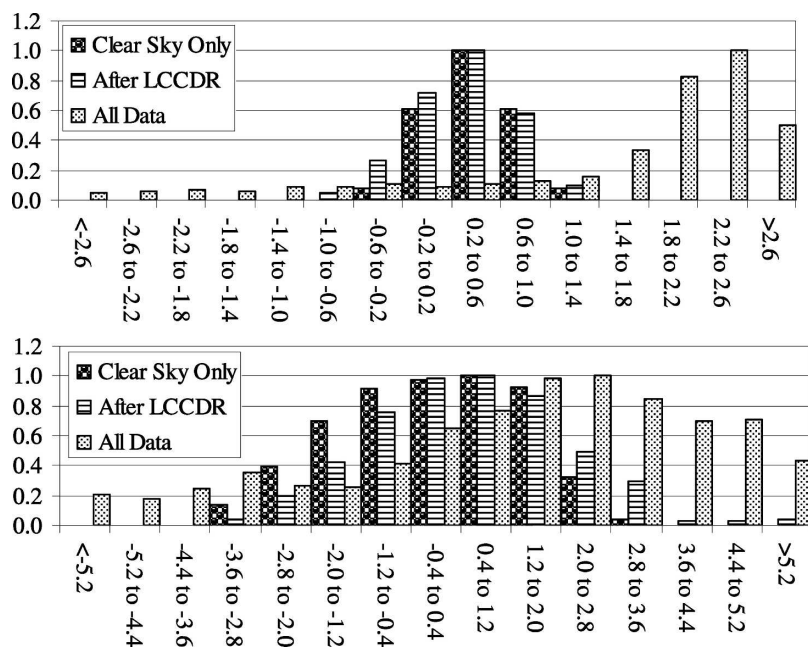


FIG. 12. A normalized PDF distribution of the differences between model-simulated and AIRS observed BTs for the AIRS channels (top) 201 and (bottom) 1583 at 1800 UTC 11 Jul 2003, including all AIRS pixels (small dots), those that remain after applying the LCCDR algorithm (lines), and the clear-sky pixels (large dots).

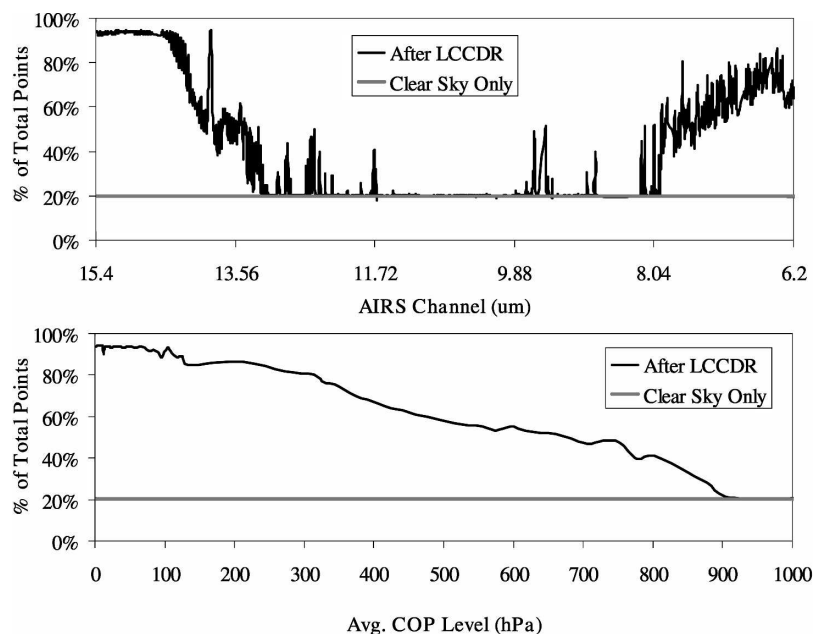


FIG. 13. Percent (of total) of clear-sky points (gray line) and domain points that have clear channel(s) (black line) for the first 1864 AIRS channels arranged in decreasing order of (top) wavelength and (bottom) COP level. Data from the following four AIRS scan swaths are included: granule 078 (0747–0753 UTC 11 Jul 2003), granule 188 (1847–1853 UTC 11 Jul 2003), granule 189 (1853–1859 UTC 11 Jul 2003), and granule 069 (0653–0659 UTC 12 Jul 2003).

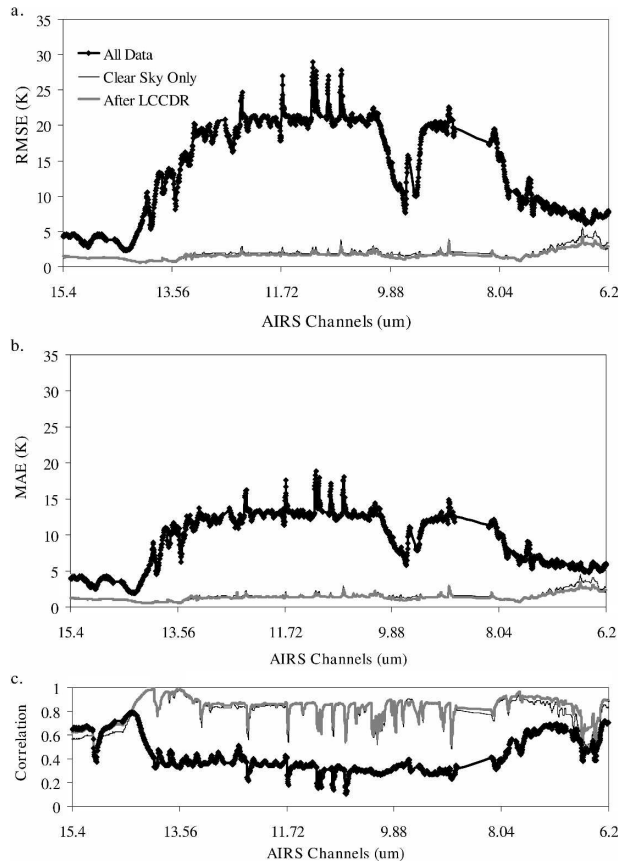


FIG. 14. (a) RMSE (K), (b) MAE (K), and (c) correlation coefficients of model-simulated and AIRS-observed BTs for the AIRS channels from 1 to 1865 averaged over four AIRS scan swaths (granule 078 at 0747–0753 UTC 11 Jul 2003; granule 188 at 1847–1853 UTC 11 Jul 2003; granule 189 at 1853–1859 UTC 11 Jul 2003; and granule 069 at 0653–0659 UTC 12 Jul 2003) for all AIRS pixels (thick black), those that remain after applying the LCCDR algorithm (gray), and the clear-sky pixels (thin black).

the sensitivity of the WF- and COP-level calculations are addressed when a reasonable amount of random error is added to the MM5 temperature and moisture profiles (2° – 3° to the temperature profile and about

10% to the mixing ratio). Figures 15 and 16 show the changes in the WFs and COP levels for two test channels, AIRS channels 201 and 1583 (14.14 and 7.13 μm , respectively). In both of these cases, the WFs do not shift much, and the corresponding COP levels shift very little as well. In fact, for the 14.14 - μm channel (Fig. 15), the COP levels are nearly the same, and for the 7.13 - μm channel (Fig. 16), the COP level drops only 20 hPa (from 515 to 535 hPa) in response to changes in the WFs.

So how would an error in the COP level or MODIS cloud-top pressure affect the LCCDR algorithm? Figures 17 and 18 show the comparison of AIRS and MM5/SARTA BTs for the two test channels (201 and 1583, respectively) when an exaggerated error is added to the MODIS cloud-top heights and COP levels. The top panels in Figs. 17 and 18 show the result when 100 hPa is added (subtracted) from the MODIS cloud-top levels (LCCDR COP levels); the bottom panels show the comparisons after 100 hPa is added (subtracted) to the LCCDR COP levels (MODIS cloud-top heights), with LCCDR clear-channel data shown in black dots, LCCDR COP test cloudy data in gray dots, LCCDR biweight outliers in open dots, and clear-sky data in gray dots. Comparing the error response for channel 201 (Fig. 17) with the original figure (Fig. 10), the result in adding 100 hPa to the MODIS cloud-top heights causes the apparent cloud-top heights to be *lower* in the atmosphere than in reality, causing more cloudy BT values to pass through the LCCDR COP test (30.1% screened before adding error, 16.4% after). This effect is equivalent to subtracting 100 hPa from the COP levels. Figure 17 shows that these cloudy data are effectively removed by the biweight test of the LCCDR algorithm (an increase from 8.6% to 16.9%). The correlation between the AIRS and MM5/SARTA values remains high (>0.98). The result of reducing the MODIS cloud-top pressures by 100 hPa (bottom panel, Fig. 17), thereby making the cloud tops appear *higher* in the atmosphere than in reality, has the opposite effect

TABLE 1. RMSE, MAE, and correlation coefficients of model-simulated and AIRS-observed BTs from channels 1 to 1864 over four AIRS scan swaths: G078 at 0747–0753 UTC 11 Jul 2003, G188 at 1847–1853 UTC 11 Jul 2003, G189 at 1853–1859 UTC 11 Jul 2003, and G069 at 0653–0659 UTC 12 Jul 2003.

Swath	All data			Clear-sky only				LCCDR-passed data			
	RMSE	MAE	Correlation coef	RMSE	MAE	Correlation coef	% Tot points	RMSE	MAE	Correlation coef	% Tot points
G078	11.94	7.02	0.48	1.57	1.32	0.86	32%	1.70	1.37	0.84	75%
G188	16.59	10.97	0.33	1.97	1.62	0.76	12%	2.37	1.88	0.73	60%
G189	12.88	9.38	0.71	1.85	1.58	0.83	19%	2.29	1.86	0.83	61%
G069	13.94	9.63	0.25	1.60	1.29	0.84	15%	1.81	1.44	0.78	65%

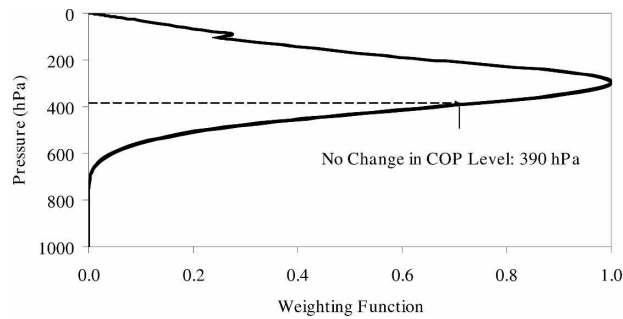


FIG. 15. The WFs for the 14.14- μm channel (AIRS channel 201) with (thin line) and without (thick line) adding errors (about 2° for temperature and 10% for specific humidity) to MM5 forecast fields. Corresponding LCCDR COP levels are indicated by dashed lines.

and is equivalent to adding 100 hPa to the COP levels. The number of pixels removed by the LCCDR COP test increases from 30.1% to 35.8%, while the number of pixels removed by the biweight step reduces from 8.6% to 6.2%. The overall number of passed pixels remains high (61.3% before adding error, 58.0% after) and the correlation is high as well. Similar results are found for channel 1583 (Fig. 18).

6. Summary and conclusions

AIRS infrared data are a highly valuable source of atmospheric information, not only because of their excellent spatial and temporal coverage but also because of their high spectral resolution. These data are already utilized in global forecasting applications; however, their usefulness in regional forecasting has not been thoroughly explored. To utilize AIRS data in a regional setting, the inherent difficulty with cloud contamination must be addressed. Operational centers, such as ECMWF, employ a “sorting” method to identify clear channels in cloudy FOV. This method, although accurate and reliable (Lavanant et al. 2004), requires the use

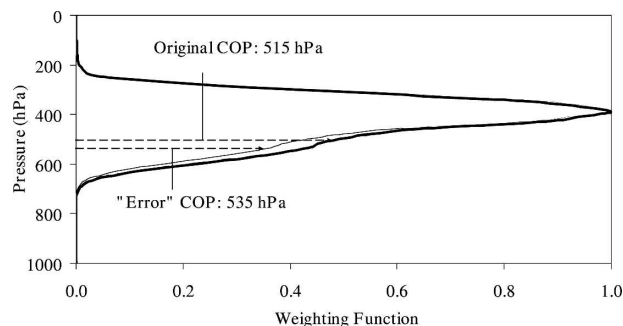


FIG. 16. Same as Fig. 15, but for the 7.13- μm water vapor channel (channel 1583).

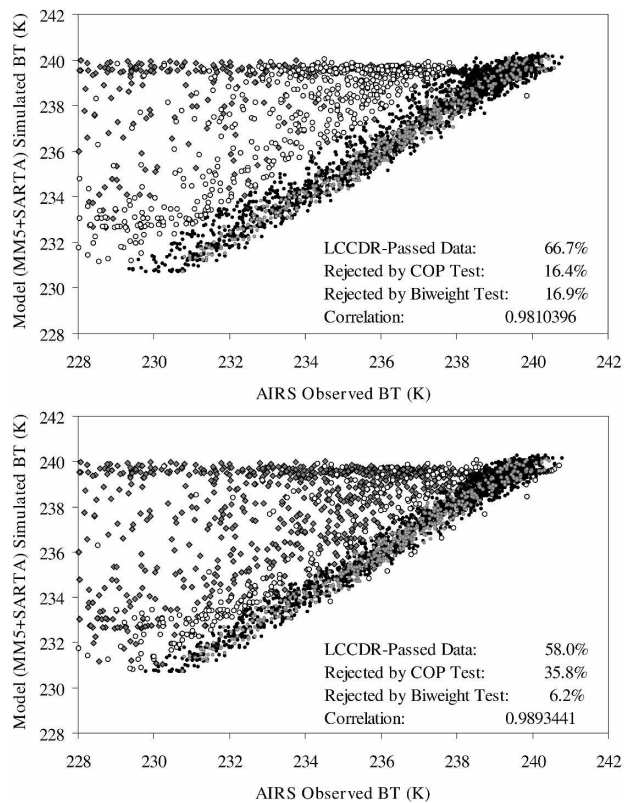


FIG. 17. Same as Fig. 10, but the MODIS cloud-top pressure levels have been (a) increased or (b) decreased by 100 hPa.

of an RTM capable of handling the scattering effects due to clouds. The RTM used for this work (SARTA) only simulates atmospheric radiances based on a non-scattering atmosphere; therefore, an alternate method is required. A channel-based cloud-contaminated data screening process, the LCCDR algorithm, is thus proposed. The LCCDR algorithm is designed to combine the MODIS cloud-top pressure information from the *Aqua* satellite with channel and domain-specific WF information to determine the pressure level at or below which a cloud can exist without contaminating the observed radiance at a specific channel. This is done for each of the AIRS channels in question, at each pixel in the domain. The usefulness of this algorithm was demonstrated using a test case from 11 to 12 July 2003 over the southeastern United States. The SARTA model is used to simulate the BTs over the domain in question, using a 36-h MM5 forecast (initialized at 0000 UTC 11 July 2003) for input data.

The results of this test case show that the LCCDR algorithm identifies many AIRS channels that do not exhibit cloud contamination. This study has shown that although the algorithm eliminates a much smaller percentage of the cloud-contaminated data, the correlation

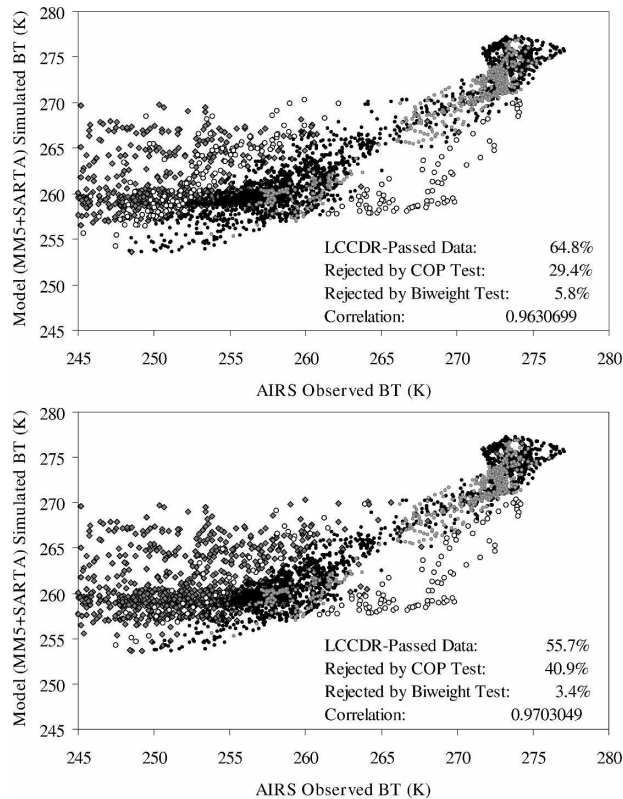


FIG. 18. Same as Fig. 11, but for the $7.13\text{-}\mu\text{m}$ water vapor channel (channel 1583).

coefficient between the AIRS observations and the MM5/SARTA-simulated BTs increases to nearly 0.9 in most cases and even exceeding 0.9 in some cases. The AIRS data is further thinned by the second part of the LCCDR algorithm, which uses a biweight method to estimate the mean and standard deviation of the dataset in order to identify and effectively remove remaining outliers. This step, in effect, removes not only those residual cloud-contaminated data, but also other outliers, which deviate greatly from model simulations because of model deficiencies (either the MM5 or SARTA). After the entire LCCDR algorithm has been applied to the AIRS dataset, the resulting data have improved agreement with the MM5/SARTA simulation, with very low RMSE and high correlation for most AIRS channels. Consisting of two steps, the COP test and biweight test, the LCCDR algorithm is not shown to be very sensitive to errors in MODIS cloud-top pressure or MM5/SARTA models.

The LCCDR algorithm for identifying clear channels from cloudy FOV, while being simple and inexpensive, ensures that a larger number of clear channels can be obtained from the AIRS data and used in a variational data assimilation scheme. In addition to this, the pro-

cessing of the MODIS cloud mask for AIRS FOV is also inexpensive as it involves a simple interpolation from MODIS space to AIRS FOV, facilitated by the finer resolution of the MODIS data (5 km) in comparison to AIRS data (13.5 km). Since the MODIS device and the AIRS instrument are both located on the *Aqua* satellite, the two types of data are available at the same time over the same area daily. This should allow the AIRS data to have a greater impact on the model initial conditions and mesoscale precipitation forecasts. This study is, however, limited to one case, and more case studies are needed. A further refinement of the LCCDR algorithm and a final testing of its impact on AIRS data assimilation for improving short-range QPFs are being conducted and results will be presented in a companion paper.

Acknowledgments. This research is supported by the NASA Project 20025157. Special thanks to Dr. Gary Jedlovec at NASA's Global Hydrology and Climate Center (GHCC) for many valuable discussions and for providing the MODIS cloud mask products.

REFERENCES

- Aumann, H. H., and Coauthors, 2003: AIRS/AMSU/HSB on the *Aqua* mission: Design, science objectives, data products, and processing systems. *IEEE Trans. Geosci. Remote Sens.*, **41**, 253–264.
- Barnes, W. L., T. S. Pagano, and V. V. Salomonson, 1998: Pre-launch characteristics of the Moderate Resolution Imaging Spectroradiometer (MODIS) on EOS-AM1. *IEEE Trans. Geosci. Remote Sens.*, **36**, 1088–1100.
- Blackadar, A. K., 1979: High resolution models of the planetary boundary layer. *Advances in Environmental Science and Engineering*, Vol. 1, J. Pfafflin and E. Ziegler, Eds., Gordon and Breach Science, 50–85.
- Chevallier, F., P. Lopez, A. M. Tompkins, M. Janisková, and E. Moreau, 2004: Prospects for assimilating cloudy radiances from AIRS. *Proc. Workshop on Assimilation of High Spectral Resolution Sounders in NWP*, Reading, United Kingdom, ECMWF, 123–132.
- Collard, A. D., 2004: Assimilation of AIRS observations at the Met Office. *Proc. Workshop on Assimilation of High Spectral Resolution Sounders in NWP*, Reading, United Kingdom, ECMWF, 63–71.
- Derber, J. C., and W.-S. Wu, 1998: The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. *Mon. Wea. Rev.*, **126**, 2287–2299.
- Dudhia, J., 1993: A nonhydrostatic version of the Penn State–NCAR Mesoscale Model: Validation tests and simulation of an Atlantic cyclone and cold front. *Mon. Wea. Rev.*, **121**, 1493–1513.
- Greenwald, T. J., R. Hertenstein, and T. Vukicevic, 2002: An all-weather observational operator for radiance data assimilation with mesoscale forecast models. *Mon. Wea. Rev.*, **130**, 1882–1897.
- Grell, G., J. Dudhia, and D. Stauffer, 1995: A description of the fifth-generation Penn State/NCAR Mesoscale Model

- (MM5). Mesoscale and Microscale Meteorology Division, NCAR Tech. Note NCAR/TN-398+STR, 117 pp.
- Haines, S. L., G. Jedlovec, and F. LaFontaine, 2004: Spatially varying spectral thresholds for MODIS cloud detection. Preprints, *13th Conf. on Satellite Meteorology and Oceanography*, Norfolk, VA, Amer. Meteor. Soc., CD-ROM, P7.9.
- Lanzante, J. R., 1996: Resistant, robust and nonparametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data. *Int. J. Climatol.*, **16**, 1197–1226.
- Lavanant, L., M. Dahoui, F. Rabier, and T. Auligné, 2004: Cloud detection for IASI/AIRS using imagery. *Proc. Workshop on Assimilation of High Spectral Resolution Sounders in NWP*, Reading, United Kingdom, ECMWF, 133–144.
- Liou, K.-N., S. C. Ou, Y. Takano, and Q. Liu, 2005: A polarized delta-four-stream approximation for infrared and microwave radiative transfer: Part I. *J. Atmos. Sci.*, **62**, 2542–2554.
- McNally, A. P., and P. D. Watts, 2003: A cloud detection algorithm for high-spectral-resolution infrared sounders. *Quart. J. Roy. Meteor. Soc.*, **129**, 3411–3423.
- MODIS Cloud Mask Team, 2002: Discriminating clear-sky from cloud with MODIS—Algorithm theoretical basis document (MOD35). Doc. ATBD-MOD-06, 112 pp. [Available online at http://modis-atmos.gsfc.nasa.gov/MOD35_L2/atbd.html.]
- Pagano, T. S., and Coauthors, 2002: On-board calibration techniques and test results for the Atmospheric Infrared Sounder (AIRS). *SPIE Int. Symp. on Optical Science and Technology*, Seattle, WA, SPIE, 4814–38.
- Strow, L. L., S. Hannon, S. D. Machado, H. Motteler, and D. Tobin, 2003: An overview of the AIRS radiative transfer model. *IEEE Trans. Geosci. Remote Sens.*, **41**, 303–313.
- Zou, X., and Z. Zeng, 2006: A quality control procedure for GPS radio occultation data. *J. Geophys. Res.*, **111**, D02112, doi:10.1029/2005JD005846.