Open access • Journal Article • DOI:10.1029/2005JD006116

Validation of Atmospheric Infrared Sounder temperature and water vapor retrievals with matched radiosonde measurements and forecasts — Source link

Murty Divakarla, Christopher D. Barnet, Mitchell D. Goldberg, Larry M. McMillin ...+4 more authors

Institutions: National Oceanic and Atmospheric Administration

Published on: 16 May 2006 - Journal of Geophysical Research (John Wiley & Sons, Ltd)

Topics: Atmospheric Infrared Sounder and Radiosonde

Validation of Atmospheric Infrared Sounder temperature and water vapor retrievals with matched radiosonde measurements and forecasts

Murty G. Divakarla,[1] Chris D. Barnet,[2] Mitchell D. Goldberg,[2] Larry M. McMillin,[2] Eric Maddy,[3] Walter Wolf,[3] Lihang Zhou,[3] and Xingpin Liu[3]

Received 21 April 2005; revised 3 November 2005; accepted 23 November 2005; published 6 April 2006.

[1] An evaluation of the temperature and moisture profile retrievals from the Atmospheric Infrared Sounder (AIRS) data is performed using more than 2 years of collocated data sets. The Aqua-AIRS retrievals, global radiosonde (RAOB) measurements, forecast data from the National Center for Environmental Prediction Global Forecasting System (NCEP_GFS), the European Center for Medium Range Forecast (ECMWF), and the operational retrievals from the NOAA 16 satellite Advanced TIROS Operational Vertical Sounder (ATOVS) instrument are used in this validation. Using RAOB observations as the reference, bias and RMS differences are computed for "sea," "land," and "all" categories for the AIRS retrievals and other collocated data sets. The results of the intercomparison reveal that temperature and water vapor retrievals from the AIRS are in very good agreement with the RAOBs. The RMS difference for clear-only cases over "sea" and "all" categories is close to the expected goal accuracies, namely, 1°K in 1 km layers for the temperature and better than 15% in 2-km layers for the water vapor in the troposphere. The overall RMS difference for the cloud-cleared cases is also close to the expected product goal accuracy except for a slight degradation at the surface. When AIRS and ATOVS retrievals are compared with the RAOBs, the AIRS temperature retrievals show an improvement over ATOVS of at least 0.5°K for all the accepted cases. Both the ECMWF and the NCEP_GFS forecasts match the RAOB temperatures within 1°K and water vapor within 14%. With respect to biases, the AIRS final retrieval shows a larger bias with the RAOBs relative to ATOVS, NCEP_GFS, and ECMWF. The bias is highly influenced by a larger bias contribution from "land" samples and shows a month-to-month and annual variation that correlates with the CO_2 variations. This coupling suggests a need to include CO_2 and possibly other trace gas climatologies in the AIRS initial guess to partially mitigate the effects in the final physical retrieval.

Citation: Divakarla, M. G., C. D. Barnet, M. D. Goldberg, L. M. McMillin, E. Maddy, W. Wolf, L. Zhou, and X. Liu (2006), Validation of Atmospheric Infrared Sounder temperature and water vapor retrievals with matched radiosonde measurements and forecasts, *J. Geophys. Res.*, *111*, D09S15, doi:10.1029/2005JD006116.

1. Introduction

[2] Satellite measured infrared radiances and the retrieved products form the basic input to numerical weather prediction models. The initialization task requires the use of satellite radiances and/or retrievals that are global by nature and that are denser and more homogeneous than other data sources. However, the measured radiances do not yield the temperature and moisture profile retrievals directly. The infrared radiances need to be corrected for cloud contamination. These corrected radiances are then used in a process of mathematical inversion to retrieve atmospheric temperature and moisture profiles, profiles of ozone and other trace gases. The accuracy of the retrievals depends on the accuracy of the observations, prescribed atmospheric transmittance functions, algorithms used for cloud clearing, and inversion algorithms. An evaluation of retrieved temperature and water vapor profiles with reference profiles such as radiosondes (RAOBs) and other forecast models helps validate the retrieval algorithms and guides efforts to alleviate any discrepancies.

[3] In this paper, we have evaluated the temperature and moisture profiles retrieved from the Aqua-Atmospheric Infrared Sounder (AIRS) instrument with collocated RAOB measurements. In addition, collocated model forecasts from National Center for Environmental Prediction Global Forecasting System (NCEP_GFS, formerly known as Aviation Model Forecast) and the European Center for Medium Range Forecast (ECMWF), and the operational retrievals from NOAA 16 satellite Advanced TIROS Operational Vertical Sounder (ATOVS) are compared with the RAOBs

[1] STG Inc., Reston, Virginia, USA.
[2] Office of Research and Applications, NOAA/National Environmental Satellite, Data, and Information Service, Camp Springs, Maryland, USA.
[3] QSS Group Inc., Lanham, Maryland, USA.

0148-0227/06/2005JD006116$09.00

to study the performance of the AIRS instrument relative to other methods of observation. The collocated data set spans a period over 2 years (September 2002 to December 2004) and is approximately global in nature with approximately 82,000 collocations. The criterion for the collocations is within ±3 hours of time coincidence and a distance window of 100 km radius between the RAOB location and the center of the AIRS retrieval. The RAOB measurements are from the National Oceanic and Atmospheric Administration (NOAA)–National Environmental Satellite Data and Information Service (NESDIS) operational meteorological database archive, which is used routinely to monitor, validate, and tune NOAA 16 ATOVS products [*Reale*, 2002, 2003]. The NOAA/NESDIS processing system uses the quality-checked RAOB reports for the mandatory and significant levels and interpolates the data to predefined ATOVS pressure levels. The RAOB temperature data are at 42 pressure levels between 1030 hPa and 0.1 hPa, and the water vapor mixing ratios are at 19 pressure levels covering the troposphere from 1030 hPa to 200 hPa. Since NOAA 16 and Aqua are both afternoon satellites with close orbital times, we have used the RAOB and ATOVS collocations to extract time and space collocated Aqua-AIRS and Advanced Microwave Sounding Unit (AMSU) radiances. The nature of the sun-synchronous Aqua satellite with its ascending and descending orbits crossing the equator at 1330 (and 0130) LT coupled with collocation criteria of ±3 hours matches mostly the 0000 and 1200 UT RAOB global collocations, and predominantly selects samples over the Europe and from the West Coast of the United States. Thus the matchup data set used in the study, although global, has a skewed geographic distribution. A detailed discussion on the sampling of RAOBs and the implications is presented in the later portions of the paper. The number of collocations is spread evenly over the 2-year period with 82% of the samples over the Northern Hemisphere. About 8.5% of the collocations are from the tropics and the remaining samples (∼10%) are from the Southern Hemisphere.

[4] To obtain AIRS retrievals, an off-line AIRS retrieval and validation system that emulates the operational version 4.0 is used. Using global RAOBs as the reference and utilizing all the accepted samples by the version 4.0 retrieval and quality assurance (QA) flags, bias and RMS differences are computed for the AIRS physical retrievals. The global samples are separated into ''land'' and ''sea'' samples and statistics are computed to study the retrieval errors for these subcategories. Statistics are also generated for the AIRS fast regression step, the first guess for the AIRS physical retrieval, to measure the improvement in the retrieval accuracy from the fast regression to the final physical retrieval. In addition, comparison statistics with RAOBs for ATOVS retrievals and for the NCEP_GFS and ECMWF model forecasts/analysis are generated for intercomparison and relative performance assessment.

[5] One of the critical steps in the retrieval process is the cloud clearing. The cloud-clearing procedure uses observed radiance of a set of channels from adjacent 3 x 3 AIRS FOVs to specify a cloud-cleared radiance [*Chahine*, 1982; *Susskind et al.*, 2003]. The noise in the cloud-cleared radiance is a function of the spatial distribution and quantity of clouds and varies widely between retrievals. Thus the cloud-clearing process, and its accompanied noise amplification (the ratio of cloud-cleared random error to the error in a single AIRS FOV) plays a major role in the retrieval uncertainty. To study this effect, we have computed statistics for all the accepted retrievals (i.e., cloud-cleared retrievals) and for ''clear-only'' retrievals using the appropriate flags from the AIRS retrieval system.

[6] Another objective of this paper is to assess the improvement in the temperature and water vapor profile accuracy of AIRS to that of ATOVS. The NOAA-ATOVS retrieval statistics with RAOBs are used as a baseline to measure the performance of the AIRS instrument and its retrieval system.

[7] *Tobin et al.* [2006] have performed AIRS validation studies for the Atmospheric Radiation Measurement program's Clouds and Radiation Testbed (ARM/CART) sites using dedicated RAOB ascents for special campaign periods. In this study, we provide a comprehensive look at the AIRS retrievals using global and spatially subset statistics for the tropics (23°N–23°S), midlatitude (50°N–23°N; 50°S–23°S), and high-latitude (90°N–50°N; 90°S–50°S) regions using quality checked operational RAOB measurements. These comparisons enable evaluation of the ability of the AIRS retrieval algorithm to deal with relatively cold and dry atmospheres in the high latitudes, and contrasting warm and humid atmospheres in the tropics. The comparison statistics presented in this paper using operational RAOBs for the tropical sea, and midlatitude land cases show considerable agreement with the results presented by *Tobin et al.* [2006] for the Tropical Western Pacific (TWP) and Southern Great Plains (SGP) sites with dedicated RAOB launches.

[8] An overview of the AIRS and ATOVS instrumentation and their retrieval algorithms is presented in section 2. The procedures used by the matchup system to generate validation data (RAOBs, AIRS and ATOVS retrievals, NCEP_GFS and ECMWF forecasts), and the characteristics of the validation data used in this study are briefly described in section 3. Description of the RAOB data set, use of version 4.0 quality flags to screen the retrievals, and the criteria used in the generation of statistics for various cases are described in section 4. Section 5 discusses in detail the results on the validation of the AIRS retrievals using the RAOB and the forecast profiles.

[9] The matchup system used to produce collocated data sets is in continuous operation at NOAA/NESDIS. A website to display validation statistics on a routine basis is located at http://www.orbit.nesdis.noaa.gov/smcd/spb/airs/index.html. Plans are on the way to study the utility of the data set in tuning the radiances, in the generation of empirical regression coefficients based on RAOB measurements, and for many other applications with future sounding instruments like the Infrared Atmospheric Sounding Interferometer (IASI) [*Diebel et al.*, 1996] and the Cross Track Infrared Spectrometer (CrIS) [*Glumb et al.*, 2003].

2. Overview of Retrievals

[10] This section presents an overview of the AIRS/AMSU and the ATOVS instruments and their retrieval systems. In this study, the ATOVS retrievals are used as a baseline standard to measure improvements with the AIRS/AMSU.

2.1. Aqua-AIRS and the Retrieval System

[11] The AIRS is the first in a new generation of high–spectral resolution infrared sounder instruments flown aboard Aqua research mission. The instrument is a cooled grating spectrometer that provides 2378 channels covering the IR spectrum from 650 to 2675 cm^{-1} (3.74–4.61 μm, 6.20–8.22 μm, and 8.8–15.4 μm infrared wavelengths) at a nominal spectral resolving power ($\lambda/\Delta\lambda$) of 1200. The AIRS is accompanied by two microwave sounding radiometers, the Advanced Microwave Sounding Unit–A (AMSU-A) and the Humidity Sounder for Brazil (HSB). AMSU-A is a 15-channel temperature sounder utilizing the 55 GHz Oxygen absorption band. HSB is mainly humidity sounder with channels centered on the water vapor line at 183.31 GHz. The HSB instrument failed in February 2003, and to be consistent, data from HSB are not used in the present study. Details of AIRS, AMSU and HSB instruments and their performance are given by *Aumann et al.* [2003]. The AIRS (and HSB) has a field-of-view (FOV) of 1.1° and provides a nominal spatial resolution of 13.5 km for IR channels and approximately 2.3 km for Vis/NIR channels. The AMSU-A, with its FOV of 3.3° has a spatial resolution of 50 km at the nadir and aligns with the 3 x 3 array of the AIRS FOVs. The geometry is frequently referred to as an AIRS golf ball. The golf ball data are first corrected for the local angle effects [*Zhou et al.*, 2005] and then used to derive retrieval products. A suite of AIRS product retrieval software (APS) has been developed to process AIRS data to derive many geophysical parameters [*Aumann et al.*, 2003]. The retrieval products include temperature and moisture profiles, IR and microwave surface emissivities as a function of frequency, total ozone, and cloud parameters [*Susskind et al.*, 2003].

[12] The APS suite has been put into operation at National Aeronautics and Space Administration (NASA) Jet Propulsion Laboratory, NOAA/NESDIS Office of Research and Applications, and at Goddard Space Flight Center (GSFC) Distributed Active Archive Center (DAAC) for dissemination of AIRS data products to user communities. The APS suite includes calibration, microwave retrieval, cloud clearing, initial IR retrieval, and a final physical retrieval. Numerous publications on the details of the modules, algorithms, and the EOS Aqua mission are available in the special issue published by *IEEE Transactions* (Institute of Electrical and Electronics Engineers, Special issue on the EOS Aqua Mission, *IEEE Transactions on Geoscience and Remote Sensing, 41*(2), 171–493, 2003). To summarize, the calibration procedure [*Pagano et al.*, 2003] converts Level-1a data to Level-1b calibrated radiances/brightness temperatures for AIRS and AMSU. These calibrated AMSU brightness temperatures are used by the microwave retrieval module to derive temperature and moisture profiles, cloud liquid water flags and microwave surface emissivity uncertainty [*Rosenkranz*, 2003]. The retrieved atmospheric state from AMSU is used to compute clear-column AIRS radiances for each of the AIRS FOVs that align with the AMSU FOV. With the assumption that the AIRS FOVs differ only in the cloud amount, the cloud-clearing procedure discards the data that violate the assumptions and produces cloud-cleared IR radiance product [*Susskind et al.*, 2003]. The initial IR fast-regression retrieval module uses the cloud-cleared radiance product to retrieve first guess profiles of temperature, moisture, ozone, and other surface parameters [*Goldberg et al.*, 2003]. The fast regression methodology is based on principal component analysis and utilizes 85 AIRS principal component scores as predictors with ECMWF forecast/analysis data as the predictant to derive physical regression coefficients. The coefficients are derived from a training data set consisting of 3 days of cloud-cleared radiances and ECMWF forecast/analysis of temperature, moisture and ozone. To qualify the data set for the regression procedure, the ECMWF data are screened to satisfy the requirement of a 2°K agreement between the AIRS measured and those simulated from ECMWF atmospheric state for 12 channels (702.7, 706.7, 711, 712.7, 715.9, 724.8, 746.0, 759.57, 965.4, 1468.83, 1542.35 and 1547.88 cm^{-1}) sensitive to different layers of the atmosphere. The requirement of radiance agreement for these channels helps to screen out spurious ECMWF atmospheric state samples that do not agree with the observed state. Details of the fast-regression and the algorithm updates are discussed by *Goldberg et al.* [2003, 2004]. The first guess retrievals are used as an initial solution by the final IR retrieval module. The procedure uses an iterative physical retrieval algorithm [*Susskind et al.*, 2003] to produce the final retrievals. Studies on the retrieval accuracies with the simulated data have revealed that temperature soundings can be produced under partial cloud cover with RMS differences of the order 1°K in 1 km layers for the troposphere and lower stratosphere [*Susskind et al.*, 2003]. For moisture, the retrieval accuracy is assessed to be better than 15% in 2-km layers in the troposphere. The AIRS validation team has also performed many validation studies with dedicated RAOB measurements for special campaign periods and sites [*Fetzer et al.*, 2003; *Tobin et al.*, 2006].

[13] The current version (version 4.0) of the AIRS APS suite is implemented at NOAA/NESDIS to produce and disseminate retrievals and special data products to many NWP centers [*Goldberg et al.*, 2003, 2004]. The AIRS/AMSU Level-1B radiance files collocated in time and space with the RAOB measurements are obtained from the near-real time AIRS processing system operated at NOAA/NESDIS. To produce AIRS retrievals for the collocated data set, an off-line research version (C. D. Barnet, ftp://ftp.orbit.nesdis.noaa.gov/pub/smcd/spb/cbarnet/reference/airsb_code.pdf) emulating version 4.0 algorithm and its accompanied statistics routines that account for the version 4.0 quality assurance (QA) flags (J. Susskind et al., Accuracy of geophysical parameters derived from AIRS/AMSU as a function of fractional cloud cover, submitted to *Journal of Geophysical Research*, 2005, hereinafter referred to as Susskind et al., submitted manuscript, 2005) is used. The QA flags and their application to the data screening are discussed in the later portions of the paper. The research version can emulate various versions of the AIRS retrieval algorithm, and comes with an array of visualization tools to assess the quality of the AIRS Level-1B and Level-2 retrieval products at various retrieval stages. The system has been evolved from the prelaunch simulation experiments and is used by many science team members including those at NOAA and at NASA-GSFC. The off-line code emulates v4.0 in all aspects, including L1b processing, QA assessment, and all computations. A significant effort has

been made to maintain the off-line code so that it produces the same answer as the AIRS APS. Thus, in terms of the retrieval results, the research version is expected to be very close to the operational AIRS version 4.0 Product Generation Executives (PGE). The HSB instrument failed in February 2003, and to be consistent, data from HSB are excluded in the retrieval process for the whole period to produce AIRS/AMSU retrievals. For simplicity, throughout the paper, the term ''AIRS retrievals'' is used to personify ''AIRS/AMSU retrievals.''

2.2. ATOVS and the Retrieval System

[14] The ATOVS system flown aboard many NOAA series satellites (NOAA 15, NOAA 16, NOAA 17, etc.) has two sounding instruments, namely, the High-Resolution Infrared Sounder (HIRS) and the AMSU-A and AMSU-B instruments. The characteristics of AMSU-A and AMSU-B aboard NOAA 16 are similar to that of AMSU and HSB instruments on board Aqua. The HIRS is a 20-channel filter-wheel radiometer with channels covering the IR bands from 4.3 μm to 15 μm with a spectral resolution of ($\lambda/\Delta\lambda$) of 75. While the HIRS is a 20 channel radiometer with coarse vertical resolution (3–5 km in the troposphere), the AIRS provides high-resolution spectral channels with improved vertical sampling. The nadir spatial resolution of HIRS channels is around 17 km, comparable to that of AIRS. Both AIRS and HIRS are cross-track scanning instruments with onboard radiometric calibration targets. The AIRS provides 90 ground footprints per scan, and the HIRS provides 56 spots per scan with a scan mirror viewing ±49.5°. The expected accuracy goals are higher for the AIRS (1°K in 1 km layers for the temperature and better than 15% in 2-km layers for the water vapor) to that of ATOVS (2°K in 1-km layers for the temperature and 25% in 2-km layers for the water vapor).

[15] The NOAA sounding products team operationally retrieves and archives ATOVS temperature and water vapor profiles, total ozone, and other surface parameters. The ATOVS retrievals and the collocated RAOB measurements are stored in the matchup database. The algorithms used for the ATOVS retrievals are described in detail by a series of publications [*Reale*, 2002, 2003]. In short, the ATOVS retrieval system uses the HIRS measurements and AMSU measurements (interpolated to the HIRS FOV) to retrieve various geophysical parameters. The HIRS and AMSU measurements are first limb adjusted to nadir view [*Allegrino et al.*, 1999; *Goldberg et al.*, 1999], and flagged for identification of effects due to precipitation for the AMSU measurements, and clouds for the HIRS measurements [*Ferguson and Reale*, 2000]. The retrieval methodology defines a first guess retrieval on the basis of a library search/eigenvector technique [*Reale*, 2003] using recent collocations from a matchup database. The final retrieval utilizes a minimum variance simultaneous solution [*Fleming et al.*, 1988; *Reale*, 2003].

3. Validation Data

[16] Daily Level-1B AIRS/AMSU radiance files generated by the AIRS near-real time processing system are used in conjunction with RAOB-ATOVS matchup files to initiate the generation of matched data sets. The RAOB-ATOVS matchup file contains RAOB measurements and ATOVS sounding retrievals collocated in time and space.

[17] The RAOB measurements include surface measurements at the RAOB station locations, temperature, and water vapor values at mandatory and significant levels. The measurements are quality checked for climatological limits [*Tilley et al.*, 2000], and are extrapolated beyond 50 hPa with AMSU temperature retrievals [*Goldberg*, 1999]. The measurements are then interpolated to predefined 42 ATOVS pressure levels to provide temperature profile covering 1000 hPa to 0.1 hPa and 19 levels of water vapor mixing ratios in the troposphere. The ATOVS retrievals provide 42 levels of atmospheric temperatures spanning 1000 hPa to 0.1 hPa and 15 levels of water vapor covering 1000 hPa to 300 hPa. A detailed description on the RAOB-ATOVS meteorological database is located at http://www.orbit.nesdis.noaa.gov/smcd/opdb/poes/atovs/docs/icd.html. Both the RAOB measurements and the ATOVS retrievals are interpolated to standard 100 level AIRS pressure levels for comparison with the AIRS retrievals. The RAOB location and time information from the matchup file is used to extract collocated golf ball data from Aqua-AIRS/AMSU instruments. The AIRS/AMSU center FOV time and location information is used to extract NCEP_GFS and ECMWF forecast profiles.

[18] The NCEP_GFS forecast/analysis data used in this study have a Gaussian grid resolution of 1° latitude × 1° longitude with temperature and relative humidity values at 26 mandatory levels from 1000 to 10 hPa. The NCEP_GFS is run 4 times a day, at analysis times equal to 0000 UT, 0600 UT, 1200 UT, and 1800 UT. For each analysis time there are 4 forecast times, F0000, F0300, F0600, and F0900 or 16 files per day in GRIdded Binary (GRIB) format. On the basis of the AIRS center FOV time information, the two files that encompasses the FOV observation time are selected. The data from these two files are interpolated in time and space to AIRS center FOV and the profile data are extrapolated with the Upper Atmospheric Research Satellite (UARS) upper atmosphere standard climatology [*Barnett and Corney*, 1985] above 10 hPa to be consistent with the standard AIRS Level-2 100 level retrievals.

[19] The ECMWF analysis data used in this study have a Gaussian grid resolution of 0.5° latitude × 0.5° longitude with temperature and mixing ratio values at 60 sigma levels. The model is run at analysis times 0000 UT, 0600 UT, 1200 UT, and 1800 UT with a 0, 3, 6, and 9 hour forecast. On the basis of the AIRS center FOV time information, only one GRIB analysis file (0000, 0600, 1200, and 1800 UT) is chosen, and profile data values closest in location and time to the AIRS center FOV are extracted and the 60 level data are interpolated to the AIRS Level-2 100 level pressures. The ECMWF data used for comparisons are only for 7 months (January to March, and July to October of 2003) because of some processing limitations.

4. Data Set Preparation and Analysis

[20] This section presents a description of the collocated data, PGE-V4.0 QA flags, their application to the data screening, and the statistical metrics used in the evaluation of AIRS retrievals.

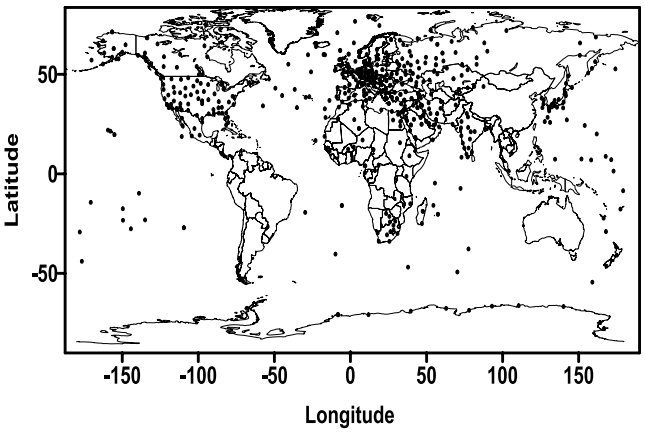

Figure 1. Locations of RAOB measurements collocated with AIRS observations. The Aqua local overpass results in a skewed distribution of collocations within 3 hours of the (predominantly 0000 and 1200 UT) RAOBs mainly in the vicinities of western Europe and the U.S. West Coast. Collocations occurring in the vicinity of the U.S. east Coast and China are with the smaller proportion of RAOBs launched at 0600 and 1800 UT.

4.1. Description of Collocated Data

[21] A data set consisting of (1) Aqua-AIRS and NOAA 16 ATOVS retrievals, (2) RAOB measurements, and (3) NCEP_GFS and ECMWF forecasts has been created for the period September 2002 to December 2004. Approximately 100,000 matchups are considered. The matchups are screened to ensure the availability of all the collocated data sets (RAOBs, AIRS, ATOVS, NCEP_GFS), and are confined to a time and distance collocation of ±3 hours and 100 km radius of the RAOB measurements. On the basis of the prescriptions provided by the NOAA sounding products team, the collocations are further screened to include 24 RAOB instrument types. This screening has provided about 82,000 collocated samples. Collocations that include ECMWF data have a sample size of 20,000 and are limited to 7 months.

[22] Figure 1 shows the RAOB locations of 538 stations that have contributed 82,246 samples over the 2-year period. The Aqua satellite crosses the equator at 1330 LT in its ascending (north) orbit and at 0130 LT while descending (south) with about 14–15 orbits a day. On the east coast, these local times correspond to approximately 1800 to 1900 UT and 0600–0700 UT, respectively. Given a collocation time window of 3 hours, and given that about 75–80% of the RAOBs are launched at 0000 UT and 1200 UT and the remainder at 0600 and 1800 UT, Aqua satellite collocations with RAOBs that adhere to the time window occur predominantly in the vicinity of western Europe and the U.S. West Coast, with relatively few in the vicinity of the U.S. east coast. However, because a small portion of the sondes have report times other than those listed, some (about 5%) collocations do occur in the eastern United States. The validity of these reports are no different, in fact, perhaps more valuable since they provide samples from a region that is otherwise unrepresented. In addition, the RAOBs typically do not report time in minutes, so a 1200 UT report time could have been launched anywhere from 1131 to 1229 UT. In summary, although there are many dots even on the eastern part of the United States (and in Japan and eastern Asia), the number of samples from these locations is significantly less than the samples from western Europe and the U.S. West Coast. The overall sampling of the RAOB matches with the Aqua orbit thus provide a skewed distribution with a very large number of samples chosen from Europe and West Coast of the United States at 0000 and 1200 UT. Thus the results presented, although global, are more representative of the highly sampled geographic regions and the type of sensors used in those RAOBs to measure temperature and moisture. Some of the areas of the planet with large numbers of RAOBs do not contribute to the results presented here. In this respect it would be optimal for more uniform global distribution [*Reale*, 2005], but is not possible at this time without extending the time window, which of course, would contribute additional errors.

[23] Figure 2 shows the percentage of samples covered by each RAOB instrument type, and the number of RAOB stations that use the instrument type. Although there are 24 instrument types in the RAOB data set, about 70% of the samples use five instrument types namely the RS80/Digicora (Fin), RS90/Digicora (Fin), VaiRS80-57H, AVK-MRZ (Russian) and VIZ-B2 (USA). Figure 3 shows the distribution of RAOB samples over different latitude zones. About 82% of the samples are from the Northern Hemisphere (90°N–23°N), 8% cover the tropics (23°N–23°S) and the remaining (∼10%) is from the Southern Hemisphere (23°S–90°S). Table 1 shows the global distribution (90°N–90°S latitude, 180°W–180°E longitude) of samples over day, night, land, sea, and coastal categories. About 29% of the samples are completely over land (day: 14.7%, night: 14.2%), 9% of samples are entirely over sea (day: 5.6%, night: 3.7%), and about 62% (day: 31.6%, night: 30%) are in coastal areas, a combination of land and sea with proportion of land less than 20%. The category "All" (Table 1) includes all the samples from land, sea, and coastal areas for day and night. The overall proportion of day and night samples is more or less equal (∼50% each). Also shown in the Table 1 are the samples that fall into tropics (23°N–23°S), midlatitude (50°N–23°N; 50°S–23°S), and high-latitude (90°N–50°N; 90°S–50°S) regions and their categories into land, sea, day and night cases. A total of 44% are over midlatitudes, and 48% pertain to high latitudes and the remaining 8% are from the tropics. The AIRS/AMSU golf ball data collocated with each RAOB sample and the surface pressure value extracted from NCEP_GFS are used in the AIRS off-line retrieval package. The retrievals are generated using an emulation of AIRS version 4.0 algorithms.

4.2. Implementation of PGE-V4.0 Quality Flags

[24] A set of QA flags have been suggested by the AIRS science team members to use as a standard in the validation of AIRS retrievals. Details of the QA flags are described by Susskind et al. (submitted manuscript, 2005). The quality flags that are relevant for temperature profile are (1) Qual_Temp_Profile_Top, (2) Qual_Temp_Profile_Mid, (3) Qual_Temp_Profile_Bot, and (4) Qual_Surf. These four quality flags assure the quality of temperature values at various pressure levels pertaining to stratosphere and upper troposphere, middle troposphere, lower troposphere, and

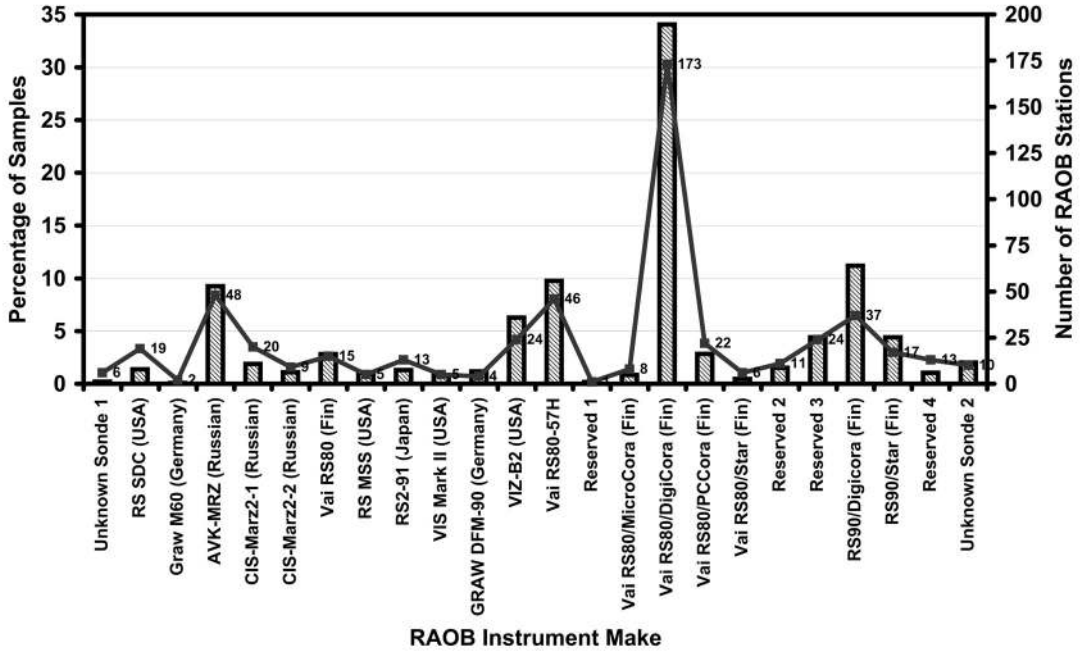

Figure 2. Percentage of RAOB samples covered by each instrument type (bars) and number of RAOB stations (solid line) that use the instrument type.

surface, respectively. Thus the number of samples accepted at each pressure level for the temperature retrievals vary depending on the acceptance criteria and the values assigned to the four quality flags. In this analysis, both the temperature and water vapor retrievals are screened for highest quality flag (final product quality flag = 0) and with these four quality flags. Table 1 shows the impact of the temperature quality flags on the global sample, and on the

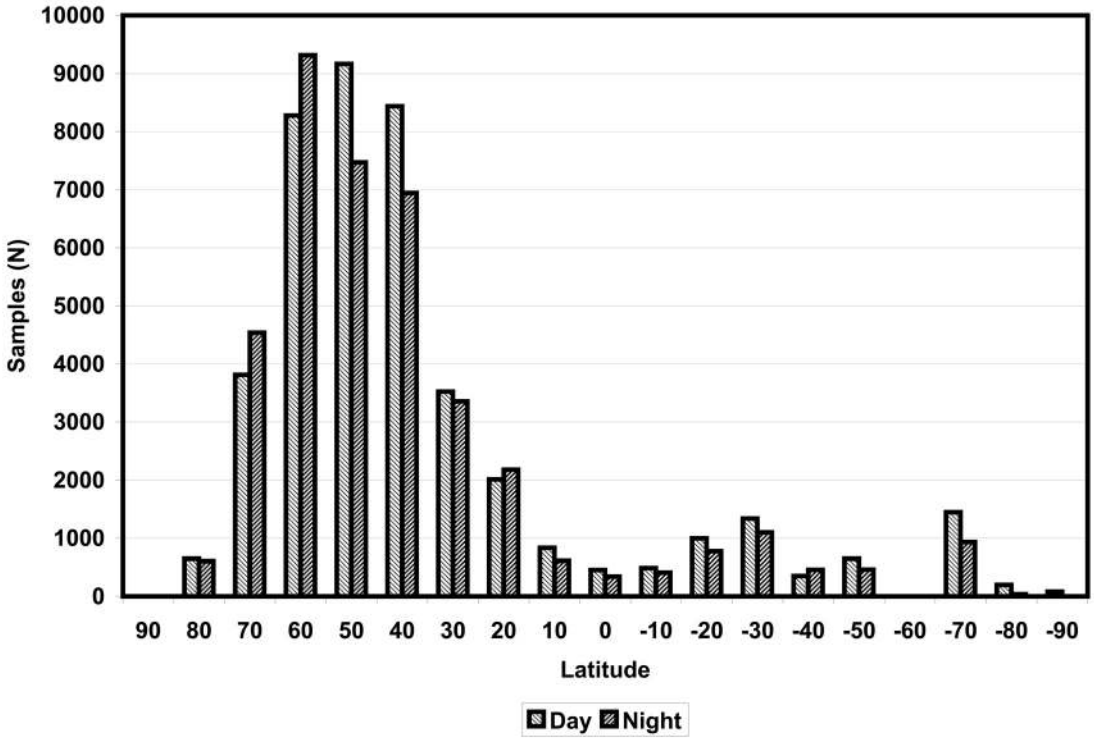

Figure 3. Histogram of the latitudinal distribution of RAOB day/night samples.

Table 1. Global Distribution of Collocated Samples and Percentage of Samples Accepted by the AIRS Retrieval Version 4.0 Emulation

	Land		Sea		Coast		All	
Quality Assurance Flags	Day	Night	Day	Night	Day	Night	Day	Night
Global (90°S–90°N, 180°W–180°E)								
Number of Samples	12102	11665	4628	3083	25983	24785	42713	39533
Samples, %	14.7	14.2	5.6	3.7	31.6	30.1	51.9	48
Qual_Temp_TOP, %	72	78	68	72	71	72	71	74
Qual_Temp_MID, %	54	62	35	40	55	55	53	56
Qual_Temp_BOT, %	48	49	35	40	54	52	50	50
Qual_Temp_SRF, %	15	16	27	32	19	20	18	20
High-Latitude (90°N–50°N, 90°S–50°S, 180°W–180°E)								
Number of Samples	3895	4127	1977	1258	13928	14138	19800	19523
Samples, %	4.7	5.0	2.4	1.5	16.9	17.2	24.0	23.7
Qual_Temp_TOP, %	72	71	66	72	74	72	73	72
Qual_Temp_MID, %	55	51	32	31	57	50	54	49
Qual_Temp_BOT, %	53	49	32	31	56	49	53	48
Qual_Temp_SRF, %	19	20	22	23	22	20	21	20
Midlatitude (50°N–23°N, 50°S–23°S, 180°W–180°E)								
Number of Samples	7463	6777	1647	1135	9944	8881	19054	16793
Samples, %	9.1	8.2	2.0	1.4	12.1	10.8	23.2	20.4
Qual_Temp_TOP, %	72	81	65	64	67	71	69	75
Qual_Temp_MID, %	55	68	35	33	52	60	52	62
Qual_Temp_BOT, %	48	50	35	33	50	54	48	51
Qual_Temp_SRF, %	13	14	28	27	14	16	15	16
Tropics (23°S–23°N, 180°W–180°E)								
Number of Samples	744	761	1004	690	2111	1766	3859	3217
Samples, %	0.9	0.9	1.2	0.8	2.6	2.1	4.6	3.9
Qual_Temp_TOP, %	65	87	74	84	68	76	69	80
Qual_Temp_MID, %	43	71	41	69	55	68	49	69
Qual_Temp_BOT, %	23	31	41	69	54	66	45	59
Qual_Temp_SRF, %	4	7	34	57	29	35	35	32

samples from the tropics, midlatitude and high-latitude regions, respectively. All the AIRS observations (N = 82,236) selected after initial screening based on distance and time collocation (±3 hours and 100 km radius) are processed through the retrieval system to generate AIRS fast regression and final physical retrievals. A quick look into Table 1 clearly indicates that the surface quality flag Qual_Surf screens out more land samples than sea and coastal regions. (This is due to the uncertainty of the land surface emissivity.) Also seen in Table 1 is the tighter control exerted by the Qual_Temp_Profile_Mid and Qual_Temp_Profile_Bot flags on the sea samples compared to land and coastal samples. The number of samples accepted by the quality flags for the nighttime is slightly higher than the daytime samples. The number of samples accepted by the Qual_Surf and Qual_Temp_Profile_Bot flags progressively increase from the tropics to midlatitude to high-latitude regions over the land areas. In contrast, the sample count progressively decreases over sea cases. In the full sample, these trends tend to compensate one another. When the RAOB database is screened with a tighter control of distance and time collocation (±1 hour and 50 km radius), the number of input samples to the retrieval system shrinks to 33,000. As expected, the percentage of accepted samples does not change appreciably.

4.3. Statistical Metrics

[25] The AIRS retrieved temperature and water vapor profile statistics are computed with reference to RAOB profiles (mean difference of AIRS-RAOB, and RMS difference) for three categories, namely, ''all'' that includes all the global samples (land/sea/coast), ''land'' and ''sea.'' Thus the first set of statistics is for all the samples that are ''cloud-cleared'' and accepted by the retrieval system for the three categories. In a similar fashion, statistics are also computed for the collocated ECMWF, NCEP_GFS forecasts, and ATOVS retrievals using RAOB as the reference and conforming to the accepted number of samples dictated by the retrieval QA flags for the three categories. Statistics with the ECMWF data comprise a sample size of 20,000 collocations in contrast to the global sample size of 82,000 for all the other collocations.

[26] To evaluate the ''cloud-cleared'' retrievals with that of ''clear'' cases, statistics are computed for ''clear-only'' cases using the ''clear'' flag that utilizes the clear test algorithm implemented in the retrieval system. The clear test uses a combination of spatial scene uniformity test, a microwave versus infrared agreement test and a shortwave IR lapse rate test [*Susskind et al.*, 2003]. Since, the number of samples accepted and designated as ''clear'' by the retrieval system are quite low (2–3%), statistics are computed for ''clear-only'' cases for the global sample with ±3 hours and 100 km radius collocation option.

[27] To assess the retrieval accuracies over the tropics, midlatitudes, and high latitudes, subsets of AIRS retrievals with matched-up collocations are stratified by zones of latitude. To study the differences in the retrieval accuracy due to proximity to the RAOBs, the time-space window is reduced from ±3 hours and 100 km to ±1 hour and 50 km radius. Finally, for the larger window sample, retrieval statistics for day and night cases are computed.

[28] The computation of statistics for all the cases is consistent with the conventions used by *Susskind et al.* [2003]. Temperature statistics are derived for 1 km layers

Table 2. RAOB Temperature Variability (Mean and Standard Deviation, STD) Over Global "Land," "Sea," and "All" Categories and Corresponding Number of Samples Accepted by the Version 4.0 Emulation

Pressure, Upper Boundary, hPa	Pressure, Lower Boundary, hPa	Land			Sea			All (Land/Sea/Coast)		
		N	Mean, °K	STD, °K	N	Mean, °K	STD, °K	N	Mean, °K	STD, °K
103	126	17799	212.5	7.6	5330	210.6	10.6	59433	213.7	8.8
126	142	17799	213.9	6.7	5330	212.7	9.0	59433	215.1	7.7
142	160	17799	215.1	6.0	5330	214.5	7.6	59433	216.3	6.8
160	190	17799	216.3	5.3	5330	216.9	5.8	59433	217.5	5.9
190	223	17799	218.2	5.6	5330	219.8	5.2	59433	219.1	5.7
223	273	13837	222.5	6.0	2854	224.9	6.1	44569	222.7	6.0
273	314	13837	228.8	7.1	2854	231.1	8.3	44569	228.2	7.4
314	344	13837	234.2	7.7	2854	236.3	9.4	44569	233.2	8.4
344	407	13837	240.8	8.2	2854	242.7	10.1	44569	239.5	9.1
407	478	13837	249.3	8.4	2854	251.0	10.4	44569	247.9	9.5
478	535	13815	256.3	8.4	2854	257.8	10.3	44546	254.8	9.6
535	618	13704	262.6	8.6	2854	264.0	10.3	44403	261.0	9.7
618	684	11472	267.7	8.7	2854	269.7	10.3	41145	266.4	9.9
684	778	6418	273.9	8.6	2586	276.1	9.4	26649	273.7	9.6
778	879	3627	276.2	9.3	2255	280.2	9.5	15742	276.0	10.4
879	1100	1830	282.0	9.3	2162	286.6	9.4	11949	282.5	10.9

for 1000 hPa to 0.01 hPa. Although statistics are computed for all the layers up to 0.01 hPa, the analysis presented here is restricted to troposphere (1000–100 hPa). For water vapor, statistics are computed for column densities converted to integrated column water in 2-km layers from the surface to 100 hPa. The percent error for each 2-km layer is computed by weighting the RMS difference with the reference (RAOB) water vapor amount in the layer. The accuracy of RAOB water vapor measurements depends on the sensor type and above 500–300 hPa, the measurement errors increase because of slower response of the sensor to the ambient humidity [*Miloshevich et al.*, 2004]. In addition, there is very little water vapor above 200 hPa. Hence statistics for water vapor are restricted to 2 km layers from surface to 200 hPa. Although temperature statistics are shown in the figures for 1 km layers, it may be noted that the expected accuracy above 300 hPa is 1°K in 3 km layers.

5. Results and Discussion

[29] Results of RMS differences are discussed in section 5.1 with individual subsections dedicated to global "all," "land," "sea," and "cloud-cleared" versus "clear-only" categories (sections 5.1.1, 5.1.2, 5.1.3, and 5.1.4). This is followed by the RMS differences for the tropics, midlatitude and high-latitude zones (section 5.1.5). The bias characteristics and possible sources for the biases seen in the AIRS retrievals are discussed in section 5.2. Section 5.3 provides a brief discussion on the statistics when all the AIRS retrievals are processed with a tighter time and distance collocation criteria (±1 hour and 50 km radius). Section 5.3 also discusses the retrieval statistics for day and night cases using subsets of samples extracted from all the accepted retrievals.

5.1. RMS Differences

[30] Tables 2 and 3 show the mean and standard deviation (a measure of RAOB variability) in temperature and water vapor in the RAOB measurements for "land," "sea," and "all" categories in conformity to the accepted number of AIRS retrievals dictated by the AIRS version 4.0 retrieval QA flags. The temperature values are in °K for 1 km layers, and the water vapor values are given as layer precipitable water (PCW) in cm for 2-km layers. The number of samples for day and night for each category is close to 50% and the variability is more or less same, and hence not shown in the tables. Although RAOB measurements have equal number samples at each pressure level, the number of samples used in calculating the statistics for each 1 km layer depends on the QA flags. One obvious feature of the quality flags is the screening-out of many land samples at the surface, and allowing more number of land samples in the middle and upper troposphere.

5.1.1. Global RMS Differences for "All" Category

[31] Figures 4a and 4b show the RMS difference for the temperature and water vapor profiles for "all" (land/sea/coastal) accepted global samples (90°N–90°S latitude, 180°W–180°E longitude). Figures 5a and 5b and Figures 6a and 6b show similar RMS difference plots using global samples for "land" and "sea" categories, respectively. The

Table 3. RAOB Water Vapor Variability (Mean and Standard Deviation, STD) Over Global "Land," "Sea," and "All" Categories and Corresponding Number of Samples Accepted by the Version 4.0 Emulation[a]

Pressure, Upper Boundary, hPa	Pressure, Lower Boundary, hPa	Land			Sea			All (Land/Sea/Coast)		
		N	Mean PCW, cm	STD PCW, cm	N	Mean PCW, cm	STD PCW, cm	N	Mean PCW, cm	STD PCW, cm
201	314	13837	0.01	0.01	2854	0.01	0.01	44569	0.01	0.01
314	407	13837	0.02	0.02	2854	0.03	0.03	44569	0.02	0.02
407	516	13837	0.06	0.06	2854	0.07	0.08	44569	0.06	0.06
516	618	13795	0.12	0.11	2854	0.14	0.14	44520	0.12	0.11
618	707	11488	0.18	0.14	2854	0.20	0.19	41161	0.17	0.15
707	853	4054	0.52	0.34	2315	0.67	0.47	17267	0.53	0.40
853	1100	3534	0.72	0.46	2255	1.35	0.79	15601	0.91	0.68

[a]Water vapor values are given as layer precipitable water (PCW).

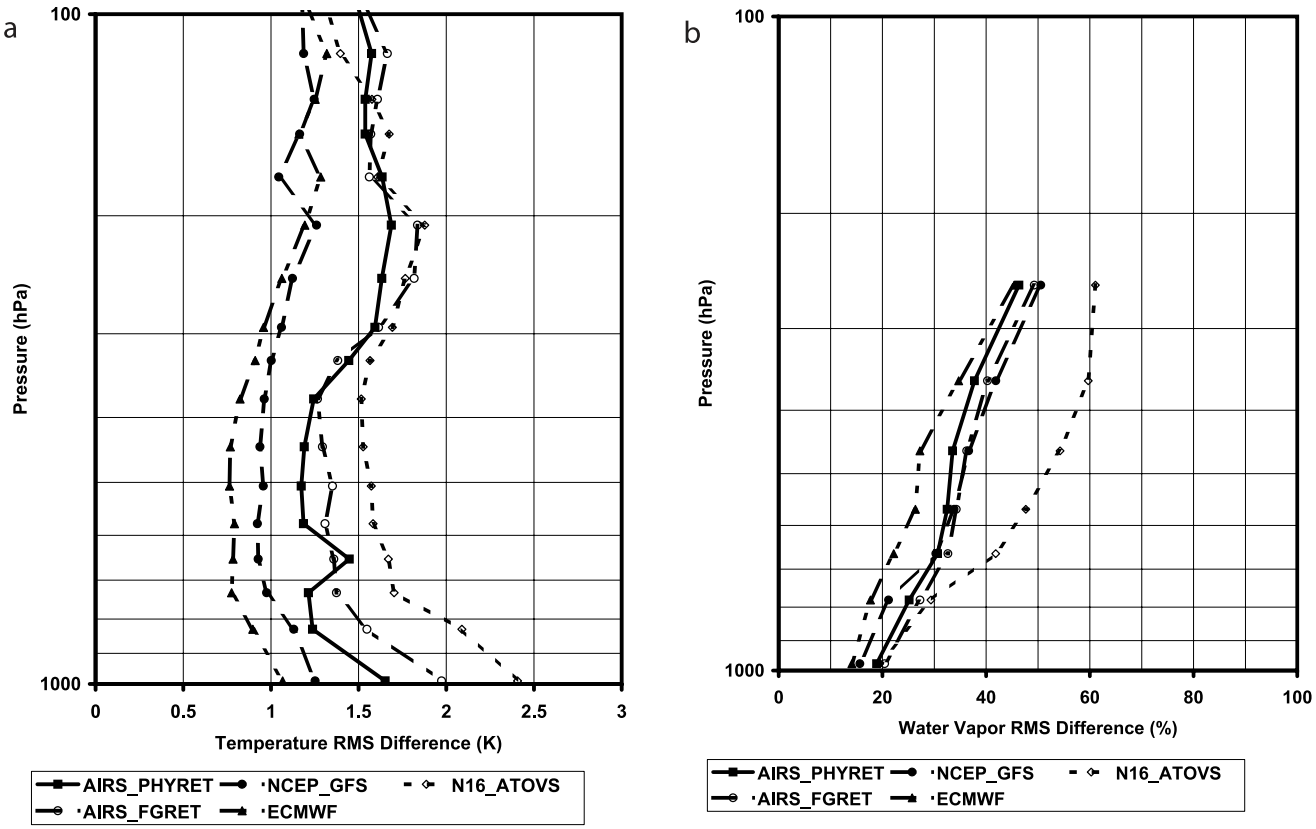

Figure 4. (a) Global temperature RMS differences for all the accepted samples for the "all" category: RAOB versus AIRS_PHYRET, solid squares; RAOB versus NCEP_GFS, solid circles; RAOB versus ATOVS, open diamonds; RAOB versus AIRS_FGRET, open circles; and RAOB versus ECMWF, solid triangles. (b) Global water vapor RMS differences for all the accepted samples for the "all" category: RAOB versus AIRS_PHYRET, solid squares; RAOB versus NCEP_GFS, solid circles; RAOB versus ATOVS, open diamonds; RAOB versus AIRS_FGRET, open circles; and RAOB versus ECMWF, solid triangles.

RAOB observation is taken as truth and bias and RMS are computed for the differences between AIRS physical retrieval and the RAOB (RAOB versus AIRS_PHYRET, solid squares). Similarly statistics are computed for NOAA 16 ATOVS retrievals (RAOB versus ATOVS, open diamonds), and for the NCEP_GFS (RAOB versus NCEP_GFS, solid circles) and ECMWF (RAOB versus ECMWF, solid triangles) forecasts. In addition, each figure has the statistics for the AIRS fast eigenvector regression solution (RAOB versus AIRS_FGRET, open circles), a retrieval step before the AIRS physical retrieval.

[32] A first glance on the global statistics (Figures 4a and 4b) indicate that both the ECMWF and NCEP_GFS model forecast temperature RMS difference from RAOBs is less than 1°K for 1 km layers in the troposphere with the ECMWF forecast agreeing a little better to that of NCEP_GFS. The RMS difference for the whole troposphere is slightly less than 1.0°K for the ECMWF and about 1.0°K for the NCEP_GFS. Both models show an increase in RMS difference with height above 300 hPa probably due to shortcomings in the RAOB data. The surface water vapor RMS difference is 14% for the 2 km layers and gradually degrades at higher levels. In the middle and upper troposphere, the ECMWF water vapor is better than NCEP_GFS by about 10%. This may be due to the finer vertical (60 levels of ECMWF versus 26 levels of NCEP_GFS) and Gaussian grid resolution (0.5° Lat x 0.5° Lon ECMWF versus 1° Lat x 1° Lon NCEP_GFS) of the ECMWF profiles used in the comparison. In addition to the performance of the models, the smaller RMS difference may also be attributed to the fact that the model forecasts heavily utilize the RAOB information in the analysis. Therefore the ensemble of RAOBs should agree with the model forecasts at the RAOB locations. The agreement between the forecasts and the RAOBs also provides an assessment of the quality of the global RAOB compilation used to compare satellite retrievals.

[33] With regards to AIRS_PHYRET, both the temperature and water vapor retrievals are in good agreement with the RAOBs. Except for the surface point, and above 300 hPa, the temperature RMS for most of the tropospheric layers is about 1.2° K. Between 700 and 600 hPa, the RMS difference shows a kink, because of bias discussed in section 5.2. The RMS difference for the whole troposphere is about 1.3°K for the AIRS_PHYRET. The AIRS_FGRET also shows good agreement with RAOBs and the AIRS_PHYRET brings an improvement of about four tenths of a degree over AIRS_FGRET. With regards to water vapor, except at the surface, the percent RMS difference from AIRS_PHYRET is even better than the NCEP_GFS, and

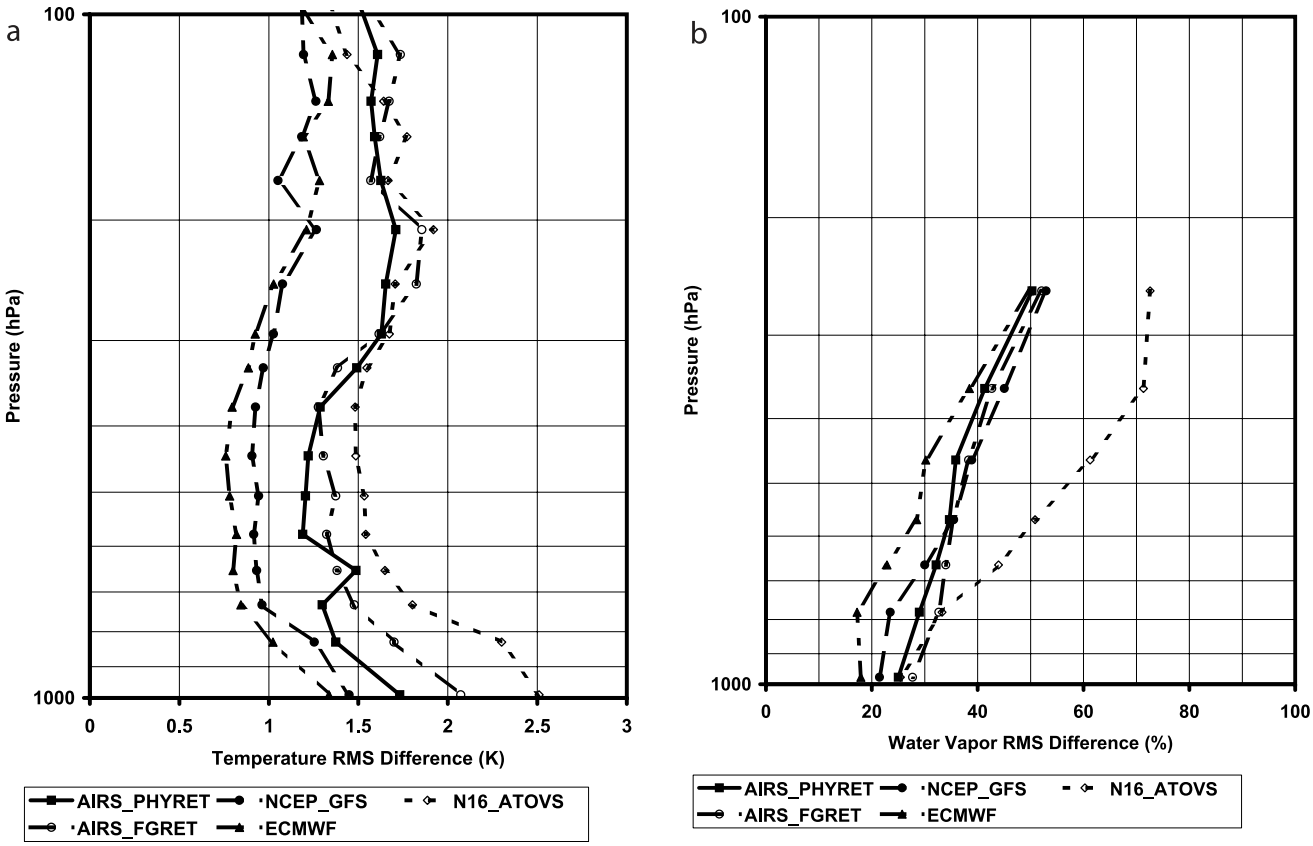

Figure 5. (a) Same as in Figure 4a but for the "land" category. (b) Same as in Figure 4b but for the "land" category.

the AIRS_FGRET RMS difference is comparable to AIRS_PHYRET. The large variability of water vapor in the middle and upper troposphere increases the RMS difference at higher levels for the retrievals as well as for the model forecasts. Added to this, the uncertainties in RAOB measured water vapor due to sensor characteristics can also affect the agreement between RAOBs and retrievals. *McMillin et al.* [2005] have discussed the issues of the RAOB sensor characteristics for a variety of RAOBs and their impact on the AIRS moisture comparisons over the contiguous United States. The observed trends of RMS differences for water vapor presented in this study reveal considerable agreement with the results presented by L. M. McMillin et al. (Radiosonde humidity corrections and AIRS moisture data validation, submitted to *Journal of Geophysical Research*, 2005, hereinafter referred to as McMillan et al., submitted manuscript, 2005). With respect to AIRS and ATOVS retrievals, the AIRS_PHYRET, in general, shows an improvement of at least 0.5°K RMS difference over ATOVS. The ATOVS statistics for this data set are similar to the comparison statistics reported by the ATOVS sounding product research team [*Reale*, 2002] and associated web pages (http://www.orbit.nesdis.noaa.gov/smcd/opdb/poes/vstats/). The RMS difference from the fast regression step (AIRS_FGRET) is in between the RMS differences of the AIRS_PHYRET and ATOVS retrievals. For water vapor, the AIRS_PHYRET difference is significantly better than the ATOVS and closely follows the ECMWF. The ATOVS water vapor retrieval algorithm uses mainly three HIRS and AMSU-B water vapor channels in the water vapor retrieval. The surface water vapor RMS difference from ATOVS coincides with that of AIRS, suggesting that the improvement in ATOVS is due to the use of AMSU-B channels. Also, the low variability of water vapor in the boundary layer might contribute to better agreement between the AIRS and ATOVS retrievals.

5.1.2. Global RMS Differences for "Land" Category

[34] An examination of the temperature RMS differences from the global "land" samples (Figure 5a) reveal the difficulties in the retrievals at the surface due to the heterogeneity of the land surface and the associated spectral emissivity variations [*Salisbury and D'Aria*, 1992, 1994]. This is also reflected in the lower percentage of samples accepted at the land surface by the surface QA flag Qual_surf (Table 2). The model forecasts also have the difficulty predicting the surface as seen by larger RMS differences at the surface. In addition to the uncertainties in the spectral emissivity, daytime convective buildup in the boundary layer, and the error in the interpolated surface pressure due to topography might have contributed to the larger RMS difference seen at the surface. Above 500 hPa, the quality flags used for the land in the version 4.0 retrieval appears to allow more samples than it should thus increasing the temperature RMS by about two tenths of a degree. The effect is more pronounced in the water vapor RMS over "land" category (Figure 5b). This is based on the experience with the AIRS retrieval version 3.7 quality flags (not shown) that accepts or rejects the entire atmospheric state

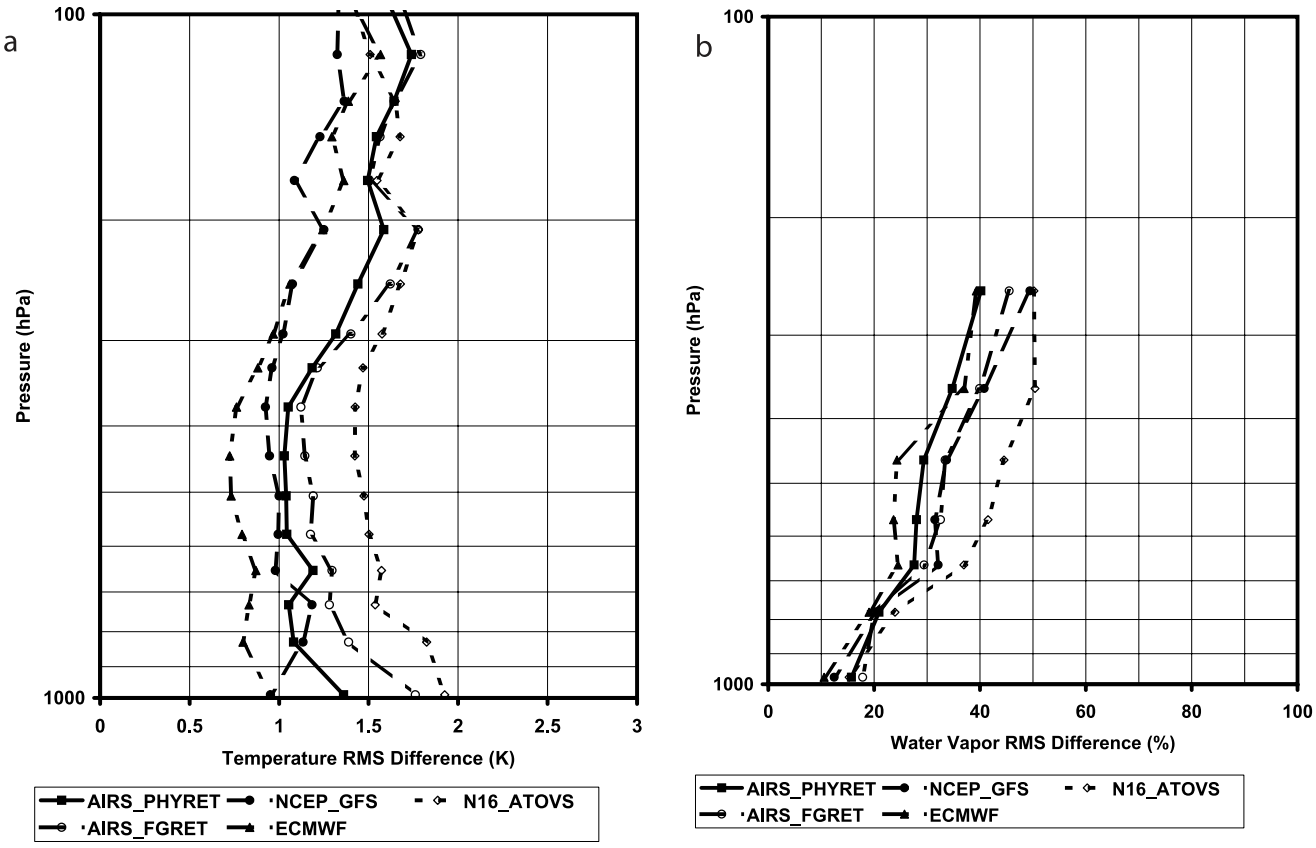

Figure 6. (a) Same as in Figure 4a but for the "sea" category. (b) Same as in Figure 4b but for the "sea" category.

over land for both the temperature and water vapor. With respect to AIRS_PHYRET and ATOVS comparisons over land, the AIRS_ PHYRET shows a larger improvement at the surface (0.65°K), and maintains an improvement of at least 0.5°K in the middle and upper troposphere. The water vapor retrieval from AIRS_PHYRET also shows significant improvement (~20%) over ATOVS for most of the troposphere.

5.1.3. Global RMS Differences for "Sea" Category

[35] Over the "sea" category, the RMS difference depicted in Figures 6a and 6b for the temperature and water vapor reveal the best that could be achieved with the AIRS cloud-cleared and accepted samples. Except for the surface point, and a kink at 680 hPa, the RMS difference for the troposphere is very close to the expected accuracy of 1°K for the 1 km layers. The water vapor error also falls in close proximity to the expected RMS difference. However, because of larger variability and uncertainties in the RAOB measurements of water vapor in the middle and upper troposphere, the RMS differences are larger than the anticipated RMS through simulations [*Goldberg et al.*, 2003].

[36] For both the land and sea cases, uncertainties in the specification of surface pressure for the retrievals through interpolation of NCEP_GFS surface pressure (the only ancillary parameter in the AIRS APS) might be causing an error term, which requires further investigation. Since the global samples contain a large number of samples from coastal and land samples, the global sample also show larger RMS difference at the surface due to surface effects from the land and coastal categories. Despite these aberrations, the AIRS_PHYRET offers superior performance in both the temperature and water vapor retrievals over the ATOVS.

[37] Above 300 hPa, both the AIRS and ATOVS retrievals, and both the models show a tendency for increased RMS difference for all the categories. This is due to the difficulty in predicting or retrieving the cold and nearly isothermal atmosphere near and around the tropopause and the variations in the tropopause height in the global sample. The RMS difference is also weighted heavily with the sample size covering the latitude belt 45°–65° where the tropopause height is lower. The increase in RMS may also be attributed to the difficulties of RAOB measurements in the jet stream regions. In general, the RAOB balloons may move with the wind. Over the jet stream regions, the RAOB drifts may be vigorous, and the RAOB measurement may not represent the true vertical retrievals made by the satellite instruments.

5.1.4. RMS Differences for "Cloud-Cleared" and "Clear-Only" Cases

[38] The noise on the cloud-cleared radiances is a function of the spatial distribution and quantity of clouds [*Susskind et al.*, 2003]. For clear golf balls (3 x 3 FOVs), the noise amplification factor ranges from 1/3 to large values in cloudy scenes. Therefore we expect retrieved statistics to have some sensitivity to clouds. To study the impact of noise amplification, statistics are computed for the AIRS_PHYRET for "clear-only" cases and are compared

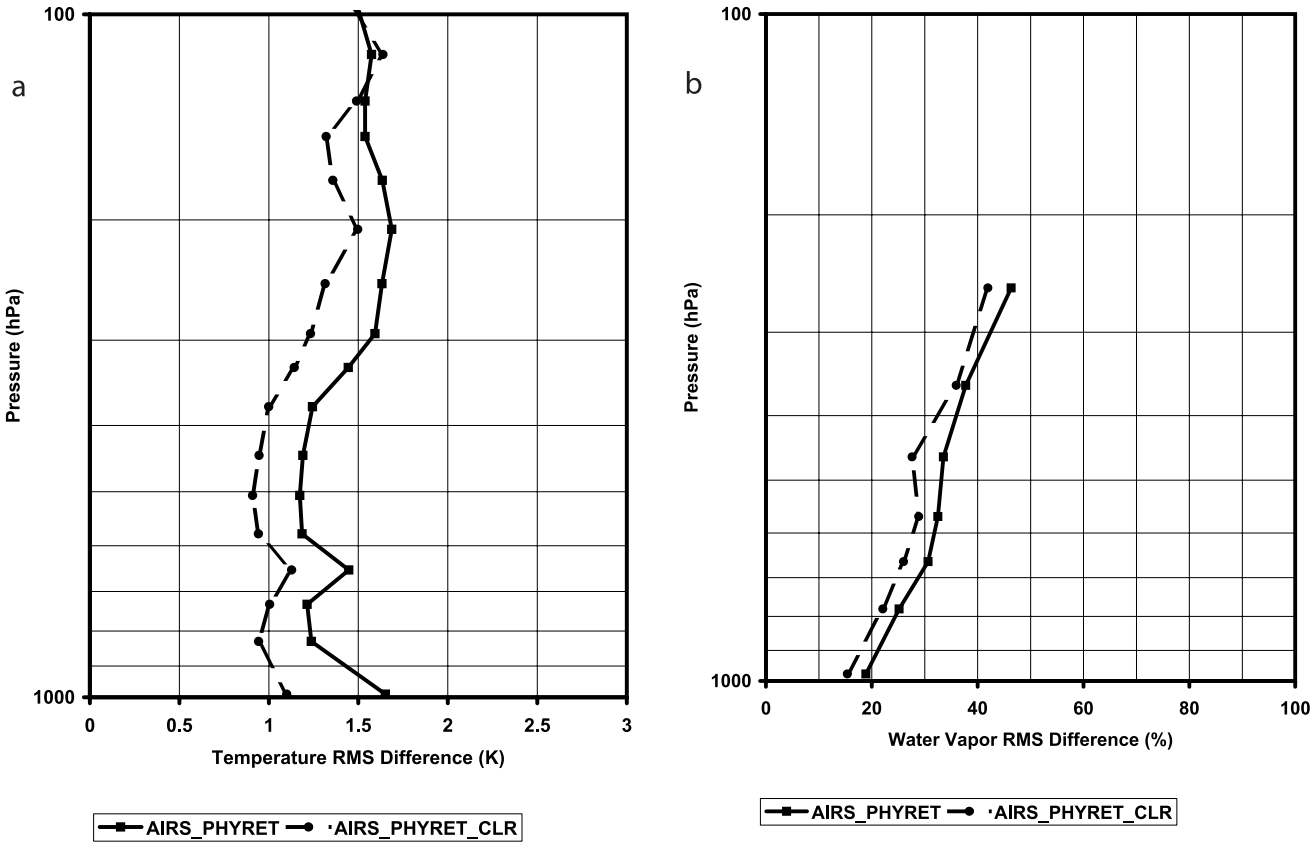

Figure 7. (a) Global temperature RMS differences for all the accepted ''cloud-cleared'' (RAOB versus AIRS_PHYRET, solid squares) and ''clear-only'' (RAOB versus AIRS_PHYRET_CLR, solid circles) samples for the ''all'' category. (b) Global water vapor RMS differences for all the accepted ''cloud-cleared'' (RAOB versus AIRS_PHYRET, solid squares) and ''clear-only'' (RAOB versus AIRS_PHYRET_CLR, solid circles) samples for the ''all'' category.

with the global accepted ''cloud-cleared'' cases. Figures 7a and 7b show these comparisons for the ''all'' (land/sea/coast) category. Similar plots using global samples for the ''sea'' only category are shown in Figures 8a and 8b. Tables 4 and 5 show the number of samples accepted by the version 4.0 QA flags, and the variability in the RAOB measurements for the clear-only samples. The sample size for the ''clear-only'' data for ''all'' category is about 1000 compared to the accepted sample size of about 60000 (Table 2). The number of ''clear-only'' samples for the ''sea'' is about 400 compared to the ''cloud-cleared'' sample size of about 5000. There are no clear-only samples over the ''land'' and thus the samples for the ''all'' category includes samples from ''sea'' and ''coastal'' categories. An examination of the temperature RMS difference from Figure 7a reveals that the ''clear-only'' retrievals (RAOB versus AIRS_ PHYRET_CLR, solid circles) show an improvement of about 0.5°K at the surface and about 0.3°K RMS at other levels over ''cloud-cleared'' samples (RAOB versus AIRS_ PHYRET, solid squares, same as in Figure 4a) for ''all'' category. The larger RMS difference in the ''cloud-cleared'' cases can be attributed to the noise amplification in the cloud-clearing process. The noise amplification factor is less than 0.5 for 60% of the ''cloud-cleared'' cases. About 35% of the cases have noise amplification in the range 0.5 to 1.0, and the remaining 5% have a noise amplification factor in the range 1.0 to 1.5. Both the ''clear-only'' and ''cloud-cleared'' RMS differences over the ''sea'' (Figure 8a, RAOB versus AIRS_PHYRET_CLR, solid circles; RAOB versus AIRS_PHYRET solid squares, same as in Figure 6a) are close to the expected accuracy, and the difference between clear and cloud-cleared result is not as pronounced as that of the ''all'' category (Figure 7a). The tendency of increased RMS difference in the temperature above 300 hPa is evident in both the cloud-cleared and ''clear-only'' cases. This again, strongly suggests that the RAOB measurements have uncertainties at those levels. Similarly, Figure 7b depicts an improvement of about 4% in the water vapor RMS difference for ''clear-only'' cases from ''cloud-cleared'' cases for ''all'' category. The water vapor accuracy for the ''sea'' cases (Figure 8b) is close to the expected accuracy at the surface (15% in 2 km layers), and there is no appreciable change in the RMS difference between ''clear-only'' and ''cloud-clear'' cases. This essentially indicates that cloud-clearing methodology over the sea is effective and the noise amplification is relatively less than the cases from a mix of samples over the land and coastal regions. The clear test algorithm uses a combination of tests to detect clear golf balls; however, the clear scenes may still have some residual cloud contamination. This is true even with the cloud-cleared cases. Despite the residual cloud contamination, the difference between the cloud-cleared results

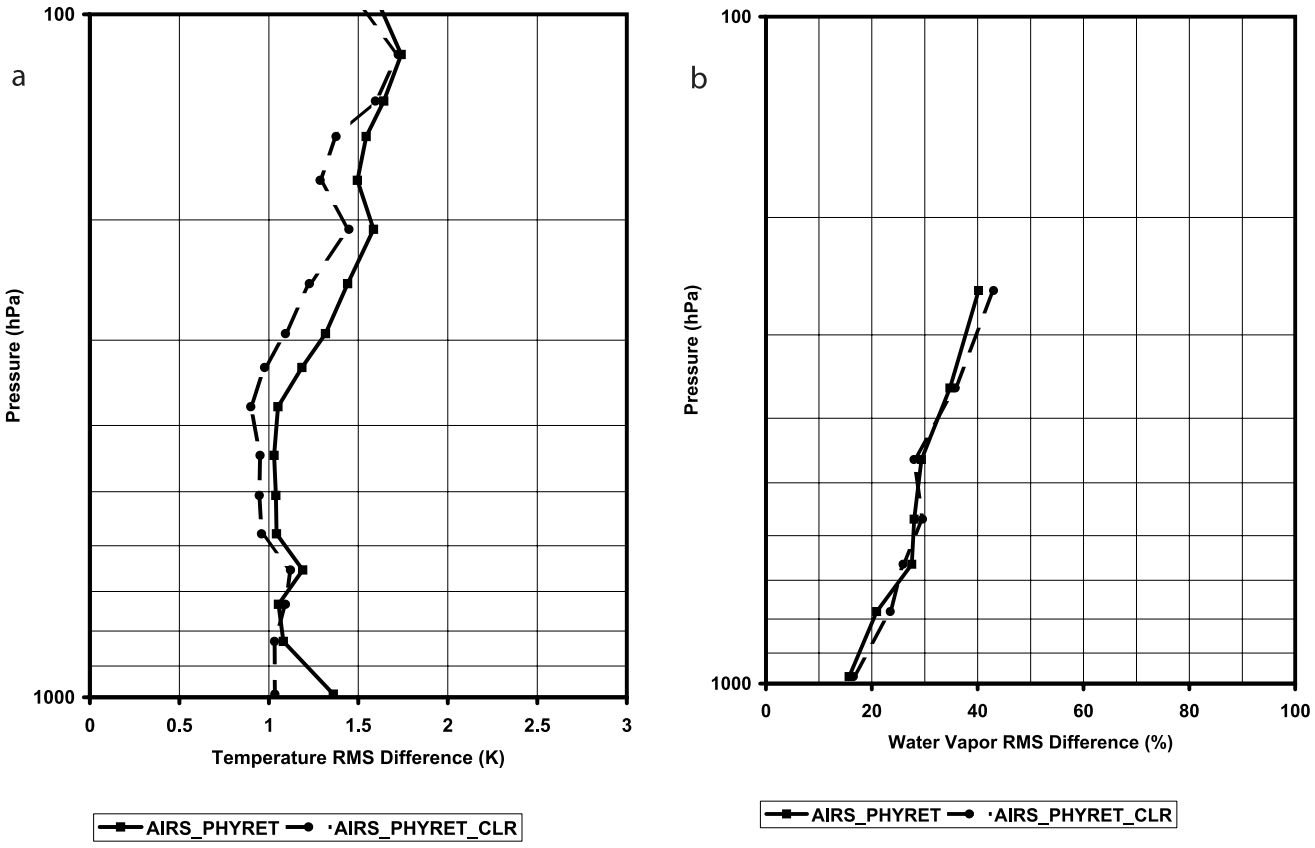

Figure 8. (a) Same as in Figure 7a but for the "sea" category. (b) Same as in Figure 7b but for the "sea" category.

(N∼60,000) and the clear-only results (N∼1000) is well within our expectations because of the noise amplification.

5.1.5. RMS Differences for Tropics, Midlatitudes, and High Latitudes

[39] The global statistics discussed so far might have been influenced by the larger sample size in the NH from middle and high-latitude land and coastal regions. When the global data set is separated into tropics, midlatitudes and high-latitude regions, the computed statistics could reveal the ability of the retrieval algorithm to retrieve contrasting atmospheric state of the region. To study this aspect, statistics are computed for AIRS_PHYRET and RAOBs for the subset of samples covering tropics (23°N–23°S), midlatitudes (50°N–23°N; 50°S–23°S), and high-latitude (90°N–50°N; 90°S–50°S) zones.

[40] Tables 6 and 7 show the mean and variability in the RAOB temperature and water vapor profiles for these three regions. The warm and humid tropical atmospheres with relatively low variability, and contrasting cold and dry high-latitude atmospheres with larger variability, and the variances in between for the midlatitudes are evident from these tables. Figures 9a and 9b show the temperature and water vapor

Table 4. RAOB Temperature Variability (Mean and Standard Deviation, STD) for "Clear-Only" Cases for "Sea" and "All" Categories and the Corresponding Number of Samples Accepted by the Version 4.0 Emulation

Pressure, Upper Boundary, hPa	Pressure, Lower Boundary, hPa	Sea			All (Land/Sea/Coast)		
		N	Mean, °K	STD, °K	N	Mean, °K	STD, °K
103	126	420	207.7	10.4	999	211.2	11.3
126	142	420	210.2	8.6	999	213.2	9.5
142	160	420	212.6	7.0	999	215.0	7.8
160	190	420	215.9	4.8	999	217.4	5.7
190	223	420	220.0	4.3	999	220.4	4.7
223	273	292	226.5	5.6	781	225.1	5.8
273	314	292	233.7	7.7	781	231.3	8.0
314	344	292	239.2	8.6	781	236.5	9.1
344	407	292	245.9	9.1	781	243.0	9.7
407	478	292	254.4	9.2	781	251.4	10.0
478	535	292	261.2	9.0	781	258.3	10.0
535	618	292	267.4	8.8	781	264.4	9.9
618	684	292	273.0	8.8	781	270.1	9.9
684	778	280	278.4	8.5	677	276.9	9.2
778	879	252	282.7	8.5	503	281.3	9.6
879	1100	240	289.4	7.7	477	287.9	9.4

Table 5. RAOB Water Vapor Variability (Mean and Standard Deviation, STD) for "Clear-Only" Cases for "Sea" and "All" Categories and the Corresponding Number of Samples Accepted by the Version 4.0 Emulation[a]

		Sea			All (Land/Sea/Coast)		
Pressure, Upper Boundary, hPa	Pressure, Lower Boundary, hPa	N	Mean PCW, cm	STD PCW, cm	N	Mean PCW, cm	STD PCW, cm
201	314	292	0.01	0.01	781	0.01	0.01
14	407	292	0.02	0.03	781	0.02	0.02
407	516	292	0.07	0.07	781	0.06	0.07
516	618	292	0.14	0.13	781	0.12	0.13
618	707	292	0.21	0.18	781	0.19	0.17
707	853	259	0.71	0.46	525	0.67	0.45
853	1100	252	1.50	0.80	503	1.38	0.82

[a]Water vapor values are given as layer precipitable water (PCW).

RMS difference for the AIRS_PHYRET for these latitude regions using "all" samples. The plot shown in Figures 9a and 9b for global retrievals (AIRS_PHYRET, solid squares) is the same as shown earlier in Figures 4a and 4b. Other statistics are for RAOB versus AIRS_PHYRET for the tropics (TROP_AIRS_PHYRET, solid circles), midlatitude (MIDLAT_AIRS_PHYRET, open diamonds), and high-latitude (HIGHLAT_AIRS_PHYRET, open circles) regions. The RMS differences for the other collocated data sets (ATOVS, ECMWF, and NCEP_GFS) portray the same trend as in global comparisons discussed earlier. The increase in RMS difference with height (above 300 hPa) is evident for all the regions and the trend resembles more or less the global RMS differences. An examination of Figure 9a reveals that for the tropical region with less variability in the truth (Table 6), the retrieval RMS difference is smaller and close to 1°K for the whole troposphere. The kink in the RMS seen at around 680 hPa for the midlatitude, high-latitude, and global cases is not evident for the tropics. Similarly, the RMS difference is also less between 300 hPa and 200 hPa, but increases approaching 100 hPa possibly because of some spurious but quality-checked RAOB measurements or because of deficiencies in the AIRS retrieval system in the detection of the tropical tropopause, which tends to be vertically localized. For midlatitude and high-latitude regions, the RMS difference is correspondingly higher between 300 hPa and 200 hPa. This is probably due to larger variability (Table 6) and lower tropopause height in the midlatitude and high-latitude regions. The global RMS follows proportionately with the sample contributions from these three regions. The water vapor RMS also show similar trend with a slightly larger RMS difference for the midlatitude compared to high-latitude because of larger percentage of land samples (~17% Table 1) contributing to the RMS difference. Figures 10a and 10b show the temperature and water vapor RMS differences for "sea" samples for these three latitude regions along with global sea retrieval RMS shown earlier in Figures 6a and 6b (AIRS_PHYRET for sea, solid squares). The temperature RMS difference for the tropical sea is close to or better than 1°K except for the surface point. The water vapor percent RMS is about 12% at the surface. These results resemble the comparison statistics presented by *Tobin et al.* [2006] for the ARM/CART TWP site using multiple RAOB ascents synchronized to AIRS overpass times. Similarly, when the retrievals over land are analyzed, both the bias (presented in section 5.2) and the RMS difference with the global RAOBs resemble the results for the SGP site with dedicated RAOB ascents.

[41] Despite quality-checked RAOB observations, the operational global RAOB observations may have many gross instrument and systematic errors because of different instrument types used by the global RAOB stations [*Eskridge et al.*, 2003]. In addition, when a time series of data is used from various RAOB stations, systematic errors

Table 6. RAOB Temperature Variability (Mean and Standard Deviation, STD) for Tropical, Midlatitude and High-Latitude Zones, and Corresponding Number of Samples Accepted by the Version 4.0 Emulation

		Tropical (23°N–23°S)			Midlatitude (50°N–23°N; 50°S–23°S)			High-Latitude (90°N–50°N; 90°S–50°S)			All (Land/Sea/Coast) (90°N–90°S)		
Pressure, Upper Boundary, hPa	Pressure, Lower Boundary, hPa	N	Mean, °K	STD, °K	N	Mean, °K	STD, °K	N	Mean, °K	STD, °K	N	Mean, °K	STD, °K
103	126	5246	198.2	3.0	25609	210.8	6.0	28578	219.2	6.8	59433	213.7	8.8
126	142	5246	202.7	2.5	25609	212.6	5.2	28578	219.7	6.6	59433	215.1	7.7
142	160	5246	207.0	2.1	25609	214.1	4.7	28578	220.0	6.5	59433	216.3	6.8
160	190	5246	213.4	1.7	25609	215.8	4.2	28578	219.7	6.7	59433	217.5	5.9
190	223	5246	221.1	1.9	25609	218.4	5.0	28578	219.5	6.6	59433	219.1	5.7
223	273	4120	230.4	2.2	20272	223.6	5.5	20177	220.2	5.3	44569	222.7	6.0
273	314	4120	239.5	2.5	20272	230.5	6.2	20177	223.5	5.4	44569	228.2	7.4
314	344	4120	245.7	2.6	20272	236.1	6.5	20177	227.7	6.2	44569	233.2	8.4
344	407	4120	252.6	2.6	20272	242.9	6.7	20177	233.5	7.2	44569	239.5	9.1
407	478	4120	260.9	2.3	20272	251.6	6.8	20177	241.5	7.9	44569	247.9	9.5
478	535	4120	267.4	2.1	20269	258.7	6.7	20157	248.3	8.3	44546	254.8	9.6
535	618	4098	273.4	1.9	20148	265.1	6.7	20157	254.4	8.4	44403	261.0	9.7
618	684	3612	279.2	1.9	17657	270.9	7.0	19876	260.1	8.5	41145	266.4	9.9
684	778	3217	284.3	2.2	12104	277.2	7.0	11328	266.9	8.6	26649	273.7	9.6
778	879	2038	288.8	2.4	5533	280.0	7.5	8171	270.1	9.0	15742	276.0	10.4
879	1100	1902	295.1	2.4	3941	286.1	7.7	6106	276.2	9.6	11949	282.5	10.9

Table 7. RAOB Water Vapor Variability (Mean and Standard Deviation, STD) for Tropical, Midlatitude and High-Latitude Zones, and Corresponding Number of Samples Accepted by the Version 4.0 Emulation[a]

Pressure, Upper Boundary, hPa	Pressure, Lower Boundary, hPa	Tropical (23°N–23°S)			Midlatitude (50°N–23°N; 50°S–23°S)			High-Latitude (90°N–50°N; 90°S–50°S)			ALL (Land/Sea/Coast) (90°N–90°S)		
		N	Mean PCW, cm	STD PCW, cm	N	Mean PCW, cm	STD PCW, cm	N	Mean PCW, cm	STD PCW, cm	N	Mean PCW, cm	STD PCW, cm
201	314	4120	0.01	0.01	20272	0.01	0.01	20177	0.00	0.00	44569	0.01	0.01
314	407	4120	0.05	0.04	20272	0.02	0.02	20177	0.01	0.01	44569	0.02	0.02
407	516	4120	0.13	0.11	20272	0.06	0.06	20177	0.04	0.04	44569	0.06	0.06
516	618	4119	0.25	0.18	20244	0.13	0.11	20157	0.08	0.07	44520	0.12	0.11
618	707	3612	0.35	0.22	17657	0.19	0.14	19892	0.12	0.10	41161	0.17	0.15
707	853	2189	1.13	0.41	6704	0.57	0.33	8374	0.35	0.25	17267	0.53	0.40
853	1100	2021	2.18	0.45	5420	0.96	0.52	8160	0.56	0.36	15601	0.91	0.68

[a]Water vapor values are given as layer precipitable water (PCW).

could result because of data reduction procedures adapted by the stations, and because of many temperature correction schemes implemented from time to time [*Luers*, 1997]. Some of these difficulties with the RAOB errors could be mitigated with synchronized RAOB ascents, and a comparison with these RAOBs might show better statistics to prove the performance of an instrument. However, the global comparisons provide a reasonable gauge for the instrument performance, and, when the RAOBs are properly quality-controlled, provide realistic results as like any special campaign periods with special RAOB ascents. Our results are similar to the results presented by *Tobin et al.* [2006] with synchronized RAOB ascents. Thus the analysis performed in this paper show a remarkable degree of confi-

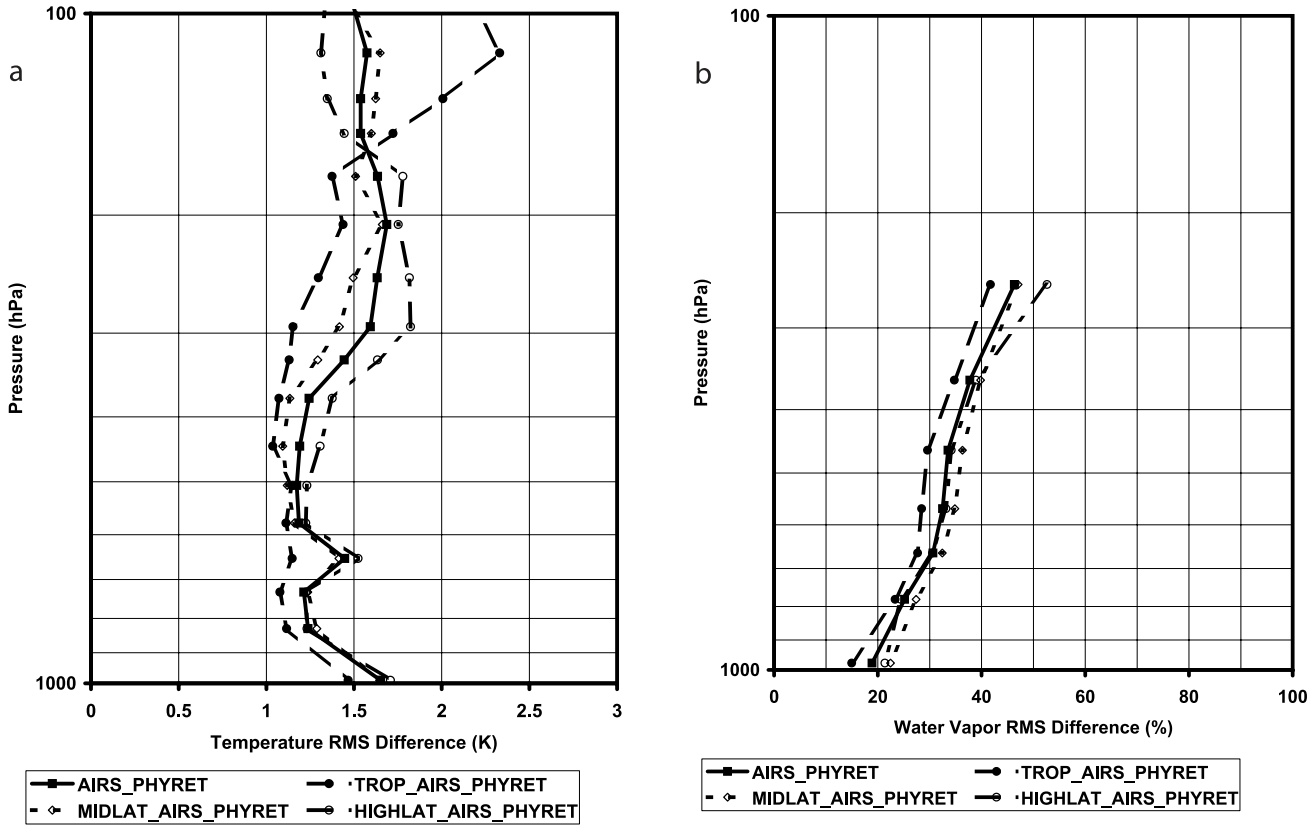

Figure 9. (a) Temperature RMS differences for different latitude zones: global (90°N–90°S, AIRS_PHYRET, solid squares), tropical (23°N–23°S, TROP_AIRS_PHYRET, solid circles), midlatitude (50°N–23°N; 50°S–23°S, MIDLAT_AIRS_PHYRET, open diamonds), and high-latitude (90°N–50°N; 90°S–50°S, HIGHLAT_AIRS_PHYRET, open circles) for the "all" category. (b) Water vapor RMS differences for different latitude zones: global (90°N–90°S, AIRS_PHYRET, solid squares), tropical (23°N–23°S, TROP_AIRS_PHYRET, solid circles), midlatitude (50°N–23°N; 50°S–23°S, MIDLAT_AIRS_PHYRET, open diamonds), and high-latitude (90°N–50°N; 90°S–50°S, HIGHLAT_AIRS_PHYRET, open circles) for the "all" category.

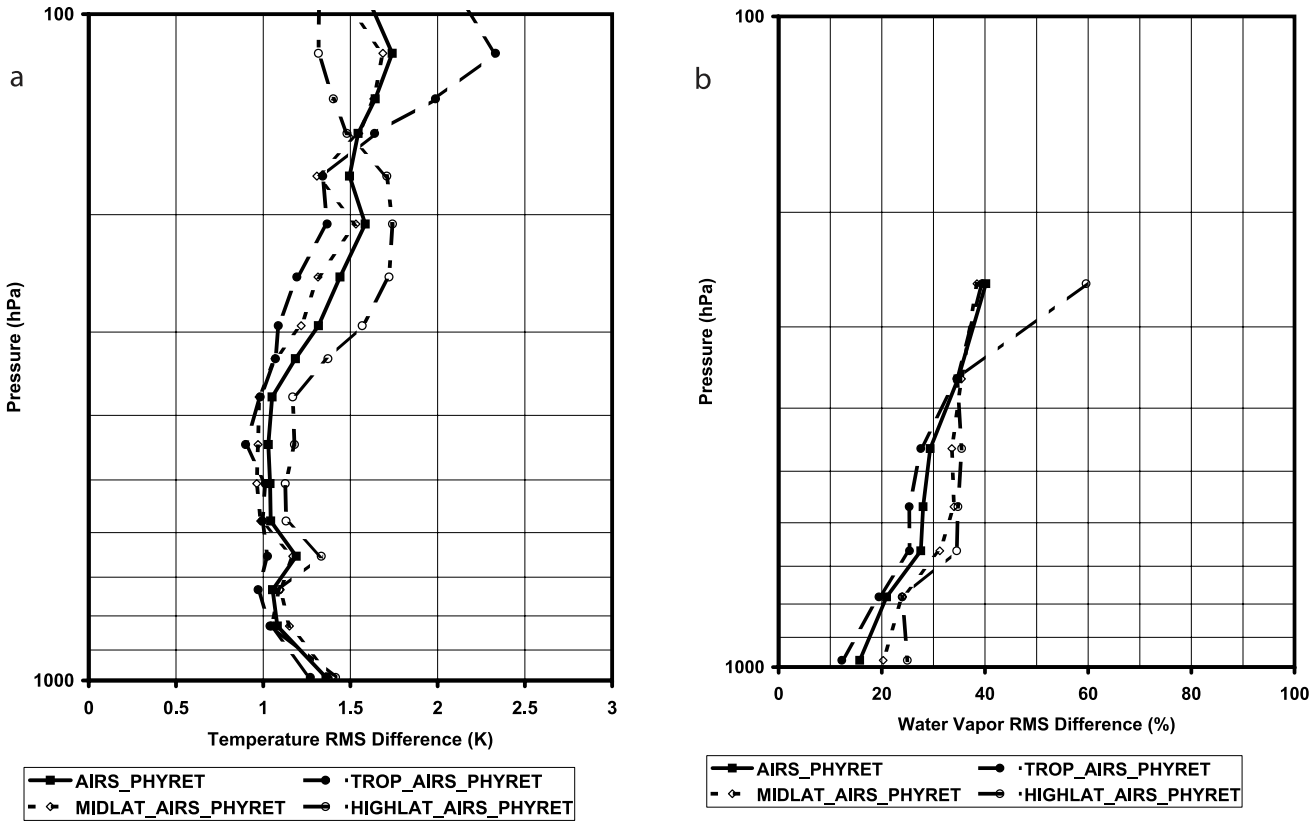

Figure 10. (a) Same as in Figure 9a but for the "sea" category. (b) Same as in Figure 9b but for the "sea" category.

dence in the AIRS retrievals despite the problems inherent in the correlative RAOB measurements.

5.2. Bias Characteristics

[42] Figures 11a and 11b show the bias characteristics for the temperature and water vapor profiles for "all" accepted global samples. Here, the RAOB observation is taken as a reference and the difference, denoted bias, is calculated between the RAOB and the following collocated data sets: the AIRS physical retrieval (AIRS_PHYRET, solid squares), ATOVS (open diamonds); NCEP_GFS (solid circles); ECMWF (solid triangles); AIRS fast regression (AIRS_FGRET, open circles). The corresponding RMS differences for these biases are shown in Figures 4a and 4b. The water vapor bias is presented as percentage to the reference (RAOB) water vapor in the 2 km layer (100 x (AIRS-RAOB)/RAOB). The AIRS_PHYRET, AIRS_FGRET, and the ECMWF show a negative water vapor bias for the upper troposphere region. This could be due to the RAOB measurements having a wet bias above 400 hPa, a tendency observed for soundings in cold polar and midlatitude air masses (T. Reale, 2005, http://foehninter. nesdis.noaa.gov/PSB/SOUNDINGS/ORA/index.html). Overall, the water vapor biases are within expected error bounds and are consistent with the results discussed by other investigators (McMillan et al., submitted manuscript, 2005). It may be noted that unlike the temperature biases, the water vapor biases (axis shown on expanded scale ±30) are a small fraction of the percent RMS indicative of the high spatial variability of water vapor.

[43] With respect to the temperature bias, the AIRS_PHYRET (Figure 11a) shows a larger temperature bias as compared to NCEP_GFS, ECMWF, and ATOVS. The ATOVS retrievals show a relatively smaller bias. This may be due to the tuning of ATOVS radiances using RAOB information, and the use of prior RAOB collocations to define first guess used in the ATOVS final retrieval algorithm [*Reale*, 2003]. Both the ECMWF and NCEP_GFS forecast models show smaller biases with RAOBs. This is expected since model forecasts utilize RAOB information in the analysis. The AIRS_FGRET bias follows the same trend as the ECMWF, probably because of use of ECMWF forecast/analysis as the training data set in the generation of regression coefficients [*Goldberg et al.*, 2003]. The AIRS_FGRET bias, although a little larger than that of ECMWF, is still smaller than the AIRS_PHYRET suggesting that the AIRS final physical retrieval step is increasing the bias from the fast regression first guess. Thus, out of all the collocated retrievals and forecasts, the AIRS retrieval is the only one that has no prior information or use of the RAOB profiles in its retrieval steps.

[44] Figure 12 shows the AIRS_PHYRET temperature bias for the "land" (LAND_AIRS_PHYRET, solid circles) and "sea" (SEA_AIRS_PHYRET, open diamonds) subsets along with the bias from the "all" category (shown earlier in Figure 11a, AIRS_PHYRET, solid squares). It is evident from Figure 12 that the land samples contribute predomi-

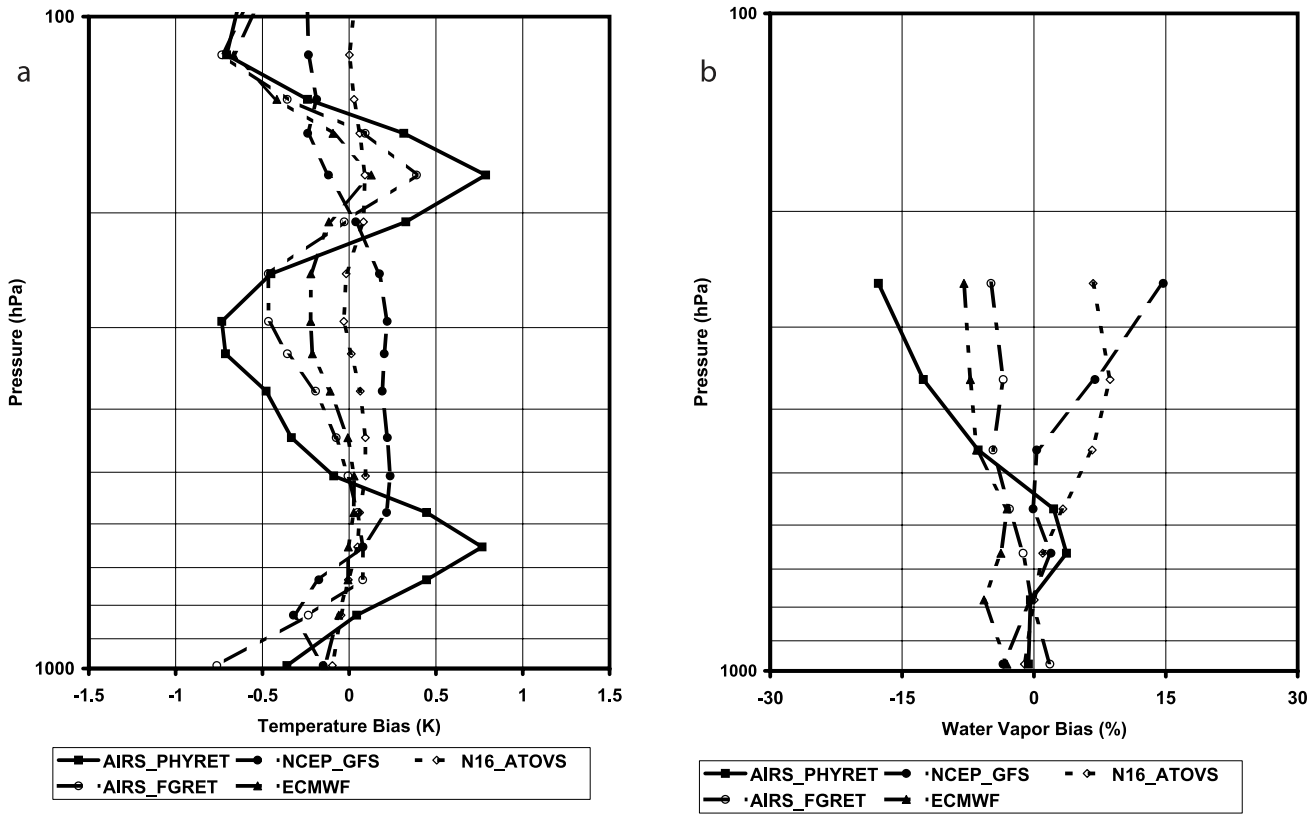

Figure 11. (a) Global temperature biases for all the accepted samples for the "all" category: RAOB versus AIRS_PHYRET, solid squares; RAOB versus NCEP_GFS, solid circles; RAOB versus ATOVS, open diamonds; RAOB versus AIRS_FGRET, open circles; and RAOB versus ECMWF, solid triangles. (b) Global water vapor biases for all the accepted samples for the "all" category: RAOB versus AIRS_PHYRET, solid squares; RAOB versus NCEP_GFS, solid circles; RAOB versus ATOVS, open diamonds; RAOB versus AIRS_FGRET, open circles; and RAOB versus ECMWF, solid triangles.

nantly to the overall bias. The sea samples also exhibit the oscillating trend in bias but the amplitude (about 0.3°K at 680 hPa) is quite small. When statistics are computed separating the global samples into many subsets (tropical-land, tropical-sea, midlatitude-land, midlatitude-sea, high-latitude-land, and high-latitude-sea), the samples from land and coastal categories for midlatitude and high-latitude regions seem to contribute mainly to the bias in AIRS_PHYRET because of the larger variability in those regions. The variability in the tropical samples (Table 6) is relatively less, and consequently the bias and the RMS differences are less for the tropical samples.

[45] Many factors might contribute to the amplification of the bias in the AIRS physical retrieval. One major factor could be the amplification of noise in the cloud-cleared radiance [*Susskind et al.*, 2003]. This is evident from Figure 13 where the AIRS_PHYRET bias is plotted for both the "clear-only" cases (AIRS_PHYRET_CLR, solid circles) and for all accepted (cloud-cleared) cases (AIRS_PHYRET, solid squares) from "all" category. The RMS differences corresponding to these biases are shown in Figure 7a. The "clear-only" retrieval also show similar trend in bias, but the magnitude of the bias is significantly lower than the cloud-cleared retrievals, suggesting that some proportion of the bias amplification is caused by the algorithm deficiencies (may retain some residual cloud amount) in cloud-clearing methodology. Samples from land and coastal categories contribute significantly to the bias; therefore the higher degree of difficulty of cloud clearing associated with the spatially heterogeneous (land and coastal) surfaces might be the contributing factor for the amplification of bias. Another plausible source could be errors in the upper atmospheric state sensed by the tails of the highest-peaking kernel functions. These errors appear to be inducing the oscillation into the upper stratosphere and lower troposphere. The biases in the lower troposphere appear to be dominated by an interaction with the microwave channels. In addition, there are other phenomena contributing to the bias.

[46] *Engelen et al.* [2001] have shown that the zonal, seasonal, and annual variations in CO_2 could cause errors in retrieved temperatures from theoretical instruments with similar spectral resolution to that of the AIRS. In addition, using the offline retrieval system for AIRS we have found that CO_2 variations of similar magnitude to expected seasonal and zonal variability will induce a bias in AIRS temperature retrievals relative to RAOB data over the entire troposphere. Because RAOB temperatures do not depend on an assumption of atmospheric CO_2 concentration, we therefore expect that differences between RAOB temperatures and AIRS retrieved temperatures in the troposphere will correlate with physical CO_2 variations.

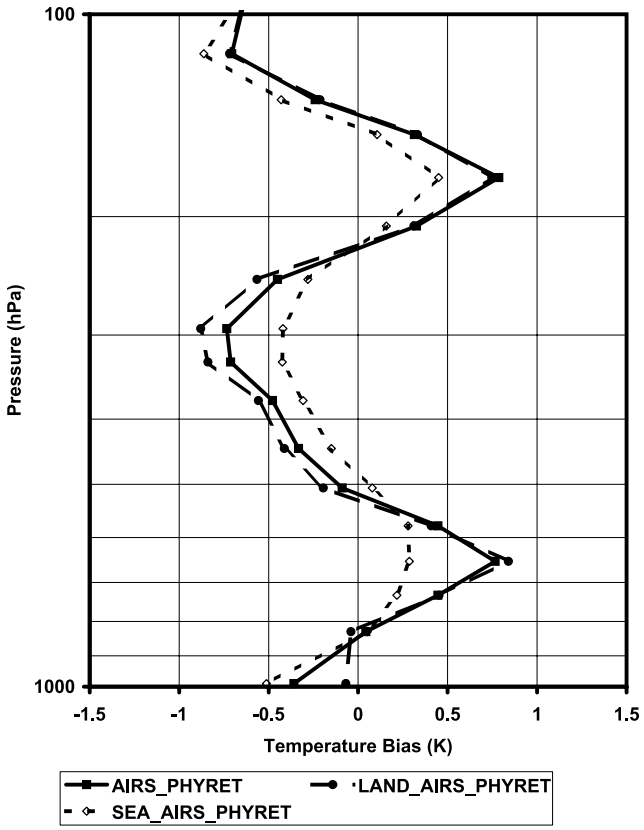

Figure 12. Global temperature bias over "all" (RAOB versus AIRS_PHYRET, solid squares), "land" (LAND_AIRS_PHYRET, solid circles), and "sea" (SEA_AIRS_PHYRET, open diamonds) categories.

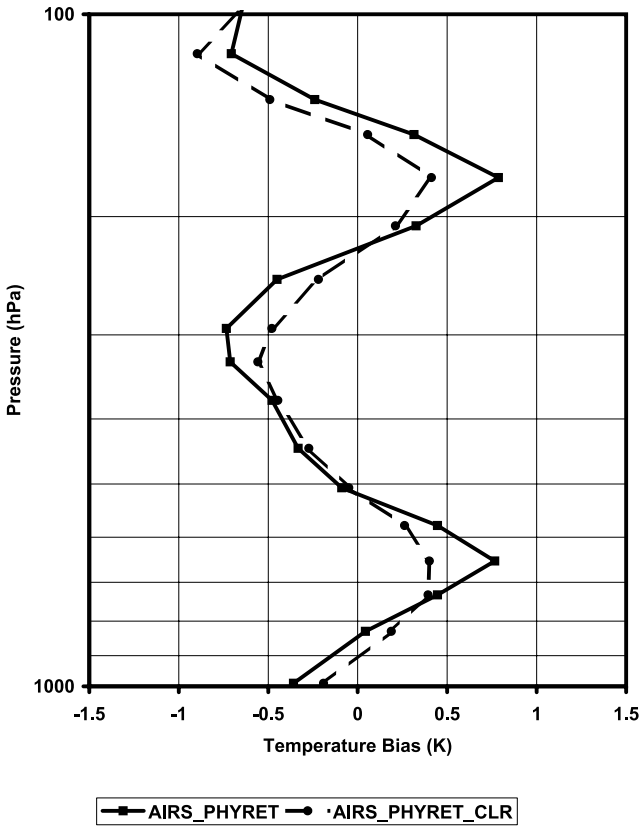

Figure 13. Global temperature bias for all the accepted "cloud-cleared" (RAOB versus AIRS_PHYRET, solid squares) and "clear-only" (RAOB versus AIRS_PHYRET_CLR, solid circles) samples for the "all" category.

[47] Figure 14 shows monthly global average differences between RAOB temperatures and AIRS retrieved temperatures (solid circles). Because IR sounders rely on thermal gradients to retrieve temperature and gas concentrations, the cold and nearly isothermal nature of the tropopause in midlatitude/polar regions represents a difficult portion of the atmosphere to resolve. Indeed most of the correlative RAOBs in our ensemble are located in the midlatitude to high-latitude regions (Table 1). Therefore, in order to mitigate uncertainties in the determination of the variable tropopause with season, we used a coarse layer between 450 hPa and the surface to calculate the trend in the difference between RAOB and AIRS retrievals over the 2-year period (September 2002 to September 2004). Also shown in Figure 14 is a smoothed representation of the raw differences using a 2-month sliding boxcar average (dotted line). The solid curve is the zonally weighted linear-least squares fit of the NOAA's Climate Monitoring Diagnostics Laboratory's (CMDL) Marine Boundary Layer (MBL) CO_2 product [*GLOBALVIEW-CO2*, 2004]. A least squares fit is used in lieu of the raw CMDL MBL product because this data set ends in December 2003; therefore trends beyond this month are extrapolated using the fitting coefficients. To produce the fitting coefficients, we calculated a global average of the raw CMDL MBL product weighted by the distribution of RAOB observation between January 2002 and December 2003 and then fit these averaged raw data to a second-order polynomial and third-harmonic terms. Differences between the averaged raw data and least squares fit are less than 0.5 parts-per-million by volume (ppmv).

[48] In Figure 14, the temperature differences and CO_2 variations show consistent annual and seasonal trends. We

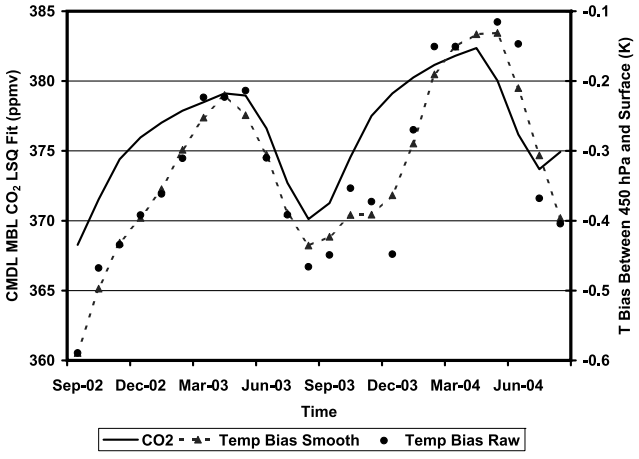

Figure 14. Seasonal trends between AIRS retrieval bias 450 hPa to surface and CMDL MBL CO_2, 90°N–90°S. Average differences between RAOB and AIRS temperatures are indicated by solid circles, smoothed differences using a 2-month sliding boxcar average are indicated by the dashed line, and zonally weighted linear least squares fit for the CMDL MBL product are indicated by the solid line.

note that CO_2 variations should correlate with temperature variations in the lower troposphere and surface due to the temperature dependence of photosynthesis; however, we do not expect the precision of temperature retrievals to vary because of these small seasonal changes. In light of these findings, we have advised the AIRS science team that the addition of a CO_2 error term in the temperature retrieval and the incorporation of a realistic CO_2 first guess into the retrieval system are necessary.

[49] Although the bias trends and CO_2 seasonal trends agree very well, the vertical oscillation apparent in all bias plots cannot be removed by incorporating a simple CO_2 first guess into the retrieval system. We are currently investigating possible causes for the vertical oscillation in the bias statistic. Some of the investigations being carried out are (1) incorporation of CO_2 and other trace gas error terms and climatologies in the AIRS initial guess profiles, (2) assessment of the contribution of upper atmospheric state errors and minimization, and (3) deficiencies in the cloud-clearing algorithm, microwave versus IR inconsistencies, etc.

5.3. RMS Difference and Bias for Tighter Collocations and Day/Night Cases

[50] When the collocated sample is chosen on the basis of a tighter control of time and distance collocation (±1 hour and 50 km radius) the number of input samples to the retrieval system shrinks to 33,000. As expected, the percentage of accepted samples by version 4.0 QA flags does not change appreciably. The temperature and water vapor statistics for AIRS_PHYRET show similar trends as seen for the ±3 hours and 100 km radius collocations. A minor improvement of about two tenths of a degree is observed in the temperature RMS. The water vapor RMS shows an improvement of 5–8% in the lower troposphere 2 km layers. There is no appreciable difference in the bias characteristics compared to the biases shown for the ±3 hours and 100 km radius collocations, and hence figures are not shown.

[51] When the collocated samples are separated into day and night cases, the statistics generated for AIRS_PHYRET for different categories did not show appreciable change from the RMS difference and biases for the corresponding categories using all samples. In general, the daytime temperature RMS difference shows an improvement of one tenth of a degree. An improvement of 3–5% in the surface water vapor RMS is observed for nighttime cases. With respect to biases, nighttime cases show larger negative bias in the temperature (about 0.4°K) at the surface. This may be due to difficulties connected with nighttime cloud clearing. This trend tapers down with height, and between 700 hPa and 400 hPa, the daytime cases show a larger positive bias of the order of 0.2°K. The water vapor shows a 5% change in bias in the upper troposphere with the daytime cases showing less negative bias than the nighttime cases indicating that daytime RAOBs are drier than nighttime RAOBs [*Whiteman et al.*, 2006; McMillan et al., submitted manuscript, 2005]. Since the differences are not that prominent, figures are not shown.

6. Summary and Conclusions

[52] The intercomparison of AIRS retrievals with RAOBs, ATOVS retrievals, and forecast profiles from NCEP_GFS and ECMWF reveals the following:

[53] First, the RAOB-AIRS temperature and water vapor retrieval accuracies for clear-only cases over "sea" and "all" categories are close to the expected product goal accuracies, namely, 1° K in 1 km layers for the temperature and better than 15% in 2-km layers for the water vapor in the troposphere. The overall RMS error for the cloud-cleared cases for "sea," and "all" categories is also in close proximity to the expected product goal accuracy except for a slight degradation at the surface. Over land, the retrieval accuracy for the cloud-cleared cases has slightly degraded in comparison to "sea" and "all" categories. The degradation may arise from the daytime convective buildup over the land surfaces, the heterogeneity of the land surface and its accompanied spectral emissivity variations, and the uncertainty in the specification of the surface pressure due to elevation differences over the land surface.

[54] Second, there is a remarkable improvement in RMS accuracy with AIRS retrievals to that of ATOVS retrievals for both the temperature and water vapor. The temperature RMS for the AIRS retrievals shows at least 0.5°K improvement over ATOVS for all the categories ("land/sea/coast," "land," and "sea"). The AIRS water vapor retrievals show a significant improvement compared to the ATOVS. This is probably due to the availability of many water vapor channels with the AIRS instrument in contrast to three HIRS and AMSU-B water vapor channels contributing mainly to the ATOVS retrievals. The AIRS water vapor error also shows improvement over NCEP_GFS, and closely matches with the ECMWF from middle to upper troposphere.

[55] Third, the NCEP_GFS and the ECMWF model forecasts show an RMS difference close to 1°K with the ECMWF faring a little better than the NCEP_GFS. The lower troposphere water vapor RMS difference is 14% for the 2 km layers. This is somewhat expected because the model forecasts weigh heavily the RAOB information in their analysis. In the middle and upper troposphere, the ECMWF water vapor is better than NCEP_GFS by about 10%. As with the satellite retrievals, the forecasts also show a little larger RMS error close to the surface, and especially over land.

[56] Fourth, the AIRS final retrieval for all the accepted cloud-cleared cases shows a larger bias with the RAOB relative to ATOVS, NCEP_GFS, and ECMWF. Clear-only retrievals also show a similar trend in bias, but the magnitude of the bias is significantly lower than the cloud-cleared retrievals, suggesting that some of the bias is caused by the algorithm deficiencies in cloud-clearing methodology or in the simultaneous use of microwave and infrared radiances. In both the cases, the final retrieval step tends to amplify the biases from the fast regression step. The bias for the "all" category is highly influenced by the larger bias contribution from "land" samples, and shows a month-to-month and annual variation that correlates with the CO_2 variations. This has to be addressed more carefully, and if true, may require the need to include CO_2 and other trace gas climatologies in the AIRS initial guess profiles. This may partially mitigate the amplification of bias in the final physical retrieval.

[57] Finally, the global statistics, and the statistics derived for different latitude zones provide a reasonable assessment for the instrument performance, and when the operational

RAOBs are properly quality-controlled, provide realistic results similar to special campaign periods with RAOB ascents synchronized to the Aqua overpass time. The results presented here are similar to the results presented using dedicated RAOB ascents and show a remarkable degree of confidence in the AIRS retrievals despite the problems that are inherent with an ensemble of operational RAOBs.

[58] It should be noted that the results presented in this paper represent the current state of the AIRS version 4.0 retrieval system. We will continue to conduct experiments with future versions of the retrieval system to investigate the source of error and improve the bias characteristics.

[59] **Acknowledgments.** We wish to express our sincere appreciation to the Information Processing Division (IPD) of the Office of Satellite Data Processing and Distribution (OSDPD), NOAA/NESDIS, for providing an access to the RAOB matchup data. We also wish to thank the European Center for Medium Range Forecasting group and the National Center for Environmental Prediction for providing ECMWF and NCEP data used in this paper. Discussions with Tony Reale (NESDIS) have helped to resolve the issues related with the radiosonde matches. Suggestions and comments from anonymous reviewers have helped to improve the scientific discussion, style and organization of the manuscript. Murty Divakarla thanks David Martin of the Space Science and Engineering Center for his comments and suggestions. The funding for this paper has been provided by the NPOESS/IPO grant and is gratefully acknowledged. The contents are solely the opinions of the authors and do not constitute a statement of policy, decision, or position on behalf of the NOAA, NASA, or the U.S. Government.

References

Allegrino, A., A. L. Reale, M. W. Chalfant, and D. Q. Wark (1999), Application of limb adjustment techniques for polar orbiting sounding data, paper presented at 10th International TOVS Study Conference, Int. TOVS Working Group, Boulder, Colo., 27 Jan. to 2 Feb.

Aumann, H. H., et al. (2003), AIRS/AMSU/HSB on the Aqua mission: Design, science objectives, data products, and processing systems, *IEEE Trans. Geosci. Remote Sens.*, *41*(2), 253–264.

Barnett, J. J., and M. Corney (1985), Middle atmosphere reference model from satellite data, in *Handbook for MAP: Middle Atmosphere Program*, vol. 16, edited by K. Labitzke, J. J Barnett, and B. Edwards, pp. 47–85, Sci. Comm. on Sol.-Terr. Phys., Boulder, Colo.

Chahine, M. T. (1982), Remote sensing of cloud parameters, *J. Atmos. Sci.*, *39*, 159–170.

Diebel, D., F. Cayla, and T. Phulpin (1996), IASI mission rationale, and requirements, *Tech. Rep. IA-SM-0000-10-CNE/EUM*, issue 4a, 35 pp., EUMETSAT, Darmstadt, Germany.

Engelen, R. J., G. Stephens, and A. S. Denning (2001), The effect of CO_2 variability on the retrieval of atmospheric temperatures, *Geophys. Res. Lett.*, *28*(17), 3259–3262.

Eskridge, R. E., J. K. Luers, and C. R. Redder (2003), Unexplained discontinuity in the U.S. radiosonde temperature data, part I: Troposphere, *J. Clim.*, *16*, 2385–2395.

Ferguson, M. P., and A. L. Reale (2000), Cloud detection techniques in NESDIS Advanced-TOVS sounding products systems, paper presented at 10th Conference on Satellite Meteorology and Oceanography, Am. Meteorol. Soc., Long Beach, Calif., 9–14 Jan.

Fetzer, E., et al. (2003), AIRS/AMSU/HSB validation, *IEEE Trans. Geosci. Remote Sens.*, *41*(2), 418–431.

Fleming, H. E., M. D. Goldberg, and D. S. Crosby (1988), Operational implementation of the minimum variance simultaneous method, preprints, Third Conference on Satellite Meteorology and Oceanography, Am. Meteorol. Soc., Anaheim, Calif., 1–5 Feb.

GLOBALVIEW-CO2 (2004), Cooperative Atmospheric Data Integration Project: Carbon dioxide [CD-ROM], NOAA Clim. Monit. and Diagn. Lab., Boulder, Colo.

Glumb, R., F. Williams, N. Funk, F. Chateauneuf, A. Roney, and R. Allard (2003), Cross-track infrared sounder (CrIS) development status, *Proc. SPIE Int. Soc. Opt. Eng.*, *5152*, 1–8.

Goldberg, M. D. (1999), Generation of retrieval products from AMSU-A: Methodology and validation, paper presented at 10th International TOVS Study Conference, Int. TOVS Working Group, Boulder, Colo., 27 Jan. to 2 Feb.

Goldberg, M. D., D. S. Crosby, and L. Zhou (1999), Limb adjustments of AMSU-A observations, 10th Conference on Satellite Meteorology and Oceanography, Am. Meteorol. Soc., Long Beach, Calif., 9–14 Jan.

Goldberg, M. D., Y. Qu, L. M. McMillin, W. Wolf, L. Zhou, and M. Divakarla (2003), AIRS near-real-time products and algorithms in support of operational numerical weather prediction, *IEEE Trans. Geosci. Remote Sens.*, *41*(2), 379–389.

Goldberg, M. D., C. D. Barnet, W. Wolf, L. Zhou, and M. Divakarla (2004), Distributed real-time operational products from AIRS, paper presented at SPIE International Symposium on Optical Science and Technology, 49th Annual Meeting, Int. Soc. Opt. Eng., Denver, Colo.

Luers, J. K. (1997), Temperature error of the Vaisala RS90 radiosonde, *J. Atmos. Oceanic Technol.*, *14*, 1520–1532.

McMillin, L. M., J. Zhao, R. V. R. Mundakkara, S. I. Gutman, and J. G. Yoe (2005), Validation of AIRS moisture products using three-way intercomparisons with radiosondes and GPS sensors, paper presented at Ninth Symposium on Integrated Observing and Assimilation Systems for the Atmospheres, Oceans, and Land Surface (IOAS-AOLS), 85th Annual Meeting, Am. Meteorol. Soc., San Diego, Calif., 9–13 Jan.

Miloshevich, L. M., A. Paukkunen, H. Vomel, and S. J. Oltmans (2004), Development and validation of a time-lag correction for Vaisala radiosonde humidity measurements, *J. Atmos. Oceanic Technol.*, *21*, 1305–1327.

Pagano, T. S., H. H. Aumann, D. E. Hagan, and K. Overoye (2003), Prelaunch and in-flight calibration of the Atmospheric Infrared Sounder (AIRS), *IEEE Trans. Geosci. Remote Sens.*, *41*(2), 265–273.

Reale, A. L. (2002), NOAA operational sounding products for advanced-TOVS, *NOAA Tech. Rep. NESDIS 107*, 29 pp., U.S. Dep. of Commer., Washington, D. C.

Reale, A. L. (2003), Scientific status for NOAA operational ATOVS sounding products, paper presented at Thirteenth International TOVS Study Conference, Int. TOVS Working Group, Ste. Adele, Que., Canada, 29 Oct. to 4 Nov.

Reale, A. L. (2005), Satellite coincident reference upper air network, paper presented at CALCON Technical Conference on Characterization and Radiometric Calibration for Remote Sensing, Space Dyn. Lab., Utah State Univ., Logan, 22–25 Aug.

Rosenkranz, P. W. (2003), Rapid radiative transfer model for AMSU/HSB channels, *IEEE Trans. Geosci. Remote Sens.*, *41*(2), 362–368.

Salisbury, J. W., and D. M. D'Aria (1992), Emissivity of terrestrial materials in the 8–14 μm atmospheric window, *Remote Sens. Environ.*, *42*, 83–106.

Salisbury, J. W., and D. M. D'Aria (1994), Emissivity of terrestrial materials in the 3–5 μm atmospheric window, *Remote Sens. Environ.*, *47*, 345–361.

Susskind, J., C. D. Barnet, and J. Blaisdell (2003), Retrieval of atmospheric and surface parameters from AIRS/AMSU/HSB data under cloudy conditions, *IEEE Trans. Geosci. Remote Sens.*, *41*(2), 390–409.

Tilley, F. H., M. E. Pettey, M. P. Ferguson, and A. L Reale (2000), Use of radiosondes in NESDIS Advanced-TOVS sounding products, paper presented at 10th Conference on Satellite Meteorology and Oceanography, Am. Meteorol. Soc., Long Beach, Calif., 9–14 Jan.

Tobin, C. D., H. E. Revercomb, R. O. Knuteson, B. M. Lesht, L. L. Strow, S. E. Hannon, W. F. Feltz, L. A. Moy, E. J. Fetzer, and T. S. Cress (2006), ARM site atmospheric state best estimates for AIRS temperature and water vapor retrieval validation, *J. Geophys. Res.*, *111*, D09S14, doi:10.1029/2005JD006103.

Whiteman, D., et al. (2006), Analysis of Raman lidar and radiosonde measurements from the AWEX-G field campaign and its relation to Aqua validation, *J. Geophys. Res.*, doi:10.1029/2005JD006429, in press.

Zhou, S., L. Zhou, W. Wolf, L. McMillin, C. Barnet, and M. Goldberg (2005), AIRS and IASI local angle correction, OSA/HISE Topical Meeting, Opt. Soc. of Am., Alexandria, Va., 31 Jan. to 3 Feb.

C. D. Barnet, M. D. Goldberg, and L. M. McMillin, Office of Research and Applications, NOAA/NESDIS, 5200 Auth Road, Camp Springs, MD 20746, USA. (chris.barnet@noaa.gov; mitch.goldberg@noaa.gov; larry.mcmillin@noaa.gov)

M. G. Divakarla, STG Inc., 11710 Plaza America Drive, Reston, VA 20190, USA. (murty.divakarla@noaa.gov)

X. Liu, E. Maddy, W. Wolf, and L. Zhou, QSS Group Inc., 4500 Forbes Boulevard, Suite 200, Lanham, MD 20706, USA. (xingpin.liu@noaa.gov; eric.maddy@noaa.gov; walter.wolf@noaa.gov; lihang.zhou@noaa.gov)