The use of error estimates with AIRS profiles to improve short-term weather forecasts

Gary J. Jedlovec*[a], Shih-Hung Chou[a], Bradley T. Zavodsky[b], William Lapenta[a]
[a]Marshall Space Flight Center, 320 Sparkman Drive, Huntsville, AL, 35805
[b]University of Alabama in Huntsville, 320 Sparkman Drive, Huntsville, AL, 35805

ABSTRACT

The hyperspectral resolution measurements from the NASA Atmospheric Infrared Sounder (AIRS) are advancing climate research by mapping atmospheric temperature, moisture, and trace gases on a global basis with unprecedented accuracy. Using a sophisticated retrieval scheme, the AIRS is capable of diagnosing the atmospheric temperature in the troposphere with accuracies of less than 1 K over 1 km-thick layers and 10-20% relative humidity over 2 km-thick layers, under both clear and cloudy conditions. A unique aspect of the retrieval procedure is the specification of a vertically varying error estimate for the temperature and moisture profile for each retrieval. The error specification allows for the more selective use of the profiles in subsequent processing. In this paper, we describe a procedure to assimilate AIRS data into the Weather Research and Forecasting (WRF) model to improve short-term weather forecasts. The ARPS Data Analysis System (ADAS) developed by the University of Oklahoma is configured to optimally blend AIRS data with model background fields based on the AIRS error profiles. The WRF short-term forecasts with selected AIRS data show improvement over the control forecast. The use of the AIRS error profiles maximizes the impact of high quality AIRS data from portions of the profile in the assimilation/forecast process without degradation from lower quality data in the other portions of the profile.

Keywords: SPoRT, AIRS, data assimilation, numerical weather prediction, quality control, forecast improvement, satellite-retrieved profiles, ADAS, WRF

1. INTRODUCTION

The Atmospheric Infrared Sounder (AIRS) on NASA's Earth Observing System (EOS) Aqua satellite is currently being used to advance climate research and improve weather prediction on a global scale (Aumann et al. 2003). With access to real time AIRS data from direct broadcast receiving stations, the radiance and profile information can be used to address short-term weather problems. The NASA Short-term Prediction Research and Transition (SPoRT) Center seeks to accelerate the infusion of NASA Earth science observations, data assimilation, and modeling research into NWS forecast operations and the decision-making process (Goodman et al. 2004). The SPoRT program has been providing output from the Weather Research and Forecasting (WRF; Michalakes et al. 2001) model to several NWS Forecast Offices in the Southern Region for the past year for use as additional guidance in their forecast and decision-making process.

One limitation for weather forecasting is the lack of observational data over data sparse regions such as the ocean. A significant portion of the eastern half of the U.S. is surrounded water with little or no upper air observations in these ocean regions. Numerous weather systems develop, intensify, and or move through these regions and impact the weather over the region. This data void problem can be addressed with the use of AIRS Level-2 temperature and moisture profiles to provide data in regions where traditional rawinsonde observations are not available. This paper will describe a procedure to optimally assimilate AIRS data over the Gulf of Mexico and eastern Atlantic Ocean region and assess the subsequent impact on the short-term weather forecasts from a regional forecast model. The EOS science team Version 5.0 prototype temperature and moisture profiles are used to demonstrate this capability. Results focus on quality control issues associated with AIRS, optimal assimilation strategies, and the impact of AIRS data on subsequent numerical forecasts.

*gary.jedlovec@nasa.gov; phone: 256 961-7966; fax: 256 961-7788; http://weather.msfc.nasa.gov/sport/

2. AIRS DATA

2.1 AIRS Specifications

Aboard the EOS polar-orbiting Aqua satellite with an early afternoon equator crossing time, AIRS coupled with the Advanced Microwave Sounding Unit (AMSU) form an integrated temperature and humidity sounding system. AIRS is a cross-track scanning infrared spectrometer/radiometer with 2378 spectral channels between 3.7 and 15.4 µm (650 and 2675 cm-1). Due to its hyperspectral nature, AIRS has the ability to provide near radiosonde-quality atmospheric temperature profiles with accuracy of 1 K in 1 km vertical layers and moisture profiles with accuracy of 20% RH in 2 km vertical layers. Because AIRS footprints coincide with AMSU footprints, the AIRS Science Team uses AMSU data in the retrieval process, producing a uniform distribution of AIRS retrievals in both clear and cloudy scenes at a spatial resolution of 50 km (Aumann et al. 2003). The superior vertical resolution and sounding accuracy make the instrument very appealing as a complementary data source to the global rawinsonde network.

For this study, we use prototype Version 5.0 supplemental retrieved soundings. Each sounding contains 101 levels between 1100 hPa and 0.1 hPa. Although, profiles from this updated version of the EOS science team algorithm have not yet been validated, we can safely assume that the relative validation errors will be similar to (or better than) those presented for the Version 4.0 data. The Version 4.0 temperature profiles have been validated by the science team over both land and ocean with RMS profile error ranging from 0.6 – 1.0 K (ocean ±50^{o} of the equator) to 0.9 – 1.3 K (global ocean and land) (Fetzer et al. 2005). The global ocean and land values will be used as part of the equation to approximate the AIRS errors in the data assimilation (see Section 3). Thus, the assignment of Version 4.0 error estimates for use with Version 5.0 prototype profiles should be conservative.

2.2 Quality Indicators in Version 5.0 AIRS Data

The Version 5.0 AIRS profiles come with quality indicators for each profile that allow the user to selectively use profiles (or parts of profiles) based on the accuracy requirements for their application. Intelligent interpretation of these quality indicators will enable use of only the most appropriate data, maximizing impact and benefit to the application. In this application, prudent use of AIRS profiles and quality indicators should yield improved 0- to 48-hour weather forecasts.

Because each retrieved profile is generated from the top of the atmosphere to the bottom, there is a specific level below which data is of questionable quality. This level is determined by careful analysis of the differences between error characteristics of each individual AIRS profile and a large-scale truth field. This difference is then compared to a predetermined threshold for retrieval failure. If the temperature value at three consecutive levels is larger than that threshold, then the uppermost failing level is deemed the maximum pressure level valid for that particular sounding with all levels below considered questionable data. The temperature and moisture thresholds are coupled meaning that the maximum pressure level for temperature is also the maximum pressure level for moisture. More information on how the errors are generated can be found in Susskind and Atlas (2006). See Section 4 for more information on how the quality indicators are being used to pre-process the data for assimilation into ADAS.

3. ANALYSIS SCHEME

The Advanced Regional Prediction System (ARPS; Xue et al. 2001) Data Assimilation System (ADAS; Brewster 1996) developed at the University of Oklahoma is used to blend the AIRS temperature and moisture information with other observations and a background field provided by WRF forecasts to produce improved initial conditions. The ADAS is designed to address regional-scale data assimilation problems in real-time numerical weather prediction. It provides a means to merge different sources of local meteorological data into a single, coherent three-dimensional description of the atmosphere. This particular analysis system was selected because it is flexible and easy to configure for the assimilation of satellite-derived atmospheric profiles. ADAS uses a Bratseth (1986) successive correction methodology, which converges to optimal interpolation without the need for complex matrix inversions that hinder more sophisticated variational assimilation techniques (Sashegyi 1993). Because the system does not use significant computation time to perform analyses, it makes an ideal candidate for assimilation of large data sets for near-real-time data assimilation on a regional scale with the resources available at the SPoRT Center.

3.1 Bratseth Successive Correction Method

The Bratseth method uses observations to iteratively update a background field provided by a large-scale model forecast for two equations: one an "updated" estimate of the grid point and the other an "updated" estimate of the observations themselves. Each observation type is weighted based on its error characteristics, data density, and proximity to analysis points. For an analysis variable, ϕ (i.e. pressure, temperature, moisture, or horizontal wind components), the grid point correction is given by

$$\phi_x(k+1) = \phi_x(k) + \sum_{i=1}^{n_{obs}} \alpha_{xi}[\phi_i^{obs} - \phi_i(k)] \ , \tag{1}$$

while the observation correction is given by

$$\phi_o(k+1) = \phi_o(k) + \sum_{i=1}^{n_{obs}} \alpha_{oi}[\phi_i^{obs} - \phi_i(k)] \, . \tag{2}$$

In the above equations, $\phi_x(k)$ and $\phi_o(k)$ are, respectively, the grid point and observation estimates, ϕ_i^{obs} is the value of the i^{th} observation, and n_{obs} is the total number of observations. The grid point weighting function, α_{xi}, in Eq. 3 is given by

$$\alpha_{xi} = \frac{\rho_{xi}}{m_i} \, , \tag{3}$$

and the observation weighting function, α_{oi}, in Eq. 4 is given by

$$\alpha_{oi} = \frac{(\rho_{oi} + \sigma^2 \delta_{oi})}{m_i} \, , \tag{4}$$

where ρ_{xi} and ρ_{oi} are the gridpoint-to-observation and observation-to-observation spatial correlation functions respectively, σ^2 is the ratio of the observation error variance to the background error variance, δ_{oi} is the Kronecker delta (i.e. 1 for $i = j$, 0 for $i \neq j$), and m_i is given by

$$m_i = \sigma^2 + \sum_{i=1}^{n_{obs}} \rho_{oi} \, . \tag{5}$$

The value of σ^2 is calculated using error covariance matrices for each data type assimilated into ADAS (in our case, these errors are only for the background field and AIRS profiles). How these errors are defined is discussed in more detail in the next section. The spatial correlation functions in Eqs. 4 and 5 are assumed to be Gaussian and are given by

$$\rho_{ij} = \exp\left(\frac{-|r_{ij}|^2}{R^2} \right) \exp\left(\frac{-|\Delta z_{ij}|^2}{R_z^2} \right) , \tag{6}$$

where r_{ij} and Δz_{ij} are, respectively, the horizontal and the vertical distances between the observation/observation pairs or the gridpoint/observation pairs, and R and R_z are, respectively, the horizontal and vertical scaling factors that define the maximum distance between observation/observation pairs and gridpoint observation pairs that will be included in the analysis. How R and R_z are selected for our setup is described in part c of this section. For further description of the Bratseth technique and how it is used in ADAS see Lazarus et al. (2002).

3.2 Determination of the error tables

As mentioned in the previous section, the error tables in ADAS are used to determine the error covariances for both the background field and each observation set. The error tables for each observational data set are an approximation of instrument error and representativeness error. Representativeness error is an estimate of inaccuracies introduced due to differences between the resolution of the observations and the grid resolution. In other words, both types of errors are approximations, so the error tables themselves are approximations. As a result, the number defined in the error table does not have a perfectly physical meaning. Rather, the true importance of the error tables is in the ratio between the background error and the observation errors (i.e. σ^2 in Eqs. 4 and 5).

Although we are using the WRF short-term (i.e. 0-1 h) forecasts as the backgrounds for our ADAS analyses, the error table used to define the background error covariance is based on the Rapid Update Cycle (RUC; Benjamin et al. 2002) 3-h forecast error. The reason we use these errors is because there is little difference between error in short-term forecasts from these regional model forecasts (Brewster personal communication), and the RUC error is well defined in the ADAS documentation. The error tables used for the AIRS profiles are the global estimates cited in the Version 4.0 validation (Fetzer et al. 2005) with some additional error added to compensate for error of representativeness. As stated in the introduction, the error estimates for the Version 4.0 data will likely be conservative for use with the prototype Version 5.0 data as improvements in the retrieval algorithm will likely lead to more accurate AIRS profiles. Figure 1 shows the error tables used for the observations and the backgrounds (again, remember that it is the ratio of the observation to background errors and not necessarily the error values themselves that determine the weights for the analysis).

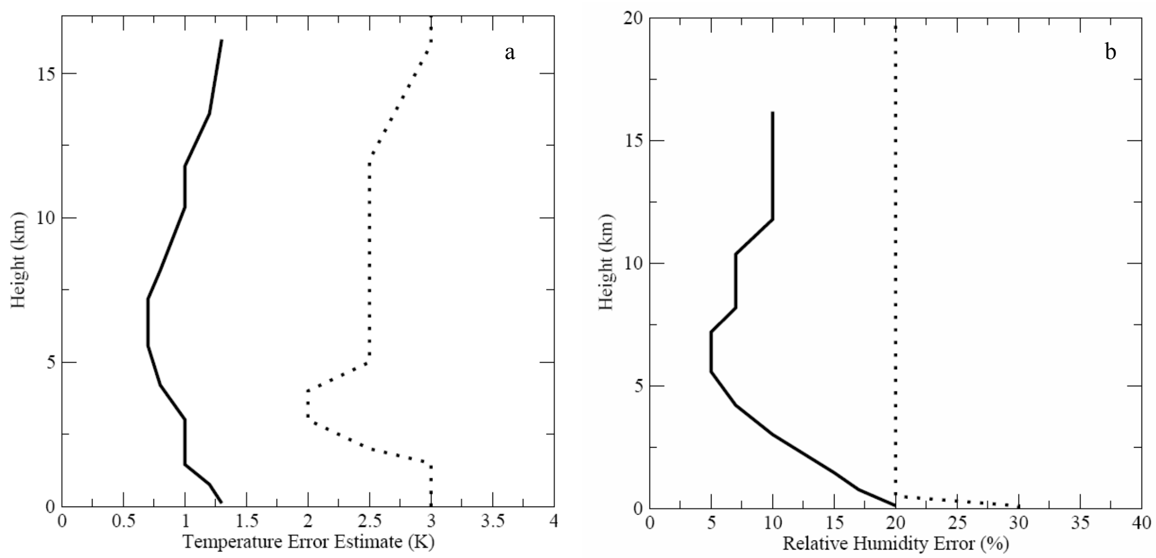

Figure 1. Error tables for (a) temperature and (b) relative humidity for assimilation of AIRS profiles into ADAS. The dotted line represents the estimated background field error. The solid line represents the AIRS estimated error. As stated in the text, the weighting in the analysis is based on the relative differences between the background and observation errors.

3.3 Determination of horizontal and vertical scaling factors

Reducing the horizontal scaling factor in subsequent iterations, draws greater detail with each successive pass (Lazarus et al. 2002). The horizontal and vertical scaling factors, R and R_z, respectively, were selected based on the horizontal and vertical resolution of the AIRS data. One wants to select scaling factors such that information from enough observations can be combined without the data becoming decorrelated over too large of a scaling distance. Two iterations of the Bratseth method are performed on the AIRS data. As a result, the two iterations are 150 km and 120 km respectively. This allows for approximately 9 AIRS profiles to be used in the final iteration for each grid point. Although the AIRS data represent a layer value due to the weighting functions and retrieval algorithm, the temperature and moisture profiles are being assimilated into ADAS in the same manner as radiosonde data. We treat the observations as point data in the three dimensional atmosphere. Although this is not the most effective method for using the data, it is the most straightforward. The effects of this are mitigated through the choice of a relatively large vertical scaling factor. For both iterations, this value is fixed at 750 m. This value was selected after testing various scaling factors from 100 m to 1000 m.

4. EXPERIMENTAL DESIGN

4.1 Case Study: 20-22 November 2005

A small weather disturbance over the northern Gulf of Mexico at 1200 UTC on 20 November 2005 played a significant role in the east coast weather for the week. Over a 48 hour period, this disturbance migrated across northern Florida and southern Georgia and emerged in the Atlantic Ocean where it quickly deepened into a significant storm off the coast of New Jersey and Long Island. The synoptic maps in Fig. 2 show 24-hour snapshots of the storm development from 1200 UTC on 20 November 2005 to 1200 UTC on 22 November 2005.

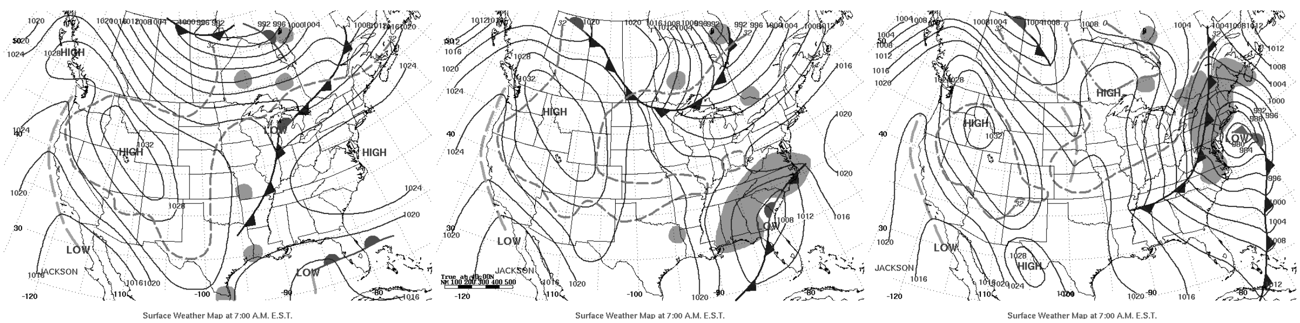

Figure 2. Surface weather maps showing the development/track of a low-pressure system from the Gulf of Mexico across northern Florida and southern Georgia and into the Atlantic Ocean south of Long Island after 48 hours. The left-most panel shows the 1200 UTC conditions on 20 November 2005, the center panel shows the 1200 UTC conditions on 21 November 2005, and the right-most panel shows the 1200 UTC conditions on 22 November 2005.

This particular storm system is of interest because of its proximity to the Southeastern United States and its rapid development in a relatively data void region. Another interesting feature about this storm is that most of its life is spent over the water, where assimilation of AIRS data should have a significant impact. The track and intensity of this system was not well represented in either 24- or 48-hr forecasts of the Global Forecast System (GFS) model, North American Model (NAM), or an operational version of the WRF run at the SPoRT Center. The poor forecasts of this synoptic system and subsequent distribution of precipitation along the east coast should provide an opportunity to assess improvements in the forecast when AIRS profiles are assimilated.

From the AIRS data perspective, because the storm is small and weak at the beginning of the time frame, there is plenty of clear to partly cloudy skies over the path of the storm whereby significant amounts of AIRS data can be assimilated. Also, there is a generous swath of AIRS data at 0700 UTC that covers much of the Atlantic. In addition, the system traverses land, and with good data coverage over land, it is possible to do sensitivity tests to see how these over-land data impact an analysis. Because the storm will impact most of the east coast at 48 hours, there is a plethora of data available for validation of model performance in this region.

As mentioned in section 2, the EOS Science Team assigns level-specific quality indicators to each retrieved profile. In order to show improvement in forecast quality, we are assimilating only over-ocean profiles because the quality of over-land profiles is still not certain even in the newest AIRS version. The profiles used are only over water with missing data values placed at all vertical levels of the AIRS profile below the maximum pressure level in pre-processing the data for ADAS. Figure 3 shows three-dimensional depictions of the AIRS profile data at 0700 UTC for 20 November 2005. The black points indicate the location of the highest quality AIRS profiles; the other shaded points indicate the maximum pressure level where AIRS data is valid.

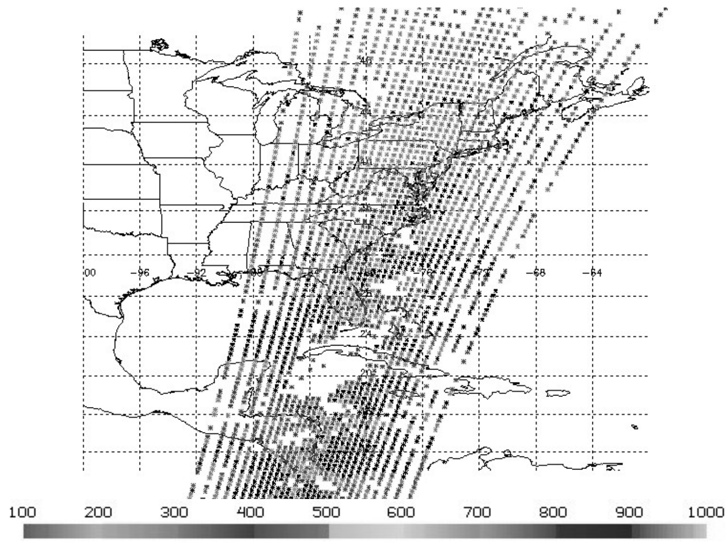

Figure 3. Three-dimensional depiction of the 20 November 2005 0700 UTC AIRS profile data assimilated into ADAS. The black points represent the highest quality data (i.e. entire sounding used) while the other shades indicate the height to which the quality data is available. Below the maximum pressure level, data is of lower quality, labeled missing, and not used in the analysis.

4.2 Forecast model configuration

The forecast model used herein is the WRF model, a next-generation mesoscale numerical weather prediction system designed to serve both operational forecasting and atmospheric research needs. It is a limited-area, non-hydrostatic primitive equation model with multiple options of physical parameterization schemes. The model domain used in this study consists of a 150 x 120 grid with 36 km spacing covering most of the contiguous United States, Eastern Atlantic Ocean, and Gulf of Mexico. It has 31 staggered terrain-following sigma levels with the top-level pressure at 100 hPa and finest resolution near the boundary layer. Table 1 summarizes the physical options used for this study

Table 1. WRF physical options

Microphysics	Ferrier (new Eta)
Longwave Radiation	RRTM
Shortwave Radiation	Dudhia
PBL Scheme	YSU
Convective Scheme	Kain-Fritsch
Soil Scheme	Noah land surface model

4.3 Analysis configuration

Special processing of the WRF and ADAS was required to link the assimilation system to the forecast model. The analysis is performed using the same horizontal domain as the WRF model. However, the ADAS has 43 sigma levels separated by an average of 500 meters with emphasis on more levels near the tope and bottom of the vertical domain. Because ADAS does not have the capability to directly define a vertical domain to match the WRF, vertical interpolation is necessary to adjust the background field prior to the analysis and then again to adjust the analysis domain to the model domain. There is likely some interpolation error introduced at these steps but it is difficult to separate the data assimilation interpolation errors from the analysis-to-model interpolation errors. The analysis is set up using the error tables and scaling factors described in Section 3.2 and 3.3.

4.4 Numerical Experiments

The first guess field and boundary conditions for the analysis are provided by the Global Forecast System (GFS), which is available every 6 hours. Currently, the operational version of the GFS assimilates radiance data from instruments aboard GOES and NOAA polar orbiting satellites. The infrared instruments (GOES sounders and HIRS) provide limited spatial coverage in cloud free regions, while the microwave instruments (AMSU) provide a more global coverage. This implies that the first-guess field used to initialize the WRF provides large-scale thermodynamic information that may be sporadically influenced by satellite data.

The SPoRT ADAS/WRF assimilation cycle begins at 0600 UTC on 20 November 2005 with a 1-hour WRF forecast initialized with the GFS analysis. This forecast is, then, used as the first-guess field for the ADAS analysis at 0700 UTC when the swath of AIRS data is valid. Two experiments are conducted. The control (CNTL) is a forecast that is run without using ADAS to assimilate any data. The second is a forecast run assimilating only AIRS data (AIRS). Although there may be some minimal impact due to the vertical interpolations described in Section 4.3, the difference between the AIRS and CNTL runs will represent the impact of the AIRS data on the forecast.

5. FORECAST IMPACT

The upper-air verification statistics for the 48-hour forecast (at 12 hour intervals coincident with rawinsonde launches) are computed by comparing the rawinsonde value to the model forecast values interpolated to the location of that specific observation. In this study, verifications are based on 17 rawinsonde station on the east coast of the United States (see Fig. 4), which is where the storm system in the case study is impacting and where AIRS data will have the largest impact on the 0 – 48 hour forecast.

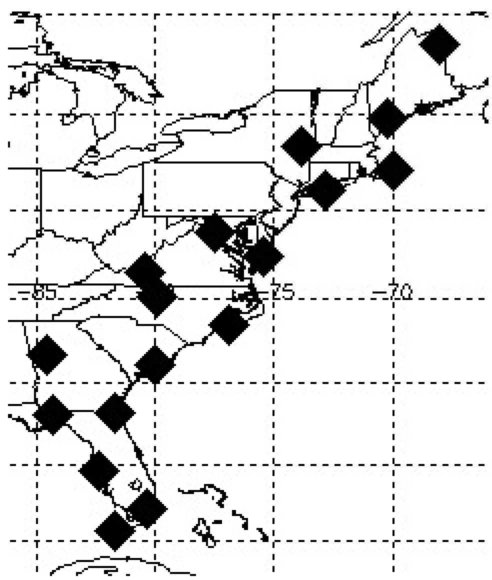

Figure 4. Rawinsonde locations for verification study.

Results for the 30h forecast are presented in Fig. 5. In Fig. 5a, the temperatures are taken as the forecast minus the rawinsonde, which means that positive biases indicate that the forecast is warmer than the observations, and negative biases indicate that the forecast is cooler than the observations. In Fig. 5b, the mixing ratio is presented as a relative value because upper atmosphere quantities are generally so small that differences are not necessarily evident in the actual mixing ratios. This relative value is obtained using

$$1/N*\sum(fcst-obs)\Big/1/N*\sum obs\ . \qquad (7)$$

Positive biases indicate that the forecast is moister than the observations; negative biases indicate that the forecast is drier than the observations.

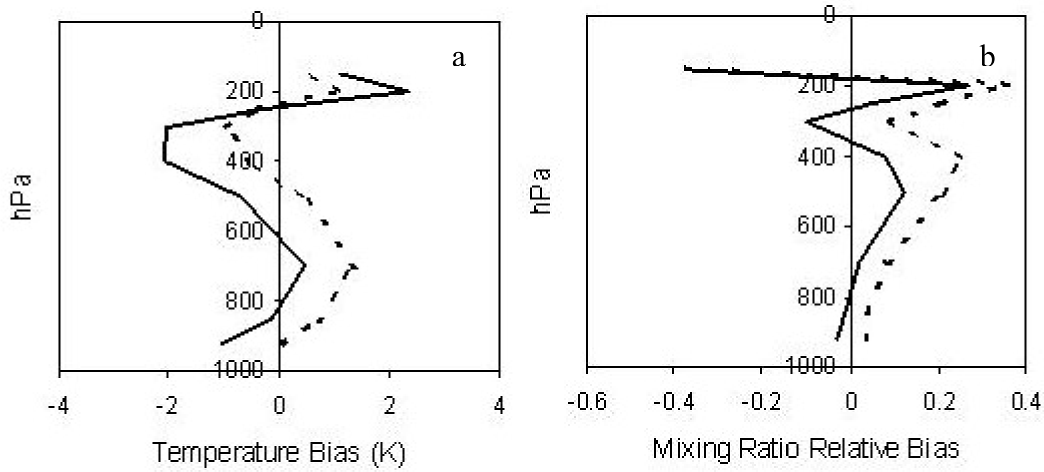

Figure 5. 30h forecast biases for a) temperature and b) relative mixing ratio ($\left[1/N*\sum(fcst-obs)\right]/\left[1/N\sum obs\right]$). The dashed line is the control (CNTL), in which no data is assimilated, and the solid line is the case where prototype Version 5.0 AIRS profiles have been assimilated (AIRS).

For most levels, the control (dotted line) performs well for both temperature (< 2K) and moisture (< 35%) for the 30h forecast. The CNTL shows a small amount of positive bias the lowest level for both temperature and moisture with this positive bias increasing throughout the middle troposphere. The temperature forecast becomes cooler above 400 hPa, but quickly becomes warmer again above 200 hPa. The moisture bias is consistently dry. Addition of the AIRS prototype Version 5.0 data leads to a general cooling and drying trend in most layers compared to the control. The control outperforms the AIRS at the lowest level and above 400 hPa for temperature, but the AIRS is less biased in the middle troposphere for temperature. For the mixing ratio, the drying of the atmosphere by the AIRS profiles produces a significantly better moisture forecast than the control at most levels—in fact, there is almost no bias at 800 hPa. In general, biases are less than 15% of the total level moisture when AIRS profiles are assimilated.

6. CONCLUSIONS

A procedure has been developed to incorporate AIRS Level-2 prototype Version 5.0 temperature and moisture sounding data into ADAS using an approach to maximize the utility of the available data. Preliminary experiments show that this technique, when applied to AIRS data over the east coast of the United States, can improve weather forecasts in regions impacted by data sparseness. Future work will examine the effects of using over-land data and how to deal with error characteristics from this different type of sounding. Also, there is data available at 1800 UTC on 20 November 2005 in the same region and experiments will be conducted to study how assimilation of further data impacts the forecasts. This pseudo-operational experiment will lead to further testing of AIRS assimilation in near-real-time for operational forecasting.

ACKNOWLEDGMENTS

This research was funded by the NASA Science Mission Directorate's Earth-Sun System Division in support of the SPoRT program at Marshall Space Flight Center.

REFERENCES

1. H. H. Aumann, M. T. Chahine, C. Gautier, M. D. Goldberg, E. Kalnay, L. M. McMillin, H. Revercomb, P. W. Rosenkranz, W. L. Smith, D. H. Staelin, L. L. Strow, and J. Susskind, "AIRS/AMSU/HSB on the Aqua Mission: Design, Science Objectives, Data Products, and Processing Systems". *IEEE Transactions on Geoscience and Remote Sensing*, 41, 253-264 (2003).

2. S. J. Goodman, W. M. Lapenta, G. J. Jedlovec, J. C. Dodge, and T. Bradshaw, "The NASA Short-term Prediction Research and Transition (SPoRT) Center: A collaborative model for accelerating research into operations". *Preprints, 20th Conference on Interactive Information Processing Systems (IIPS)*, Seattle, WA, Amer. Meteor. Soc., CD-ROM, P1.34 (2004).

3. Michalakes, J., S. Chen, J. Dudhia, L. Hart, J. Klemp, J. Middlecoff, and W. Skamarock, "Development of a next-generation regional weather research and forecast model". *Developments in Teracomputing: Proceedings of the Ninth ECMWF Workshop on the Use of High Performance Computing in Meteorology*, W. Zwiefnofer and N. Kreitz, Eds., World Scientific, 269–276 (2001).

4. M. Xue, K. K. Droegemeier, and V. Wong, "The Advanced Regional Prediction System (ARPS) - a multiscale nonhydrostatic atmospheric simulation and prediction tool. Part I: Model dynamics and verification". *Meteor. Atmos. Phys.*, 75, 161-193 (2001).

5. K. Brewster, K, "Implementation of a Bratseth analysis scheme including Doppler radar". *Preprints, 15th Conf. on Weather Analysis and Forecasting*. Amer. Meteor. Soc., Boston, MA, 596-598 (1996).

6. J. Susskind and R. M. Atlas, "Improved soundings and error estimates using AIRS/AMSU data". SPIE Conference Proceedings 6233 (2006).

7. A. M. Bratseth, "Statistical interpolation by means of successive corrections". *Tellus*, 38A, 439-447 (1986).

8. K. D. Sashegyi, "Application of the Bratseth scheme for the analysis of GALE data using a mesoscale model". *Mon. Wea. Rev.*, 121, 2331-2350 (1993).

9. S. M. Lazarus, C. M. Ciliberti, J. D. Horel, and K. A. Brewster, "Near-real-time applications of a mesoscale analysis system to complex terrain". *Wea. Forecasting*, 17, 971-1000 (2002).

10. S. G. Benjamin, G. A. Grell, J. M. Brown, T. G. Smirnova, and R. Bleck, "Mesoscale weather prediction with the RUC hybrid isentropic—terrain-following coordinate model". *Mon. Wea. Rev.*, 132, 473-494 (2004).

11. E. J. Fetzer, A. Eldering, E. F. Fishbein, T. Hearty, W.F. Irion, and B. Kahn, "Validation of AIRS/AMSU/HSB Core Products for Data Release Version 4.0", JPL, [Online]. Available: http://disc.gsfc.nasa.gov/AIRS/documentation/v4_docs/V4.0_Validation_Report.pdf (2005).