AIRS Version 7 Level 2
Performance Test and Validation Report

Edited by:

Qing Yue and Bjorn Lambrigtsen

Contributions by:

John M. Blaisdell, Alireza Farahmand, Eric J. Fetzer, Evan Fishbein, Erica Griffin, Lena Iredell, Fredrick W. Irion, Brian H. Kahn, Peter Kalmus, Evan Manning, Yuliya Marchetti, Thomas Pagano, Nadia Smith, Joel Susskind, Joao Teixeira, Heidar Th. Thrastarson, Tao Wang, Yixin Wen, R. Chris Wilson, Sun Wong

April, 2020

Jet Propulsion Laboratory, California Institute of Technology

Pasadena, CA

Submit Questions to:
https://airs.jpl.nasa.gov/data/support/ask-airs/

Table of Contents

1. Executive Summary 15
2. Introduction 17
3. Summary of Algorithm Improvements in V7 18
3.1 Changes to Stochastic Cloud Clearing Neural Network (SCCNN) 18
3.2 Changes to Surface Classification in the IR-only Algorithm 19
3.3 Changes to Ozone Retrieval Algorithm 19
3.4 Changes to the Water Vapor Retrieval Algorithm 20
3.5 Changes to Temperature and Surface Retrieval Algorithm 20
3.6 Changes to Temperature and Water Vapor Quality Control (QC) 21
4. Details on Major Product Improvements, Changes, and Performance Limitations 22
4.1 Better Agreements between IR-only Surface Classification and NSIDC data 22
4.2 Removal of Spurious Day-time Bias and Day-Night Performance Difference on Water Vapor Retrieval 24
4.3 New Stochastic Cloud Clear Neural Network (SCCNN) for V7 AIRS IR-only Algorithm by Including AMSU-A1 29
4.4 Other Major Improvements in the V7 Retrieval Products 33
4.5 Performance Limitations of Challenging Scenes in V7 33
4.5.1 Larger Surface Temperature Error over Antarctic Sea Ice in V7 SCCNN 34
4.5.2 Water Vapor Profiles Near Surface Over Land 35
4.5.3 V7 IR-only Retrieval Near the Surface Over Frozen Surface Classes 35
4.5.4 Underestimation of Observation Uncertainty by Error Estimate Reported in the AIRS L2 Product 35
4.5.5 Thin and Broken Clouds with Effective Cloud Fraction < 0.05 in the AIRS Cloud Products 36
5. Analyses and Results 37
5.1 Analyses on L2 Temperature (T) and water vapor (Q) vertical profiles 37
5.1.1 Pixel-Scale Evaluations Against ECMWF Model Analyses: Yield, Retrieval Bias, and Sampling Bias Characteristics 37
5.1.2 Pixel-Scale Evaluations Using IGRA Radiosondes 47
5.1.3 Comparisons with Radiosonde Measurements Taken During the MAGIC Field Campaign 52
5.1.4 Pixel-scale Comparisons with Field Campaign Measurements in the Central Arctic . 56
5.1.5 Temperature and Water Vapor Profile Bias and Bias Trends Compared with PREPQC Radiosondes 64
5.2 Analyses on AIRS Cloud Product 70
5.2.1 Granule-scale Level 2 Comparisons of Cloud Products Between V6 and V7 70
5.2.2 Global-scale Monthly Level 3 Comparisons of Cloud Products Between V6 and V7 78
5.3 Comparisons of L2 Near-surface Air Temperature and Humidity with In-situ Measurements 85
5.4 Analyses on AIRS Ozone Vertical Profile and Total Ozone 90
5.4.1 Comparisons of L2 Ozone Vertical Profiles with Ozonesonde Observations 90
5.4.2 Comparisons of AIRS L2 Total Ozone with OMI Retrievals 94

5.4.3 Evaluation of V7 AIRS O_3 Retrieval and Information Content Over Antarctica 95
5.4.4 AIRS L3 Total Ozone Compared to OMPS 105
5.5 Analyses on AIRS L2 Cloud-Clear Radiance (CCRad) 108
5.6 Analyses on AIRS L3 Outgoing Longwave Radiation (OLR) 114
5.7 Analyses on Drought and Flu Season Prediction Applications of AIRS Data 128
5.7.1 Drought Application of AIRS Data 128
5.7.2 Applications of AIRS Data in a Prediction System for Seasonal Influenza 135
6. Averaging Kernels and Other Information Content Metrics for Water Vapor and Temperature Profiles in the AIRS L2 Products 139
6.1 Degree of Freedom and Verticality for Water Vapor Profile Retrievals Shown by AIRS V7 Averaging Kernel Functions 139
6.1.1 Introduction to Averaging Kernel Matrix 139
6.1.2 Averaging Kernel in the AIRS V6 and V7 Retrieval 140
6.1.3 The Change of Averaging Kernels and DOF for Temperature and Water Vapor in AIRS V7 L2 Product 143
6.2 AIRS Temperature and Water Vapor Vertical Structures Examined Using the Empirical Averaging Kernel Method and PREPQC Radiosonde Observations 146
7. References 154

Table of Tables

Table 4.1.1 AIRS Surface classification in V6 and V7 for IR+MW and IR retrievals.....23
Table 5.1.1. Frequency of occurrence for different surface types used to categorize the data. The statistics are based on the 2003 July data from the AIRS V6.6 IR+MW algorithm. 41
Table 5.1.2: The three Central Arctic field campaigns used in this study.....56
Table 5.3.1 Number of matched data-pairs between AIRS and MesoWest and AIRS and ICOADS. 86
Table 5.4.1: Summary of average relative biases and ancillary results for AIRS V6 and V6.5.5 against OMI total ozone columns.....94
Table 5.5.1. Bias, standard deviation, and percent of matchups between -0.5 and 0.5K for cases in January 14, 2003.....109
Table 5.5.2. Differences between the AIRS V655 L2 CCRs for each MODIS channel for cases in January 14, 2003. Both IR+MW and IR only results are shown.....110
Table 5.5.3. Same with Table 5.5.2 but for AIRS V6 L2 CCRs.....110
Table 5.5.4. AIRS V6.55 CCRs vs. MODIS Clear V655 for January 15, 2015. Frac 0 is the percentage of cases with QC equal to zero that had an error greater than 1K. Frac 1 is the is the percentage of cases with QC equal to one that had an error greater then 2.5K. Frac 2_1 is the percentage of cases with QC equal to two that had an error less then 2.5K. The first value is IR+MW and the value in parenthesis is IR.....113
Table 5.5.5 Same with Table 5.5.4 but for AIRS V6. 113
Table 5.6.1. Flux (W/m^2) and Percentage Contributions of Spectral Bands to Clear Sky OLR, OLR, and LWCRF.....128
Table 6.2.1: Numbers of matched temperature and water vapor profiles in the polar, mid-, and tropic latitudes, for January 2011.....147

Table of Figures

Figure 1.1. Comparison of AIRS V6 and V7 temperature (left) and water vapor (right) profiles with collocated radiosonde measurements during the Marine ARM GPCI Investigation of Clouds (MAGIC) campaign. Both bias (solid) and RMSE (dash) are shown with different colors corresponding with various AIRS algorithms: light blue for AIRS V6 IR-only, dark blue for AIRS V6 IR+MW, orange for AIRS V7 IR-only, and red for AIRS V7 IR+MW retrievals, respectively. The MAGIC campaign (9/2012–10/2013) included 19 round trips between Los Angeles and Honolulu. 550 vertical profiles for each variable are collocated with AIRS and included here. V7 products show a decreased mean bias and RMSE than V6. 15
Figure 3.1. Incremental changes in the AIRS retrieval algorithms from V6 to V7. 18
Figure 4.1.1. Daily mean surface classifications from AIRS retrievals for ascending node and the daily data from NSDIC. Gray color (8) means no data. 24
Figure 4.1.2. Daily mean surface classifications from AIRS retrievals for descending node and the daily data from NSIDC. Gray color (8) means no data. 24
Figure 4.2.1. The retrieval yield of January 2013 AIRS total precipitable water vapor from V7 IR+MW, V6 IR+MW, V7 IR-only, and V6 IR-only products. V7 is labeled as V6.6 in the panels. Results are shown using the ascending (Asc: left) and descending (Dsc: middle) passes, separately, with the difference between the two passes shown in the right column. 26
Figure 4.2.2. Differences of retrieval yield for AIRS total precipitable water vapor between V6.6 IR+MW and V6 IR+MW (a), V6.6 IR-only and V6 IR-only (b), and V6.6 IR-only and V6.6 IR+MW (c). Ascending (left) and descending (right) passesare shown separately. Data from January 2003 is used. Note that V7 is labeled as V6.6 in the figure. 27
Figure 4.3.1: Improvements in air temperature at 850 mb for newer IR-only version of V7 candidate (v6.54 NN) versus previous v7 candidate (v6.46 NN) and v6 NN 30
Figure 4.3.2: Improvements in global RMS temperature error performance for newer IR-only version of V7 candidate (v6.54 NN) versus previous v7 candidate (v6.46 NN) and v6 NN.. 30
Figure 4.3.3: Global temperature RMS error vs ECMWF for newer AIRS+A1version of V7 candidate (v6.54 NN) versus v6 AIRS+AMSU product. 31
Figure 4.3.4: Global temperature bias error vs ECMWF for newer AIRS+A1version of V7 candidate (v6.54 NN) versus v6 AIRS+AMSU product. 31
Figure 4.3.5: Global water vapor %RMS error vs ECMWF for newer AIRS+A1version of V7 candidate (v6.54 NN) versus v6 AIRS+AMSU product 32
Figure 4.3.6: Global water vapor %bias error vs ECMWF for newer AIRS+A1version of V7 candidate (v6.54 NN) versus v6 AIRS+AMSU product 32
Figure 4.3.7: Improvements in PBL temperature bias versus sondes for 01/15/15, a challenging test..... 33
Figure. 4.3.8 The surface temperature difference between AIRS SCCNN and ECMWF on July 1, 2015. Two operational versions are shown on the left and middle panels for V6 IR-only and V7 IR-only, respectively. The right panel shows the result with proposed improvements on SCCNN..... 34

Figure 5.1.1. Retrieval yield over non-frozen and frozen ocean and land surfaces as a function of effective cloud fraction (ECF) bins. ECF from AIRS V6.6 IR+MW retrieval is used. In each plate, the top panels are for QC=0 and 1, and the bottom panels are for QC=0. 40
Figure 5.1.2. Spatial distributions of retrieval yield for the four AIRS algorithms at four pressure levels: 100, 500, 700, 850 hPa. Maps are produced using data from January 2003. ECF (%) from V6.6 IR+MW retrieval is also shown. 41
Figure 5.1.3. The observation bias (first row), NN bias (second row), and the sampling bias (third row) for temperature and humidity profiles over non-frozen ocean. Each column represents results from different retrievals. Results using data in January 2003 are on the left and July 2003 data on the right. The top plate is for temperature (K) and the bottom plate is for water vapor (%). The water vapor layer quantities in the L2 product are converted to level specific humidity by simple log(pressure) interpolation. Differences in humidity are calculated as mean(satellite – ECMWF)/mean(ECMWF)x100%.....42
Figure 5.1.4. Similar to Figure 5.1.3, but showing results over non-frozen land instead. 43
Figure 5.1.5. Similar to Figure 5.1.3, but showing results over sea ice (frozen ocean) surfaces instead. 43
Figure 5.1.6. Similar to Figure 5.1.3, but showing results over frozen land surfaces instead. 44
Figure 5.1.7. The RMS error of AIRS temperature and humidity profiles in January 2003 from different AIRS retrievals over non-frozen ocean. Results for final retrieval is shown in the first row and NN (initial guess) in the second row. Each column represents results from different retrievals. The top plate is for temperature (K) and the bottom plate is for water vapor (%). The water vapor layer quantities in the L2 product are converted to level specific humidity by simple log(pressure) interpolation. RMS in humidity are calculated as RMS(satellite – ECMWF)/mean(ECMWF)x100%. 44
Figure 5.1.8. Similar to Figure 5.1.7, but showing results over non-frozen land surfaces instead. 45
Figure 5.1.9. Similar to Figure 5.1.7, but showing results over sea ice surfaces instead.....45
Figure 5.1.10. Similar to Figure 5.1.7, but showing results over frozen land surfaces instead. ...46
Figure 5.1.11. Number of pixels in the Arctic (poleward of 60N) binned by V6 IR+MW AIRS ECF data (A) and cloud fraction by CloudSat and CALIPSO 2B-CLDCLASS-Lidar products collocated with AIRS footprints (CCL: B). Results are generated using data from 01/2006 and 12/2010. The three bins used for the CCL cloud amounts are defined as below: clear if >90% of CCL pixels within the collocated AIRS footprint are clear; partly cloudy if 10~90% of CCL pixels within the collocated AIRS footprint are cloudy; Overcast if >90% of C-C pixels within the collocated AIRS footprint are cloudy. Results over nonfrozen ocean (open ocean) and sea ice are shown separately with different colors. (C) Same as (A) except that only AIRS FOVs corresponding to CCL FOV clear-sky groups are included. (D) Same as (C) expect for the CCL FOV group being single layer overcast conditions by CCL. More details see Peterson et al. (2020). 46
Figure 5.1.12 Global distribution of the IGRA radiosonde launch sites. The black box encircles the area influenced by the 2003 July heat wave, a focus of the study in this section. 48
Figure 5.1.13 Specific humidity (q, in g/kg) profiles averaged over European continent (10°-40°E, 30°-50°N) for July of 2003 (top) and 2011 (bottom). The profiles are plotted in different ranges of near-surface temperatures as shown in the titles of each panel. The

profiles from IGRA radiosonde measurements are in black, V6 retrievals in turquoise, V6IR retrievals in yellow, V6.X retrievals in blue, and V6.XIR in red. The corresponding 1-σ variations are in dashed lines. Numbers on the left axes are sample sizes. The V7 retrievals with AMSU microwave channels are named V6.X, while the one with IR-only channels are named V6.XIR. 49
Figure 5.1.14 Specific humidity (q, in g/kg) RMSE profiles averaged over European continent (10°-40°E, 30°-50°N) for July of 2003 (top) and 2011 (bottom). The profiles are plotted in different ranges of near-surface temperatures as shown in the titles of each panel. The profiles for V6 retrievals in black, V6IR retrievals in yellow, V6.X retrievals in blue, and V6.XIR in red. Numbers on the left axes are sample sizes. The V7 retrievals with AMSU microwave channels are named V6.X, while the one with IR-only channels are named V6.XIR. 50
Figure 5.1.15 For the near-surface T in 300-310 K, q RMSEs from different versions (different columns) of retrievals (red) and NN (blue). The dashed lines are the error estimates reported in the AIRS archive. The top panel is for July of 2003 and the bottom for July of 2011. Numbers on the left axes are sample sizes. The V7 retrievals with AMSU microwave channels are named V6.X, while the one with IR-only channels are named V6.XIR. 50
Figure 5.1.16 Similar to Fig. 5.1.13, but for T profiles (in K). 51
Figure 5.1.17 Biases (top) and RMSEs (bottom) of T retrievals binned according to ranges of near-surface T (shown in the title of each panel). The profiles for V6 retrievals in black, V6IR retrievals in yellow, V6.X retrievals in blue, and V6.XIR in red. Numbers on the left axes are sample sizes. The V7 retrievals with AMSU microwave channels are named V6.X, while the one with IR-only channels are named V6.XIR. 51
Figure 5.1.18. MAGIC data used in this analysis. Background contours are MODIS mean cloud fraction. 53
Figure 5.1.19. Comparisons with MAGIC sondes and ECMWF (Fig. 5.1.18 for location) for temperature bias (a) and RMSE (c); and water vapor bias (b) and RMSE (d), **for AIRS+MW retrievals**. Black lines are for ECMWF reanalysis. Blue lines are for V6 (solid) and V6.55 (dotted); red lines are for v6.55 SCCNN. Note that absolute difference in K and g/kg is shown. 54
Figure 5.1.20. Comparisons with MAGIC sondes and ECMWF (Fig. 5.1.18 for location) for temperature bias (a) and RMSE (c); and water vapor bias (b) and RMSE (d), **for AIRS only retrievals**. Black lines are for ECMWF reanalysis. Blue lines are for V6 (solid) and V6.55 (dotted); red lines are for v6.55 SCCNN. Note that absolute difference in K and g/kg is shown. 54
Figure 5.1.21. Comparisons with MAGIC sondes (Fig. 5.1.18 for location) for q bias for (a) v6 AIRS+MW; (b) v6.55 AIRS+MW (c) v6 AIRS only; (d) v6.55 AIRS only. Black lines show boundary layer top estimated from radiosonde temperature inversions. Note that absolute difference in g/kg is shown.....55
Figure 5.1.22. Comparisons with MAGIC sondes (Fig. 5.1.18 for location) for T bias for (a) v6 AIRS+MW; (b) v6.55 AIRS+MW (c) v6 AIRS only; (d) v6.55 AIRS only. Black lines show boundary layer top estimated from radiosonde temperature inversions. Note that absolute difference in K is shown.....55

Figure 5.1.23. Comparisons with MAGIC sondes (Fig. 5.1.18 for location) for T and q RMSE for the AIRS-only product. (a) v6 q RMSE; (b) v6.55 q RMSE (c) v6 T RMSE; (d) v6.55 T RMSE. Black lines show boundary layer top estimated from radiosonde temperature inversions. Note that absolute difference in g/kg or K is shown. 56
Figure 5.1.24: the location of the collocated sondes during the three field campaigns (indicated by different symbols) specified by classification indicators: the surface class (a), month when the observation was conducted (b), and the effective cloud fraction (ECF) from AIRS (c). The histogram shows the number of observations by different values of classification indicators. 58
Figure 5.1.25. Yield (a), bias (b and d), and RMS (c and e) for temperature (T) and humidity (Q) profiles of various versions of AIRS products and ECMWF model analyses as indicated by different colors. ErrEst stands for the reported error estimate in the AIRS L2 products. The solid and dashed lines in Panel a) correspond with QC = 0 or 1 and QC=0. They correspond with biases of AIRS retrieval and AIRS SCCNN in Panels b and d. The sold, dashed, and dotted lines in Panels c and e correspond with RMS of AIRS retrieval and AIRS SCCNN, and ErrEst reported in the AIRS L2 products. 59
Figure 5.1.26 Retrieval yield profiles as a function of AIRS ECF for various versions of AIRS products. Top row shows results for QC=0 & 1 and bottom shows QC=0. 60
Figure 5.1.27. Similar to Figure 5.1.25, except by five ECF bins: 0~0.2, 0.2~0.4, 0.4~0.6, 0.6~0.8, 0.8~0.9. 61
Figure 5.1.28. Retrieval yield profiles as a function of month for various versions of AIRS products. Top row shows results for QC=0 & 1 and bottom shows QC=0. Data are grouped into three bins: January-February-March (JFM), April-May-June (AMJ), and July-August-September (JAS). Results are presented as a function of the center point of each bin. 62
Figure 5.1.29. Similar to Figure 5.1.25, except by month of observation taken. Data are divided into three bins: January-February-March (JFM), April-May-June (AMJ), and July-August-September (JAS). 63
Figure 5.1.30: Number of AIRS retrievals successfully matched to radiosonde temperature observations by radiosonde launch site, 2004 – 2016 inclusive.....66
Figure 5.1.31: Averaged AIRS V6 and V6.5.5 temperature bias drift against coincident radiosondes as a function of pressure for AIRS *a priori* profiles (left panels) and final retrievals (right panels). Rows show binning of results by latitude bands, 60°N – 90°N, 30°N – 60°N, and 15°N – 30°N.....67
Figure 5.1.32: Averaged AIRS V6 and V6.5.5 temperature bias drift against coincident radiosondes as a function of pressure for AIRS *a priori* profiles (left panels) and final retrievals (right panels). Rows show binning of results by latitude bands, 15°S – 15°N, 30°S – 15°S, and 60°S – 30°S. 68
Figure 5.1.33: Averaged AIRS V6 and V6.5.5 temperature bias drift against coincident radiosondes as a function of pressure for AIRS a priori profiles (left panels) and final retrievals (right panels) for latitude band, 90°S – 60°S. 69
Figure 5.1.34: Map of number of AIRS observations matched to Vaisala RS92 or RS41 radiosondes as used in the determining AIRS water vapor bias trend, 2004 through 2016. (See text for selection criteria.) 69

Figure 5.1.35, left panel: Trend in relative bias of AIRS *a priori* water vapor vs coincident radiosondes. Right panel: Trend in relative bias of AIRS final retrieval. Points are averages of individual monthly trends, with standard deviations of such monthly trends shown as error bars (see text). For clarity, only every fourth error bar is shown. 70
Figure 5.2.1 Upper row: cloud_phase_3x3 for v6 (left), v6.55 (middle), and brightness temperature at 1231 cm^{-1}. Middle row: upper layer TCldTop for v6 (left) and v6.55 (middle) with the difference of v6-v6.55 (right). Bottom row: lower layer of TCldTop for v6 (left) and v6.55 (middle) with the difference of v6-v6.55 (right). The averages and standard deviations of CTT (K) for the full granule are reported below each sub-panel in the left and center columns, while the differences between v6-v655 average and standard deviations of CTT (K) are reported below each sub-panel in the right column. 73
Figure 5.2.2. Same as Figure 5.2.1 except for all clouds with ECF < 0.05 filtered out. The upper row is a repeat of the upper row in Figure 5.2.1. 74
Figure 5.2.3. Upper row: total effective cloud fraction (ECF) which is a sum of the two layers of ECF (if a second layer exists within a given AIRS pixel), shown are v6 (left), v6.55 (middle), and v6-v655 (right). Middle row: upper layer ECF for v6 (left), v6.55 (middle), and v6-v6.55 (right). Bottom row: lower layer of ECF for v6 (left), v6.55 (middle), and v6-v6.55 (right). The averages and standard deviations of ECF for the full granule are reported below each sub-panel. 75
Figure 5.2.4. Same as Figure 5.2.3 except for all clouds with ECF < 0.05 filtered out. 76
Figure 5.2.5 Three more example granules on July 1, 2011 for (i) granule 004 (upper row), (ii) granule 005 (middle row), and (iii) granule 008 (lower row) for v6 upper level CTT (left column), v655 upper level CTT (center column), and the 1231 cm^{-1} Tb (right column). As previous figures show, the average and standard deviations for the full granule are reported for upper level CTT. All values of ECF are shown with no filtering in order to illustrate the problematic cloud retrievals. If a filter with ECF < 0.05 is applied (not shown), these issues are largely eliminated. 77
Figure 5.2.6. Left column v6, middle column v6.55 and right column v6.55 minus v6 for total CloudFrc_A (top row), CloudFrc_D (second from top row), layer weighted CldTopTemp_A (second from bottom row) and CldTopTemp_D (bottom row). 81
Figure 5.2.7. FineCloudFrc for v6 (left column), v6.55 (middle column), and v6 minus v6.55 (right column). The upper two rows are for 346 hPa, the middle two rows for 648 hPa, and the lower two rows for 887 hPa. Within each pair, the upper (lower) row represents the ascending (descending) orbit. 82
Figure 5.2.8. Cloud_phase_3x3_A for v6 (left column), v6.55 (middle column), and v6 minus v6.55 (right column). From top to bottom is the phase tests -2 and -1 (liquid), 0 (unknown), and +1, +2, +3, and +4 (ice). 83
Figure 5.2.9 Cloud_phase_3x3_D for v6 (left column), v6.55 (middle column), and v6 minus v6.55 (right column). From top to bottom is the phase tests -2 and -1 (liquid), 0 (unknown), and +1, +2, +3, and +4 (ice) 84
Fig. 5.3.1. Number of reports from (a) MesoWest and (b) ICOADS in July 2011. The number is in dB scale in each grid. 85
Fig. 5.3.2. Scatterplots with colored data density of near surface air temperature (left) IR-only and (right) IR+MW. 87

Fig. 5.3.3. Scatterplots with colored data density of near surface relative humidity from MesoWest and specific humidity from ICOADS (left) IR-only and (right) IR+MW. 88
Fig. 5.3.4 The mean bias (left) and RMSE (right) with the collocated (a) MesoWest and (b) ICOADS near surface air temperature measurements (unit in K) by effective could fraction (ECF) from AIRS. Different colors indicate different AIRS retrievals: blue for V6 IR only, red for V6 IR+MW, yellow for V655 IR only, and purple for V655 IR+MW. 89
Fig. 5.3.5 The relative bias (left) and RMSE (right) with the collocated (a) MesoWest near surface relative humidity (%) and (b) ICOADS near surface specific humidity (g/kg) by effective could fraction (ECF) from AIRS. Different colors indicate different AIRS retrievals: blue for V6 IR only, red for V6 IR+MW, yellow for V655 IR only, and purple for V655 IR+MW. 90
Figure 5.4.1: Location and number of ozonesonde launches matched to AIRS V6 observations as used in this study. 92
Figure 5.4.2, Left panel: Median profile bias of V6 AIRS ozone profile retrievals against coincident ozonesondes. Right panel: Same for V6.5.5 AIRS ozone. See text for explanation of data selection and reduction. 93
Figure 5.4.3, Left panel: Median profile bias of V6 AIRS ozone profile retrievals against ozonesondes with AIRS averaging kernel applied (see text for formula). Right panel: Same for V6.5.5 AIRS ozone profiles. See text for explanation of data selection and reduction. 93
Figure 5.4.4. Left panel: AIRS V6-OMI relative ozone column bias for January 1, 2007. Right panel: AIRS V6.5.5-OMI relative ozone column bias. Note that the color scale has been restricted for clarity. 95
Figure 5.4.5, left panel: AIRS V6-OMI relative ozone column bias for July 1, 2007. Right panel: AIRS V6.5.5-OMI relative ozone column bias. Note that the color scale has been restricted for clarity. 95
Figure 5.4.6: AIRS ozone spectra from granule 59, scan-line 59, footprint 14 on 06 July 2015. Green crosses are the channels used in the V7, red circles are used in V6, and green crosses over red circles are used in both. 96
Figure 5.4.7: The upper panel shows the trapezoids and vertices used to represent ozone profile perturbations in V6, while the lower panel shows them for V7. 97
Figure 5.4.8: Comparison of v6 (magenta) and v7 nominal (black) and depleted (green) ozone climatologies. The bottom panel is the latitude-dependence of 50 hPa O_3 during August, the middle panel is the time-dependence at 67˚S (DUS) at 50 hPa, and the upper panel are the profiles at DUS. The red curve is the v7 depleted minus nominal. The red curves in the middle panel show the threshold temperatures for the nominal to depleted climatologies. 98
Figure 5.4.9: Total O_3 was unusual during winter 2015. Frame (a) compares total O_3 maps from 28 August 2007, 2012 and 2015. Frame (b) shows the location of total O_3 stations and frame (c) compares time series of total O_3 at Dumont d'Urville Station versus 20-year median. Figures are from Braathen, G., 2015: *WMO Antarctic Ozone Bulletin* no. 1 and 5, 2015. 99
Figure 5.4.10: Averaging kernel vertices 100
Figure 5.4.11: Comparison of version 6 and 7 ozone averaging kernels in the upper troposphere and lower stratosphere for version 7 (solid lines) and version 6 (dotted lines) for summer

(2015-02-26), winter (2015-07-06) and spring (2015-09-26). Averaging kernels are show for the trapezoids with bottom vertices at 20, 40 70, 100 and 200 hPa.....101
Figure 5.4.12: Comparison of version 6 and 7 ozone profiles for the same footprints in Figure 5.4.11 (left to right, Summer, Winter and Fall). The upper row shows ozonesonde profiles. Black and black diamonds are the ozonesonde and layer-mean ozonesonde profiles. Version 6 IR+MW are red curves, version 6 IR-only are magenta curves, version 7 IR+MW are green curves and version 7 IR-only are cyan curves. Kerned profiles use the same colors, but are shown as dashed curves. The lower row shows differences, solid curves are differenced with kerned ozonesondes; dashed curves are differenced with layer-mean profiles. The line colors are the same as in the upper row. 102
Figure 5.4.13: Comparison of version 6 and 7 total O_3. The upper row shows total O_3 from left to right as histograms of version 6 O_3, version 7 and time series. The lower row are histograms and time series differenced from the SAOZ up-looking UV spectrometer. For the histograms, black curves are total O_3 from the SAOZ up-looking UV spectrometer, the blue curves are the V6 and v7 first guess, red curves are the IR-only retrieved total, magenta curves are the IR+MW total O_3, green curves are the kerned ozonesonde and yellow curves are the "corrected" kerned ozonesonde (see text for details). For the kerned profiles, the IR-only curves are marked with "+", but the differences between IR and IR+MW products are small and the curves over-lay. The mean and standard deviation for each histogram is shown on the upper left of each panel. The right column shows the times series of V6 and V7 total O_3 (upper panel) and differenced with SAOZ total O_3 (lower panel). Version 7 are shown as the green curves, V6 by the red curves, and kerned by the cyan (V7) and magenta (V6). Corrected kerned are marked with diamonds. IR+MW are shown as solid lines, while IR-only are shown as dotted lines. 104
Figure 5.4.14 Daily gridded ozone comparisons for August 6, 2018, early in the formation of the 2018 ozone hole. Note the enormous improvements in the latest version in the low ozone region over the tropical western Pacific Ocean, where the overdamped Version 6 could not reach sufficiently low total ozone values and retrieval improvements have ameliorated this issue. Note also the improvement over northern Africa and the adjacent Atlantic Ocean, where the QC rejection of dusty cases has removed the incorrect ozone values from the Saharan dust outbreak area. These improvements in the tropical areas were necessary for the AIRS total ozone product to be competitive with other data sources. Note also the hint of a forming ozone hole in the OMPS data, with three small areas along the terminator with values below 250 DU. These areas are precisely consistent with the northernmost extent of the low ozone region depicted in the Version 6.55 total ozone product, which for this day is not very different from the Version 6 depiction.....106
Figure 5.4.15. Daily gridded ozone comparisons for August 31, 2018, somewhat later in the formation of the 2018 ozone hole. On this day we see similar improvement as in Figure 5.4.14 for the tropical areas, but here we have a significant change in the ozone hole depiction. The areas measured by OMPS as below 200 DU are matched in the AIRS data, and we can see that these low values do not extend as far as the South Pole. On this day there are significant differences over Antarctica between Version 6 AIRS-Only and Version 6.55 AIRS. Anomalously high retrievals of total ozone over Antarctica in Version 6 have been corrected by a combination of using the new climatology and excluding the strongest

ozone channels from the retrieval in regions where the surface temperature is nearly equal to the stratospheric temperature. 107
Figure 5.4.16. Daily gridded ozone comparisons for October 1, 2018, near the lowest ozone values and greatest spatial extent of the 2018 ozone hole. The vast majority of the points over Antarctica are now within 20 DU of the OMPS values, even in this extreme case, and examination of the V6.55 minus V6 difference plot shows that the retrieval algorithm updates in Version 6.55 made significant improvements in nearly all parts of the globe. Some smaller discrepancies remain in cloudy regions of the Southern Ocean which are particularly challenging for IR retrievals. 108
Figure 5.5.1. The spectral response function for MODIS channels. 109
Figure 5.5.2. Difference between AIRS CCRs and MODIS channel bright ness temeperatures for cases on January 15, 2003 for AIRS V655 (left) and AIRS V6 (right) with different QC flags. 111
Figure 5.5.3. Same with Figure 5.5.2 but for cases on January 15, 2015. 111
Figure 5.5.4. The percent of cases that are QC level 0, 1, and 2 for the AIRS CCRs spectrally convolved to MODIS channel 22. The left panel shows results for AIRS V6.55 and the right panel for AIRS V6. 112
Figure 5.5.5. Same with Figure 5.5.4 but showing results with respect to MODIS Channel 28. 112
Figure 5.5.6. Same with Figure 5.5.4 but showing results with respect to MODIS Channel 33. 112
Figure 5.6.1. Differences between AIRS OLR and CERES OLR for every month of 2018, in W/m^2. 115
Figure 5.6.2. Seasonal variation of OLR in the CERES reference data set. 116
Figure 5.6.3. The OLR differences from CERES EBAF_Ed4.0 for the month of January 2018 are small, and very slightly smaller in Version 6.55 than in Version 6. 117
Figure 5.6.4. The OLR differences from CERES EBAF_Ed4.0 for the month of April 2018 are very slightly smaller in Version 6.55 than in Version 6, with the largest differences along the Antarctic ice edge where the surface class determination has been improved. 118
Figure 5.6.5. The OLR differences from CERES EBAF_Ed4.0 for the month of July 2018 are very slightly smaller in Version 6.55 than in Version 6, with the only significant changes occurring along the Antarctic ice edge. 119
Figure 5.6.6. The OLR differences from CERES EBAF_Ed4.0 for the month of October 2018 are very slightly smaller in Version 6.55 than in Version 6, with the largest differences occurring along ice edges. 120
Figure 5.6.7. Difference of AIRS Clear Sky OLR from CERES Clear Sky OLR for every month of 2018, in W/m2. 121
Figure 5.6.8. Seasonal variation of Clear Sky OLR in the CERES reference data set. 122
Figure 5.6.9. Clear Sky OLR for January 2018. Changes in retrieval methodology and QC for AIRS have resulted in reduced global mean Clear Sky OLR values compared to Version 6, primarily in convective areas and along ice edges. 123
Figure 5.6.10. Clear Sky OLR for April 2018. Changes in retrieval methodology and QC for AIRS have reduced the global mean Clear Sky OLR values compared to Version 6, with the largest changes in convective regions and off the coast of Antarctica. 124
Figure 5.6.11. Clear Sky OLR for July 2018. Changes in retrieval methodology and QC for AIRS have slightly reduced the global mean Clear Sky OLR values compared to Version 6. The

largest differences from Version 6 are in convective regions and off the coast of Antarctica, but there are significant differences in the northern mid-latitude and polar regions as well. 125
Figure 5.6.12. Clear Sky OLR for October 2018. Spatial patterns of differences between Version 6.55 and Version 6 are similar to those in July. 126
Figure 5.6.13. Long-wave Cloud Radiative Forcing obtained by differencing OLR and Clear Sky OLR. 127
Figure 5.7.1: Mesowest number of records for July 2011 129
Figure 5.7.2: L2 relative humidity comparisons with Mesowest for July 2011 130
Figure 5.7.3: L2 temperature comparisons with Mesowest for July 2011 130
Figure 5.7.4: L2 VPD comparisons with Mesowest for July 2011 130
Figure 5.7.5: Relative humidity actual (left) vs anomalies (right) for July 2011. Top: V6, Middle: V6.5.5, Bottom: V6 minus V6.5.5 131
Figure 5.7.6: Comparisons on the monthly V6 vs V6.5.5 relative humidity actual (left), anomaly (middle), and the histogram of relative humidity anomaly differences between V6 and V6.5.5 (right) for July 2011 over Texas. 131
Figure 5.7.7: The time series of RH actual and anomalies during the Texas drought event in 2011. Red: V.6.5.5. Blue: V6. 132
Figure 5.7.8: Temperature actual (left) vs anomalies (right) for July 2011. Top: V6, Middle: V6.5.5, Bottom: V6 minus V6.5.5 132
Figure 5.7.9: Comparisons on the monthly V6 vs V6.5.5 temperature actual (left), anomaly (middle), and the histogram of temperature anomaly differences between V6 and V6.5.5 (right) for July 2011 over Texas. 133
Figure 5.7.10: The time series of T actual and anomalies during the Texas drought event in 2011. Red: V.6.5.5. Blue: V6. 133
Figure 5.7.11: VPD actual (left) vs anomalies (right) for July 2011. Top: V6, Middle: V6.5.5, Bottom: V6 minus V6.5.5 134
Figure 5.7.12: Comparisons on the monthly V6 vs V6.5.5 VPD actual (left), anomaly (middle), and the histogram of VPD anomaly differences between V6 and V6.5.5 (right) for July 2011 over Texas. 134
Figure 5.7.13: The time series of VPD actual and anomalies during the Texas drought event in 2011. Red: V.6.5.5. Blue: V6. 135
Figure 5.7.14: Comparison of time series for Los Angeles 2017-2018 (left) and 2018-2019 (right). Top: Forecasted influenza activity from the seasonal influenza prediction system driven by V6 (red) and V6.55 (blue) near-surface water vapor mass mixing ratio (H2O_MMR_Surf_A). The time series are composed of ensemble means of seven-day forecast values. The green line shows influenza surveillance data for the multi-state region. Middle: Time series of H2O_MMR_Surf_A for V6 (red) and V6.55 (blue) that are used as input for the influenza prediction system. Bottom: Relative difference of H2O_MMR_Surf_A values from V6 and V6.55 at each point in time. The magenta line shows the mean of the relative differences. The number of valid points in the time series for each version is shown in the plot titles. 137
Figure 5.7.15: Comparison of time series for Dallas 2017-2018 (left) and 2018-2019 (right). Top: Forecasted influenza activity from the seasonal influenza prediction system driven by V6

(red) and V6.55 (blue) near-surface water vapor mass mixing ratio (H2O_MMR_Surf_A). The time series are composed of ensemble means of seven-day forecast values. The green line shows influenza surveillance data for the multi-state region. Middle: Time series of H2O_MMR_Surf_A for V6 (red) and V6.55 (blue) that are used as input for the influenza prediction system. Bottom: Relative difference of H2O_MMR_Surf_A values from V6 and V6.55 at each point in time. The magenta line shows the mean of the relative differences. The number of valid points for each version is shown in the plot titles.....138
Figure 6.1.1. Maps of degree of freedom (DOF) calculated from temperature (left) and water vapor (right) averaging kernels in AIRS IR-only L2 retrievals. Data that pass the AIRS quality control (QC=0 or 1) from January 14-16, 2015 is used in the analysis. V6 and V6.55 results are shown in the top two rows. The bottom three tows show results from the three test runs.....145
Figure 6.1.2. Zone mean of degree of freedom (DOF) calculated from temperature (top) and water vapor (bottom) averaging kernels in AIRS L2 retrievals. Data that pass the AIRS quality control (QC=0 or 1)from January 14-16, 2015 is used in the analysis. Dashed lines show IR+MW results and solid lines show IR only results. Colors correspond with different versions of AIRS retrievals: black for V6, blue for V6.55, and green, orange, and red for three different test cases. 146
Fig. 6.2.1: Examples of matched radiosonde observations and AIRS retrievals for temperature in the polar, mid-, and tropic latitudes, for January 2011. Color scale indicates temperature in Kelvin. 148
Fig. 6.2.2. Degree of freedom for temperature (upper panel) and water vapor (low panel) in the tropic (green lines), mid- (blue lines), and polar (red lines) latitudes. Dashed lines represent data for V655 and solid lines represent data for V6. 149
Fig. 6.2.3. EAKs for temperature in the polar, mid-, and tropic latitude regions in January 2011, comparing V6 (left column) and V655 (right column) for rank 12. Empirical averaging kernel profiles with darker grey-colored lines corresponding with lower pressure. 150
Fig. 6.2.4. Similar to Fig. 6.2.2, but for water vapor. 151
Fig. 6.2.5 Row sums of EAKs for temperature at rank 12 (left column) and for water vapor at rank 7 (right column). Blue lines represent data for V6 and red lines represent data for V655.....152

1. Executive Summary

This report characterizes the differences between Version 6 (V6) and Version 7 (V7) of various AIRS core data products and includes comparisons to other well validated data sources such as radiosondes, surface station, and satellite measurements. The objective was to determine whether the core products in the AIRS V7 **Level 2** retrievals are within reasonable range of well-established reference data sources and to measure its performance against the previous versions. Limited results for Level 3 data products are presented as well. A complete review of Level 3 data products can be found in the AIRS V7 Level 3 User Guide (Tian et al. 2020).

This document reports on the analyses performed to detect errors during the development of V7 and to characterize the performance of the final products. An example is shown in Figure 1.1, which shows the reduction of mean bias and root mean square error (RMSE) in V7 temperature and water vapor profiles, particularly in the boundary layer, compared to those in V6 as evaluated against the collocated radiosonde measurements during the Marine ARM GPCI Investigation of Clouds (MAGIC) field campaign.

Major algorithm changes between versions are discussed in Section 2 and include:

- an improved Stochastic Cloud Clearing Neural Network (SCCNN) used as a first guess;
- removal of ambiguity in surface classification in the infrared-only (IR-only) retrieval algorithm;
- algorithm improvements that lead to better ozone, water vapor, and temperature products.

For a detailed description of the V7 algorithm, please refer to the V7 Algorithm Theoretical Basis Document (Fetzer et al. 2020). All AIRS documents can be obtained at the NASA Goddard Earth Science Data and Information Services Center (GES DISC) at https://disc.gsfc.nasa.gov/information/documents?title=AIRS%20Documentation.

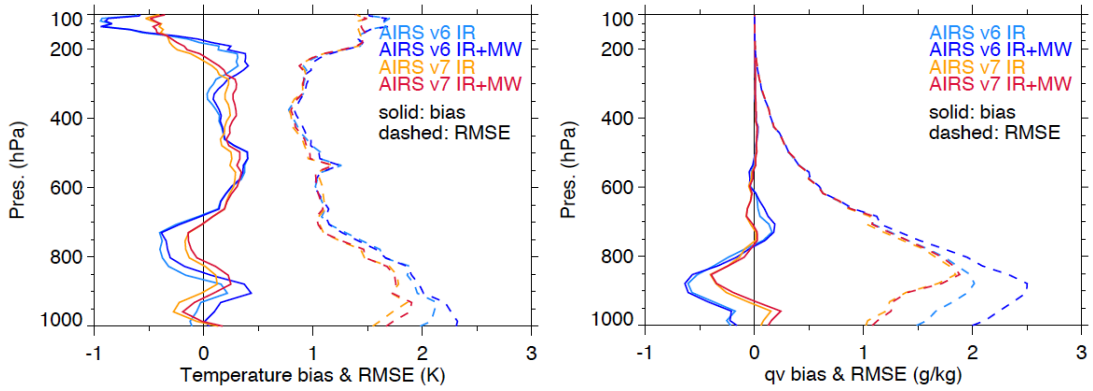


Figure 1.1. Comparison of AIRS V6 and V7 temperature (left) and water vapor (right) profiles with collocated radiosonde measurements during the Marine ARM GPCI Investigation of Clouds (MAGIC) campaign. Both bias (solid) and RMSE (dash) are shown with different colors corresponding with various AIRS algorithms: light blue for AIRS V6 IR-only, dark blue for AIRS V6 IR+MW, orange for AIRS V7 IR-only, and red for AIRS V7 IR+MW retrievals, respectively. The MAGIC campaign (9/2012–10/2013) included 19 round trips between Los Angeles and Honolulu. 550 vertical profiles for each variable are collocated with AIRS and included here. V7 products show a decreased mean bias and RMSE than V6.

The results in this report are organized into four main sections:

Section 3 summarizes the *algorithm* development and improvements;
Section 4 summarizes the *product* improvements as well as performance limitations;
Section 5 provides detailed *analyses* of AIRS core retrieval products;
Section 6 provides a discussion of *averaging kernels* and related variables for water vapor and temperature profiles.

Highlights of Version 7 are as follows.

- Analysis of AIRS V6 water vapor products after its release revealed spurious bias in day-time products and large difference in the retrieval performance between day- and night-time cases. That bias has been removed in V7 as a result of removal of shortwave (2607~2657 cm^{-1}) channels used in the water vapor retrieval of both profiles and column integrals.
- With the loss of AMSU-A2 and several AMSU-A1 channels, one of the goals with V7 has been to improve the quality of infrared-only algorithm, which continues to provide high quality global retrievals with reduced dependence on microwave observations. The current V7 infrared-only product uses a SCCNN first guess that utilizes stable microwave channels of AMSU-A1, while its physical retrieval step only uses infrared channels and does not use microwave information. Plans are in place to transition to a pure infrared-only SCCNN first guess if there is further loss of microwave channels in the future.
- The ambiguity in differentiating frozen and non-frozen surface classification in the IR-only algorithm has been removed by using ice concentration over ocean and water equivalent accumulated snow depth over land obtained from the ancillary Global Forecasting System (GFS) files. We note that until the loss of AMSU-A2 in 2016, this information was derived from the microwave observations.
- Comparisons of temperature and water vapor vertical profiles with radiosonde and reanalysis data show an improved performance over all in V7.
- For ozone, V7 includes more channels sensitive to O_3 in the upper troposphere and lower stratosphere in the 9.6μm O_3 band and uses additional vertical basis functions to better represent the profile. The ozone retrieval also benefits from improved polar stratospheric temperatures in the V7 SCCNN and better quality control of the retrievals. Comparisons with OMPS and OMI both indicate an improved column ozone retrieval in V7.

Section 4.5 discusses the performance limitations in the V7 retrievals that have been identified. These issues, which affect only a limited set of particular regions or conditions, have no discernible impact on global statistics. We note that there are larger errors in V7 SCCNN surface skin temperature over the Antarctic sea ice, the challenges of near surface profile retrievals over snow and ice covered surfaces, and ambiguity in cloud retrievals when effective cloud fraction is less than 5%, all of which are highly regional in nature. A summary on how the current algorithm (both V6 and V7) calculates the temperature and water vapor averaging kernel matrices and error estimate are presented in Section 6. *Users who conduct science investigations with AIRS data under these conditions may wish to do a more thorough assessment of those cases than is warranted here.*

2. Introduction

Since the release of AIRS V6, significant improvements and modifications have been applied to the AIRS retrieval algorithm, especially the IR-only, which continues to provide high quality observations and retrievals after the complete loss of AMSU-A2 on September 24, 2016 (See AIRS V6 Test Report Supplement: Performance of AIRS+AMSU vs. AIRS-only retrievals Analysis of the impact of the loss of AMSU-A2). As a result of these algorithm changes, the V7 retrieval products are different from those of V6. The purpose of this report is to document these changes and the differences in the retrievals between the two versions. Both the IR+MW (AIRS/AMSU infrared and microwave combined) and IR-only (AIRS-only) retrievals are examined. A brief summary of major algorithm improvements is given in Section 3. Major changes and improvements as well as several known performance limitations in the V7 retrieval products are summarized in Section 4, followed by the detailed analysis results for a set of AIRS core products (primarily Level 2 but also Level 3) in Section 5. Section 6 presents a summary on the averaging kernels reported in both V6 and V7 Level 2 products and the difference between the temperature and water vapor profile averaging kernels between the versions, as well as an application of empirical averaging kernels using AIRS retrievals.

Conventions of version numbers used in this report (see Figure 3.1):

- ***The studied version of the product is V6.55 (later numbered as V6.6), which has become the AIRS official V7 data release.***
- ***AIRS/AMSU infrared and microwave combined retrieval is denoted as "IR+MW".***
- ***AIRS-only retrieval is denoted as "IR-only".***

The AIRS product files contain a large number of parameters, ranging from quality flags to retrieved data values. This report provides analyses on the following categories of core retrieval variables:

- Thermodynamic quantities (profiles of temperature and humidity, near surface properties, and cloud parameters)
- Cloud cleared radiances
- Trace gases (O_3)
- Surface types
- Outgoing longwave radiation (OLR).

Complete lists of all parameters relevant for Level 2 can be found in the V7 Level 2 Product User Guide (Thrastarson et al. 2020), and for Level 3 in the V7 Level 3 Product User Guide (Tian et al. 2020). The reader is referred to the table of contents in this report for the parameters and their processing levels included in these analyses.

Two examples are given in Sec. 5.7 to illustrate the changes for two applications based on AIRS data: the AIRS drought detection application and the flu season prediction application. *Analysis results on ozone retrievals are presented in Sec. 5.4, but most of the composition products as well as the dust flag are covered in a supplemental report to be released at a later date.*

3. Summary of Algorithm Improvements in V7

Based on the evaluation of the performance of V6, numerous scientific modifications have been made to the retrieval algorithm for V7 to improve the retrieval quality. Figure 3.1 summarizes the major changes implemented at the various stages of the development, which are elaborated below..

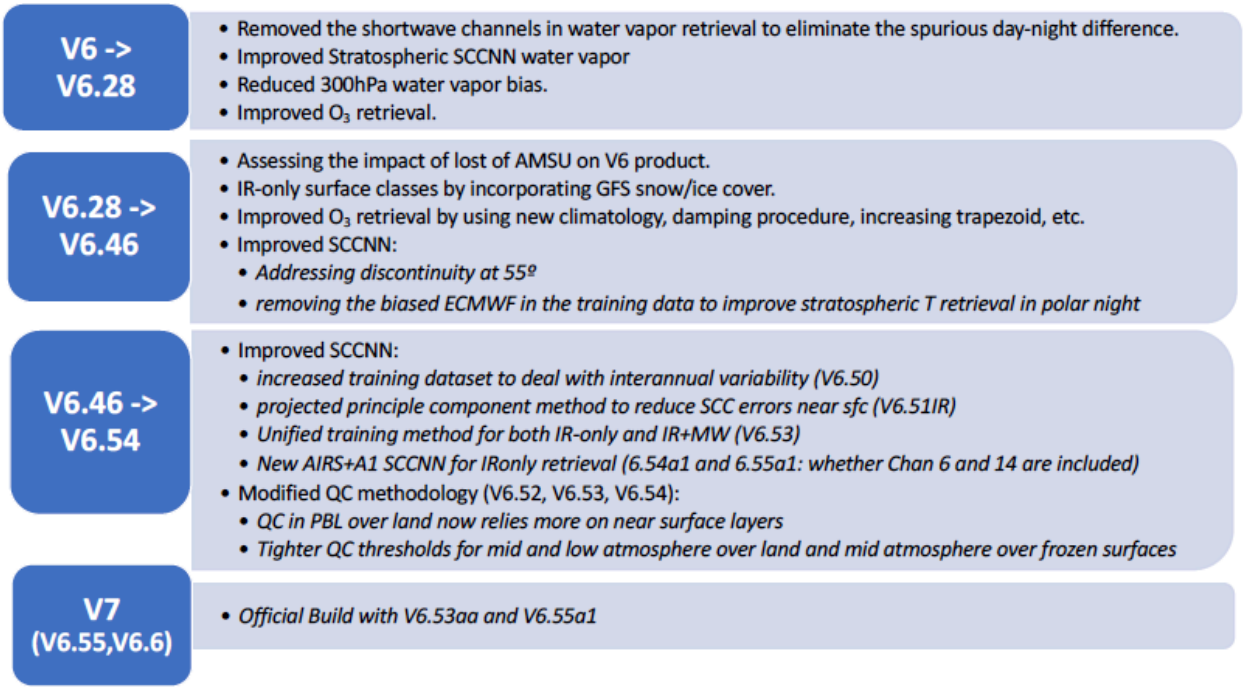


Figure 3.1. Incremental changes in the AIRS retrieval algorithms from V6 to V7.

3.1 Changes to Stochastic Cloud Clearing Neural Network (SCCNN)

The SCCNN algorithm is a statistical method for retrieval of temperature and water vapor profiles as well as surface parameters using radiance observations from AIRS and AMSU. The algorithm consists of two key parts: SCC (Stochastic Cloud Clearing) estimates the cloud-cleared infrared spectrum using series of linear and nonlinear operations on AIRS/AMSU radiances from a 3×3 "golf ball" footprint. The NN (neural network) estimates temperature and water vapor profile from projected principal components of the cloud-cleared spectrum. SCCNN surface skin temperature, temperature and water vapor profiles are used as first guess in V6 and V7 AIRS retrievals and are also provided in the AIRS Level 2 Support product. The NN algorithm is a regression, trained once, then applied over the entire mission. Its products are generated in near-real-time. Because it is trained once and for all, the NN first guess is expected to behave similarly throughout the whole mission, while model-based first guesses can show significant change in bias structure over mission duration due to model changes.

A new SCCNN algorithm version was developed and delivered as the AIRS V7 IR+MW and IR-only algorithm first guesses. The new SCCNN algorithm has been trained on newer, more diverse training sets. A series of versions with incremental changes during the V7 development are summarized as the following:

- New coefficients trained on 2013-era ECMWF (v6.46)

  - Addressed previously noted issues from prior V7 candidates (V6.28) related to training set interpolation artifacts, discontinuities in sea ice flagging, and training set endpoints that truncated a portion of the summer.
- Increased training set comprehensiveness to capture a wide range of variability (v6.50)
  - Now trained using data from 2013, 2010, and 2005.
- Improved principal component representation for SCC algorithm, which improved cloud clearing results in IR-only algorithm (v6.51).
- Improved SCCNN for V7 AIRS IR-only algorithm by including AMSU-A1 (v6.54/v6.55):
  - 6.5.5 SCCNN, for AIRS+AMSU-A1-only, uses AMSU channels 3, 8-13, 15. (V6 SCCNN, in comparison, used channels 1-3, 6, 8-15.), which provides an option for addressing the degradation and failure of AMSU+A2.
  - Removed flagged invalid channels from IR channel list used in SCCNN and replaced with other, valid channels that sound near the PBL.
- Updated training set screening in poles (v6.54/v6.55) which causes modest polar improvements versus v6.51 overall.

Section 4.3 and the analyses presented in Section 5 show the comparisons between V6 and V7 SCCNN results and comparisons with reference datasets.

3.2 Changes to Surface Classification in the IR-only Algorithm

In the AIRS-Only system, V6 uses solely the criteria of whether surface temperature from SCCNN is above or below the freezing point to determine whether the surface is frozen or not, which is ambiguous near the freezing point and causes reduced retrieval performance (Yue and Lambrigtsen, 2017). The V7 IR-only algorithm determines the surface classification using the ice concentration over ocean and water equivalent accumulated snow depth over land from the Global Forecasting System (GFS). Section 4.1 shows the details of GFS ancillary surface conditions and the comparison with the data from National Snow and Ice Data Center (NSIDC).

3.3 Changes to Ozone Retrieval Algorithm

A new climatological first guess was developed which separates ozone hole cases from other cases to yield two distinct profile shapes over Antarctica during spring. On a profile-by-profile basis, the 50hPa temperature given by the SCCNN is used to select the profile shapes to use as first guess in the retrieval process. The final retrieval shows improved ozone profiles. The ozone first guess is now included in the V7 L2 product.

Outside the ozone hole the ozone first guess profile is adjusted to the tropopause given by the SCCNN. A shape-preserving transformation is made to align the climatological tropopause to the case-by-case tropopause in the ozone first guess, with no adjustment applied above 1hPa. The ozone first guess is now included in the V7 L2 product.

More trapezoids are added to allow better representation of the vertical structure where the information is available. The damping was decreased everywhere and was made even less during summer mid-latitude and polar months, recognizing that very large excursions from the first guess are normal in propagating waves during those months.

Many of the strongest ozone lines were added to the retrieval for V7 except in the relatively few very cold cases where there is little contrast between the stratosphere and the surface, channels between 1031 cm^{-1} and 1061 cm^{-1} are not used. In all other spatial regions these channels increase the degrees of freedom over V6 and improve the total ozone retrieval.

Two additional spectral hinge points in the 9.6 μm region were added to the surface retrieval (along with channels in this region), and their emissivity values are also updated in the ozone retrieval step.

The surface emissivity noise covariance contribution was removed from ozone retrieval, since emissivity is now part of the solution space. The water noise covariance contribution was also removed because it was found that this reduced the overdamping present in all previous versions.

Quality control specific to ozone was added. The ozone retrieval is rejected under the following conditions: the attempted change on the first iteration is too large indicating the temperature or water profiles may be in error, radiative closure is not achieved in the ozone channels, the dust test indicates the presence of significant dust, and/or the change in emissivity from the first guess is very different in the ozone spectral region and the adjacent spectral regions.

These algorithm improvements lead to an increase of information content and vertical resolution of the ozone product (Section 5.4.3). Together with improved ozone quality control, these changes enabled a much better match to OMI and OMPS data (Section 5.4.2 and 5.4.4).

3.4 Changes to the Water Vapor Retrieval Algorithm

In V7, the shortwave channels in the range of 2607~2657 cm^{-1} are removed from the water retrieval, which eliminated the spurious day-night bias reported in V6 water products (Section 4.2).

Several other changes were also made:

- The number of humidity trapezoids was increased to better represent the boundary layer humidity and the trapezoid boundaries were made a subset of the temperature trapezoid boundaries to increase the stability of the retrieval.
- The damping in the water retrieval was increased to reduce the degradation of near surface water vapor profiles relative to SCCNN results.
- Channel changes: The strongest water lines were removed from the retrieval as it is shown their utilization had a negative effect. Water continuum channels were added to the water retrieval following the suggestion of the science team.
- In the water vapor first guess, we now taper the SCCNN profile to climatology at four levels above the tropopause. This has removed spurious structure found in the stratospheric water profile present in V6.

3.5 Changes to Temperature and Surface Retrieval Algorithm

The temperature retrieval trapezoids were increased in the boundary layer and were made more uniform vertically, which led to a better representation of the boundary layer and of the tropopause.

Longwave channels with cloud clearing corrections larger than 5K for a given profile are now excluded from the temperature retrieval for that profile.

In V6, the surface temperature was modified in the temperature retrieval step. This feature was removed for V7.

Seven additional 4μm channels were added to the temperature retrieval to increase information at the top of the atmosphere, which also affects the ozone retrieval.

The emissivity channel list was modified to ensure information in the ozone band is used to adjust the surface emissivity there.

3.6 Changes to Temperature and Water Vapor Quality Control (QC)

The following modifications have been made:

- The error estimate used as the decision point for atmospheric temperature and water profile QCs is moved from 6 layers above the surface in V6 to 2 layers above the surface over frozen and land surfaces in V7. This effectively uses the entire profile error estimate information and allows the algorithm to make finer distinctions of quality over land and frozen areas near the surface.
- The numerical threshold points for marking profile levels with QC=2 is tightened over land from the mid to lower atmosphere, and for frozen cases in the middle atmosphere, while the numerical threshold for frozen cases near the surface is slightly loosened which increases the yield over frozen surfaces.
- One relative humidity quality check was modified. In V6, cases with unrealistic relative humidity and low cloud fractions are identified and the profiles are marked as QC=2 up to 100hPa. In V7 this test is revised to address the uncertainty associated with the cloud top pressure in a less restrictive way by moving PGood and PBest only a few levels instead of to 100hPa.

4. Details on Major Product Improvements, Changes, and Performance Limitations

Note: In the following discussion frequent reference is made to intermediate versions of the AIRS V7 retrieval system. The reader is advised to consult the discussion of these various versions in Section 3 and in particular Figure 3.1.

4.1 Better Agreements between IR-only Surface Classification and NSIDC data

Contributor: Qing Yue, Evan Manning, Bjorn Lambrigsten

Data and Methodology:

AIRS surface classification in the IR+MW and IR-only retrievals from V6 and V7 is evaluated against the data from the National Snow and Ice Data Center (NSIDC). The surface classification in the AIRS data is given by the SurfClass variable in the L2 product and summarized as the SurfClass_Count_A(D) in the L3 products for ascending (descending) node. Table 4.1.1 shows how surface types are determined in the IR+MW algorithm and the major changes in the IR algorithms from V6 to V7. Land/water/coast are all determined from geolocation and a digital elevation model in all retrievals. In the IR+MW retrieval (for both V6 and V7), microwave radiances are used to distinguish snow and ice from bare land or liquid water surfaces. Since microwave emissivity information is not available for the IR retrieval, in V6 the surface temperature from forecast (tsurf_forecast), or climatology if forecast is not available, is used to differentiate frozen and non-frozen surfaces (frozen land if tsurf_forecast < 273.15 K; frozen water if tsurf_forecast < the sea water freezing point temperature). As a result, V6 IR-only retrievals have less accurate discrimination of the surface types, and hence the surface emissivity in the high latitude regions is less accuarate, which affects the retrieval yield, accuracy, and precision of various parameters in these regions. The V7 IR-only algorithm uses auxiliary surface type information from the Global Forecasting System (GFS). Over ocean, sea ice is assigned if GFS Ice concentration > 0.05. Over land, the Water Equivalent Accumulated Snow Depth (WEASD) from GFS is used. Since GFS surface auxiliary data is available on a daily resolution, there is no temporal variation in the V7 IR-only surface classification within the same day. This report illustrates the differences among different retrieval algorithms by qualitatively comparing the L3 SurfClass_Count_A(D) with the daily snow and ice cover data from the NSIDC.

One day of data is used in this study: January 02, 2015. Over Northern Hemisphere (NH), NSIDC's Interactive Multi-sensor Snow and Ice Mapping System (IMS) NH Snow and Ice Analysis at 24-km resolution is used. For a list of sensors and instruments used to produce this data, please refer to the following link: https://nsidc.org/data/docs/noaa/g02156_ims_snow_ice_analysis/#sensor-table. Over the Antarctic Area, NOAA Antarctic Sea Ice Extent Data is used, which is based on SMMR, SSMI and SSMIS observations. This data is also downloaded from NSIDC and its land surface type corresponds with frozen land type. Only daily data are available from these two datasets. Therefore, AIRS L3 daily summarization on the surface classes are used in this study.

Results and Conclusions:

Figures 4.1.1 and 4.1.2 show the maps of one day averaged surface classifications from AIRS L3 daily data on Jan 02, 2015 (ascending node: Figure 4.1.1; descending node: Figure 4.1.2) and

the NSIDC daily data. Maps are colored coded by the surface classifications shown in Table 4.1.1. Compared with V6 IR-only algorithm, the V7 (labeled as V6.6) IR-only surface classification is very similar to the NSDIC data and more consistent with IR+MW algorithms in the high latitude. The IR+MW algorithms identify Europe as nonfrozen land with spatial extent varying significantly between day and nighttime, while NSDIC data identifies this region as frozen land. Other small difference between the IR+MW surface type and NSDIC is noticed especially over the mountain regions (such as Tibetan Plateau and Sierra Nevada), which is probably due to the resolution difference between the datasets.

Table 4.1.1 AIRS Surface classification in V6 and V7 for IR+MW and IR retrievals

Surface Classification	V6 and V7 IR+MW	V6 IR	V7 IR
0	Coast	Coast	Coast
1	Land	Land	Land
2	Ocean	Ocean	Ocean
3	High microwave emissivity sea ice	Sea ice if tsurf_forecast < sea water freezing point	Sea ice if GFS Ice Concentration > 0.05
4	Low microwave emissivity sea ice	Not Assigned, filled by -9999	Not Assigned, filled by -9999
5	Snow Covered Land (high microwave emissivity)	Frozen land if tsurf_forecast < 273.15 K	Frozen land if GFS WEASD > 0
6	Glacier/Snow	Not Assigned, filled by -9999	Not Assigned, filled by -9999
7	Snow Covered Land (Low microwave emissivity)	Not Assigned, filled by -9999	Not Assigned, filled by -9999

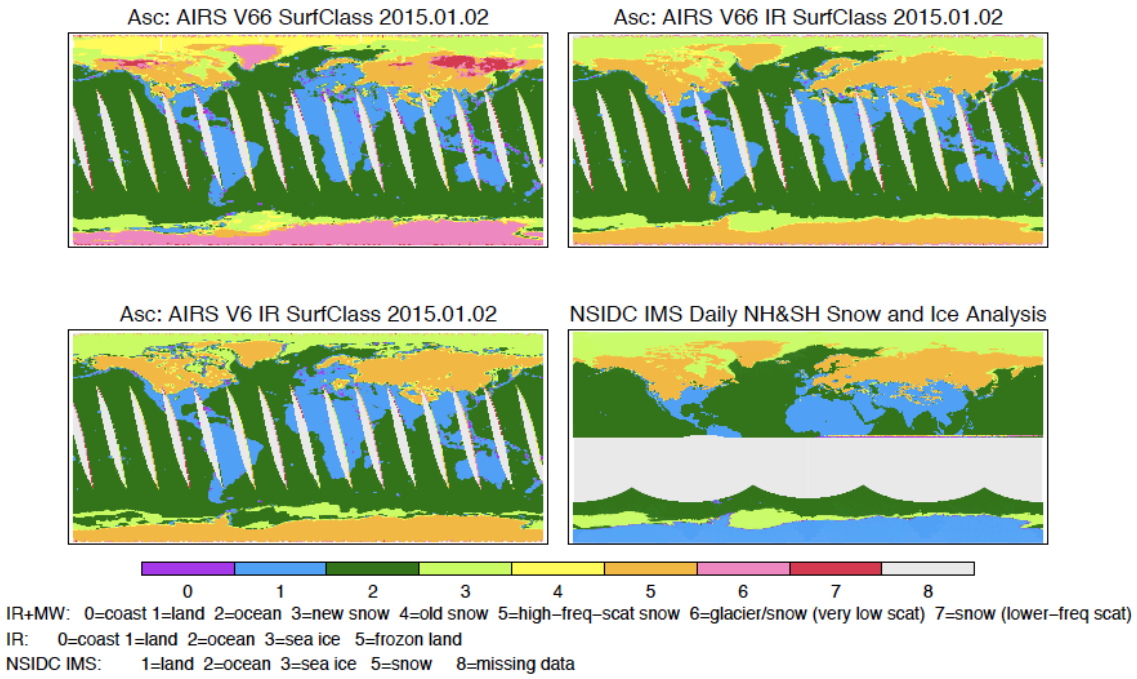


Figure 4.1.1. Daily mean surface classifications from AIRS retrievals for ascending node and the daily data from NSDIC. Gray color (8) means no data.

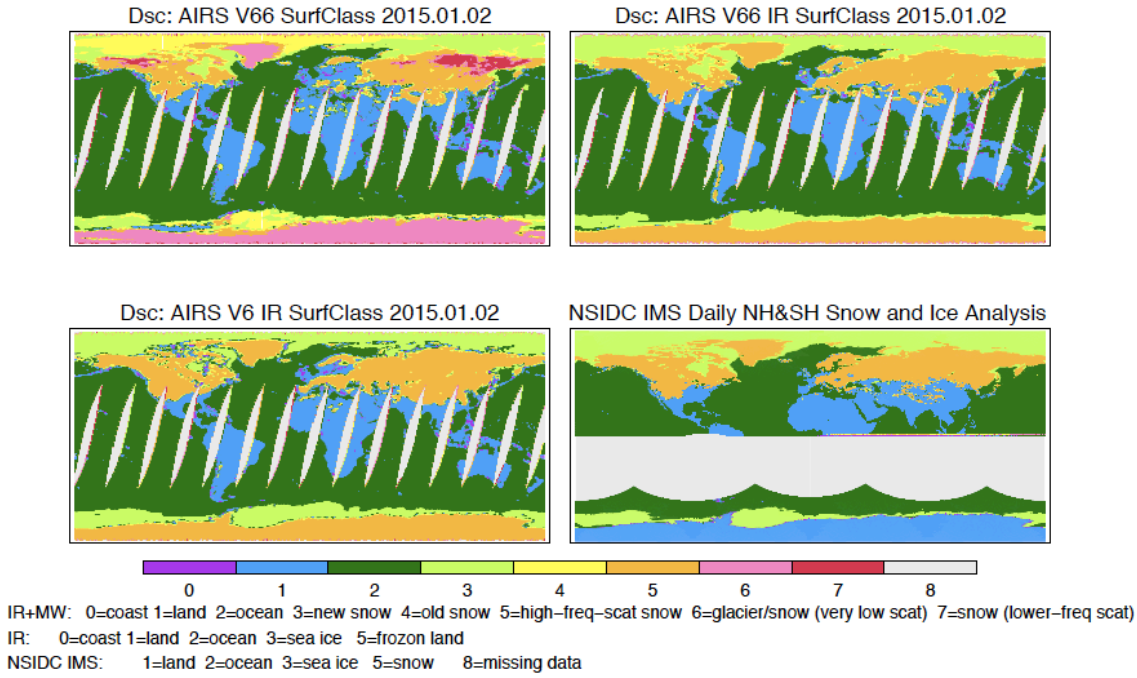


Figure 4.1.2. Daily mean surface classifications from AIRS retrievals for descending node and the daily data from NSIDC. Gray color (8) means no data.

4.2 Removal of Spurious Day-time Bias and Day-Night Performance Difference on Water Vapor Retrieval

Contributor: Qing Yue, Eric Fetzer

Data and Methodology:

Previous evaluations on AIRS V6 water vapor products revealed spurious bias in day-time products and large difference in the retrieval performance between day- and night-time cases. Such bias has been removed in the V7 retrieval as a result of removal of shortwave channels (2607~2657 cm^{-1}) in water vapor retrieval on both profiles (Sec. 5.1.1) and column integrals. In this section, to highlight such improvements, AIRS IR+MW and IR-Only total precipitable water vapor (TPW) retrievals from both Version 6 and V7 (labeled as V6.6) are compared with the measurements from the Advanced Microwave Scanning Radiometer - Earth Observing System (AMSR-E) on board of Aqua obtained from Remote Sensing System. Daily gridded TPW in the Ascending (Asc) and Descending (Dsc) passes from the two instruments are used in this study to examine the difference between the day- and night-time AIRS retrievals. Measurements in January and July of 2003 and 2011 are used. Comparisons with other microwave-based TPW products have been conducted, such as SSMI, GMI, and TMI. Similar results are obtained with the AMSR-E comparisons reported here.

Both AIRS and AMSR-E are on board NASA's *Aqua* satellites, which ensures the collocation of the measurements. Monthly mean differences between the AMSR-E and the AIRS TPW data are calculated for Asc and Dsc passes separately. Results are presented on maps of 1×1° longitude by latitude grid boxes. AMSR-E TPW is derived over ocean only; therefore, comparisons between instruments are only available over oceanic regions.

AIRS retrieval yield for TPW is calculated as the percentage of successful TPW retrievals (QC=0 or 1) in the total number of observations in each grid box. Since AIRS profile products have a level-by-level QC flag, a successful retrieval of TPW in a given pixel generally corresponds with successful retrievals of nearly complete vertical profiles of atmospheric temperature and humidity.

Results and Conclusions:

Similar conclusions are drawn from results using data from different months. Only the January 2003 results are shown in this report.

1) AIRS TPW Retrieval Yield Day and Night Differences and Changes among Retrievals

Figure 4.2.1 shows the yield maps from the four AIRS retrievals for Ascending (Asc) and Descending (Dsc) passes, respectively. The difference between the two (Asc-Dsc) is also shown to highlight the day and night differences on yield. Similar spatial patterns are seen for all four retrievals, which shows yields > 80% over the tropical and subtropical ocean except for the regions with extensive cloud cover, such as ITCZ and the subtropical stratocumulus regions. Yield is reduced in the high latitude regions, especially during the cold season. The difference between the Asc and Dsc passes is also similar among algorithms. A reduction of yield during the nighttime (Dsc pass) is apparent in the subtropical oceanic low cloud region, which is consistent with the thickening of these clouds during the nighttime in this region. Over the polar area, generally the yield in the Asc pass is smaller than that in the Dsc pass for all algorithms. The differences change signs from negative to positive along the 70°N latitudinal line. This feature is also present in results using other months. It is probably related to the complex spatial and temporal variation of surface classes within one field of regard in these regions and the limitations of SCCNN training on the transition between frozen and non-frozen surfaces, which warrants further study.

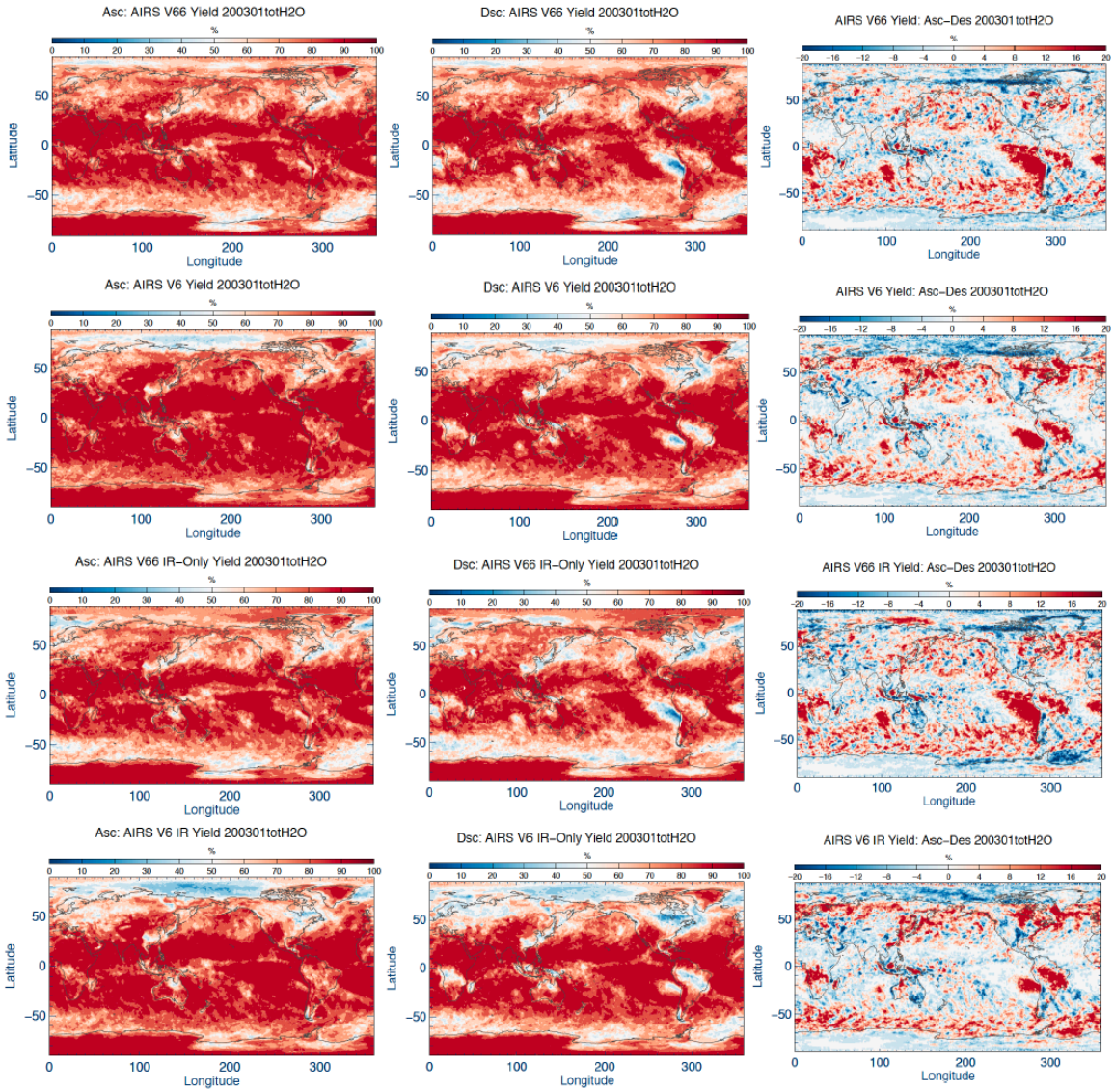


Figure 4.2.1. The retrieval yield of January 2013 AIRS total precipitable water vapor from V7 IR+MW, V6 IR+MW, V7 IR-only, and V6 IR-only products. V7 is labeled as V6.6 in the panels. Results are shown using the ascending (Asc: left) and descending (Dsc: middle) passes, separately, with the difference between the two passes shown in the right column.

Figure 4.2.2 highlights the differences of yields among different AIRS retrievals. The new versions produce smaller yields than their V6 counterpart over ocean, especially in the subtropical low cloud regions and the midlatitude storm tracks. However, a higher yield in the Arctic regions is seen in the new versions, which is related to the improvements of SCCNN in the polar region and the change of quality control methodology in the high latitudes. The difference between the IR-only and IR+MW retrievals in V7 is mostly over high latitude ocean regions and land regions such as eastern China, Amazon Basin, and central Australia. An increase of yield is seen in the high polar region.

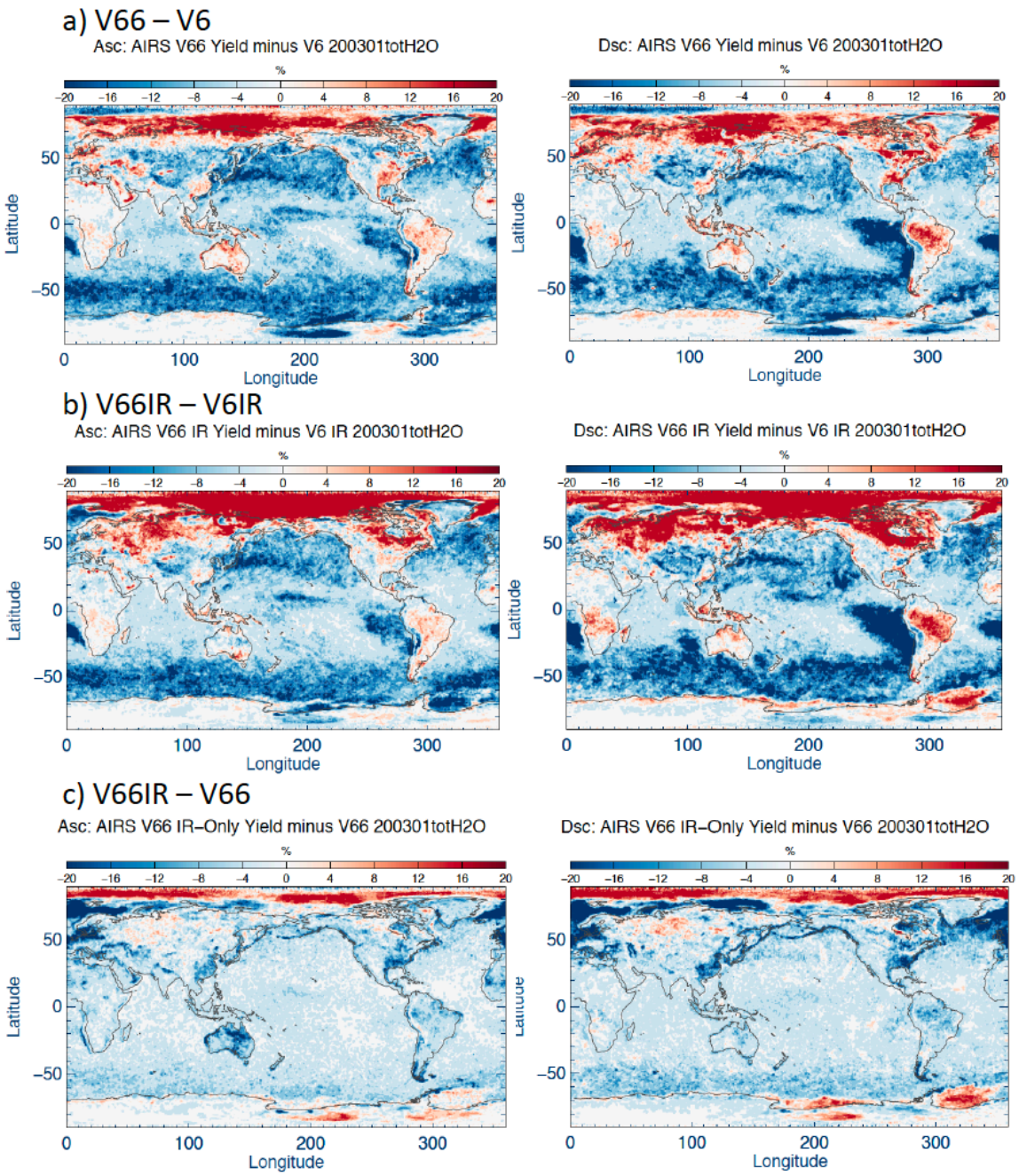


Figure 4.2.2. Differences of retrieval yield for AIRS total precipitable water vapor between V6.6 IR+MW and V6 IR+MW (a), V6.6 IR-only and V6 IR-only (b), and V6.6 IR-only and V6.6 IR+MW (c). Ascending (left) and descending (right) passesare shown separately. Data from January 2003 is used. Note that V7 is labeled as V6.6 in the figure.

2) Removal of Day and Night Retrieval Performance Difference in the AIRS TPW Products and Validation Using AMSR-E TPW Products

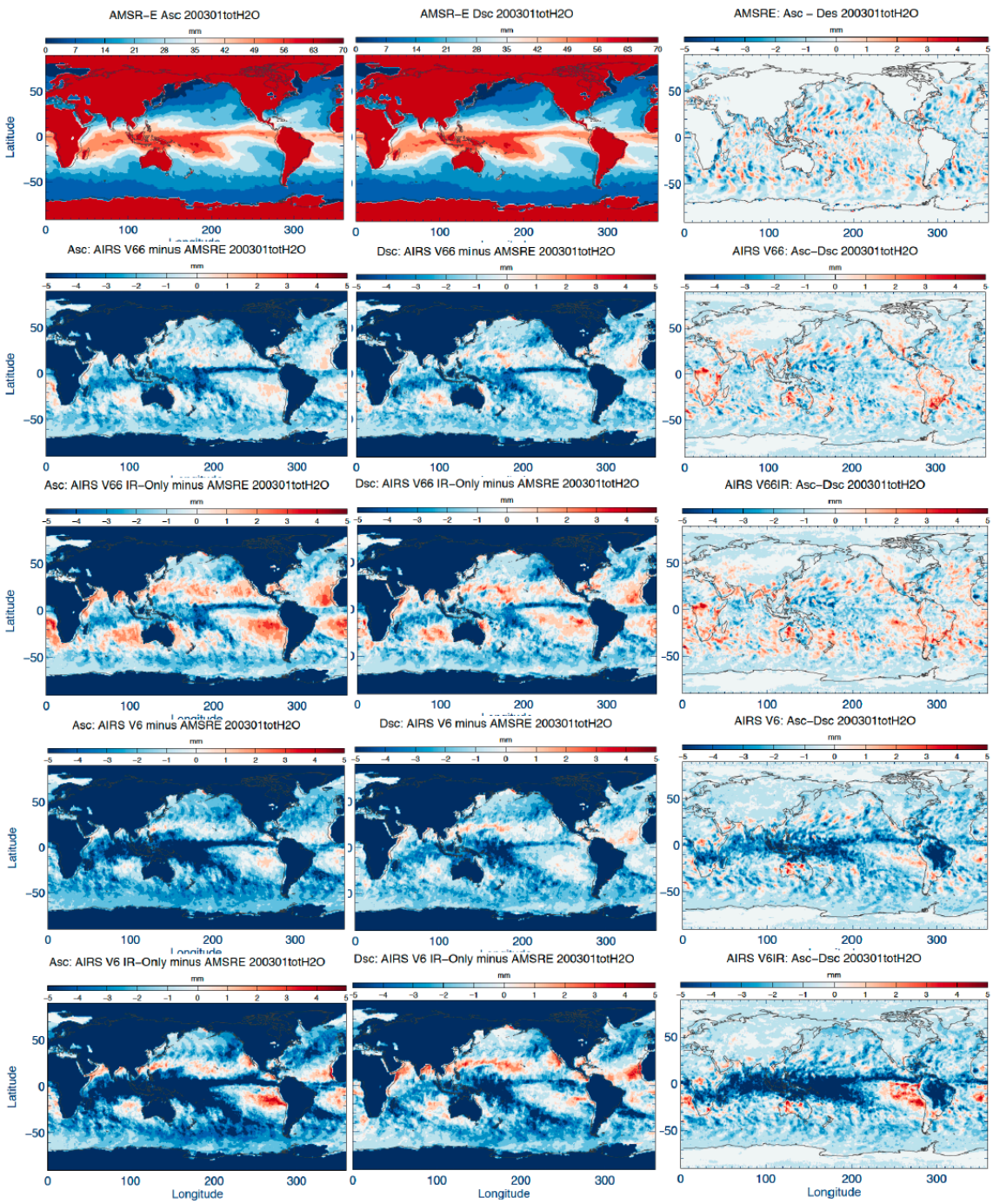


Figure 4.2.3. The monthly mean TPW from AMSR-E for Asc and Des passesare shown on the top together with the differences between the two passes. The AIRS TPW biases are shown for different versions of AIRS retrievals by comparing with the AMSR-E observation. Results for V6.6 IR+MW, V6.6 IR-only, V6 IR+MW, and V6 IR-only are shown by separating ascending (left) and descending (middle) passes. The day and night TPW differences from different data products are shown using the January 2003 data.

Figure 4.2.3 shows the monthly mean TPW from AMSR-E and the difference between the Asc and Dsc passes. The difference between AIRS retrievals from different algorithms is also

shown, as well as the difference between the Asc and Dsc passes from AIRS data. Compared with the microwave-based products such as AMSR-E TPW, all four AIRS TPW products show a dry bias in the deep convective cloud regions and a wet bias in the subtropical oceanic low cloud regions. However, compared with previous versions (V6), the magnitude of dry biases is greatly reduced in V7 for both Asc and Dsc passes. The V6 TPW show biases with a much larger magnitude in the daytime than nighttime, causing a significant day-night TPW difference in the V6 products. This artifact is caused by different retrieval performance during the day and night and has been removed in the new versions. However, the V7 IR-only product produces a larger wet bias in the low cloud topped subtropical oceans, which can be as large as 20% relative to the monthly mean TPW in this region.

4.3 New Stochastic Cloud Clear Neural Network (SCCNN) for V7 AIRS IR-only Algorithm by Including AMSU-A1

Contributor: Adam B. Milstein

Data and Methodology:

The SCCNN produced temperature, water vapor, and surface temperature that are used as the first guess in the AIRS L2 retrieval algorithm. They are also are distributed to users in the L2 support product as "TAirSCCNN", "H2OCDSCCNN", and "TSurfSCCNN". For more details on the AIRS SCCNN algorithm and the differences between V6 and V7, readers can refer to Section 3.1. In this section, the SCCNN profiles have been compared with the "Truth" data, which include ERA-interim reanalysis, IGRA radiosonde profiles, and ARM SGP radiosonde profiles.

AIRS and ECMWF are matched by applying the nearest neighbor method to select the nearest ECMWF profile. AIRS and radiosondes are collocated with <3 hour temporal tolerance and 100km spatial tolerance. Results are shown with and without application of the "PBest" flag (QC=0) down to the surface.

Results and Conclusions:

V7 (labeled as V6.55, which is synonymous with 6.54 for the NN) is significantly improved overall versus previous V6, V6.4x versions, including difficult 01/2015 sonde PBL cases. Figures 4.3.1 and 4.3.2 illustrate the global improvement in v6.55 IR-only temperature profile retrieval performance (bias and RMS) versus v6 and the previous v7 candidates. Figures 4.3.3-4.3.6 show, respectively, global temperature RMS error, temperature bias, water vapor %RMS, and water vapor RMS statistics for six example days spanning different times of year, all showing expected performance, with bias structure improved versus AIRS V6.

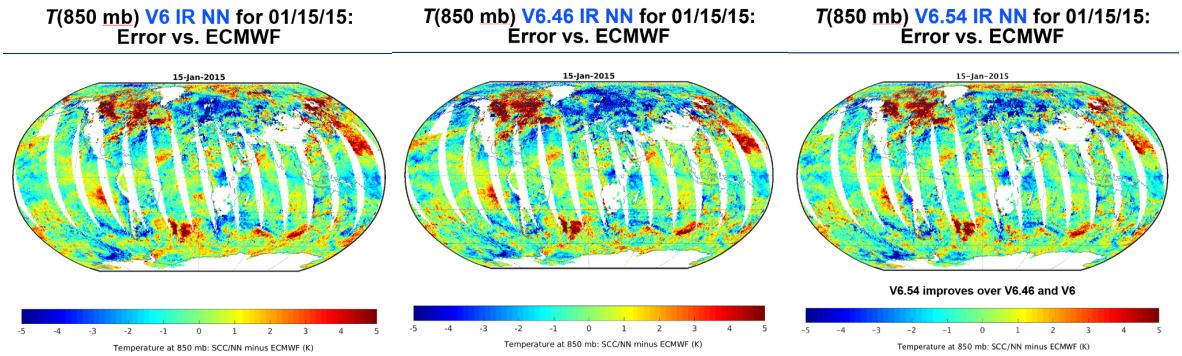


Figure 4.3.1: Improvements in air temperature at 850 mb for newer IR-only version of V7 candidate (v6.54 NN) versus previous v7 candidate (v6.46 NN) and v6 NN

Figure 4.3.7 shows the improvements in bias performance versus radiosondes in the PBL for v6.54 relative to previous V7 candidates and V6. The IR-only case improves versus previous versions due to the improvements in principal component transforms for the SCC algorithm introduced in v6.51. The AIRS+AMSU-A1-only first guess performs almost as well as the full AIRS+AMSU version. V6.54 was delivered as the most recent version of SCCNN, as a first guess to that and subsequent AIRS V7 candidate physical retrieval algorithms. Comparisons with SGP sondes show very similar PBL performance for the AIRS+AMSU-A1 SCCNN with AIRS+AMSU SCCNN despite the loss of AMSU-A2 channels in the former.

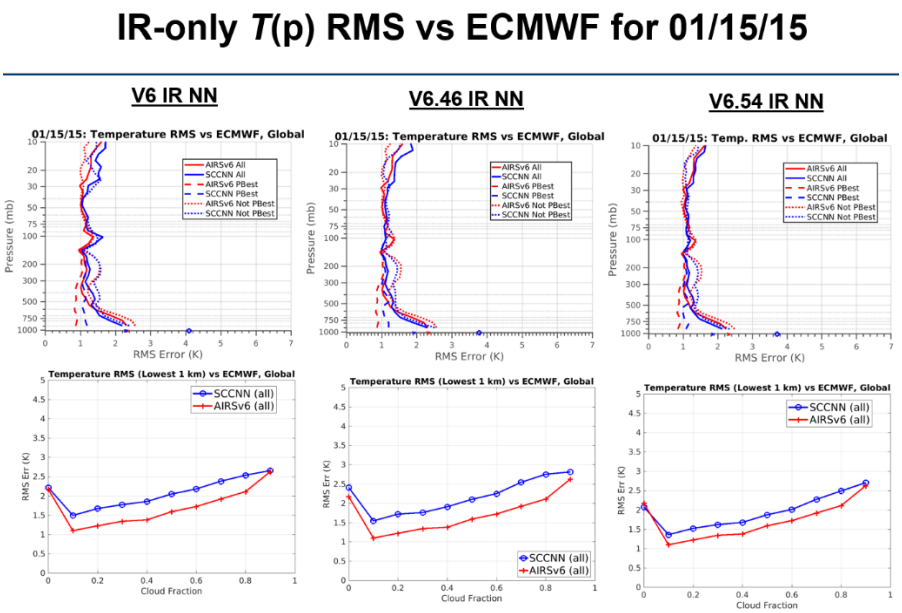


Figure 4.3.2: Improvements in global RMS temperature error performance for newer IR-only version of V7 candidate (v6.54 NN) versus previous v7 candidate (v6.46 NN) and v6 NN

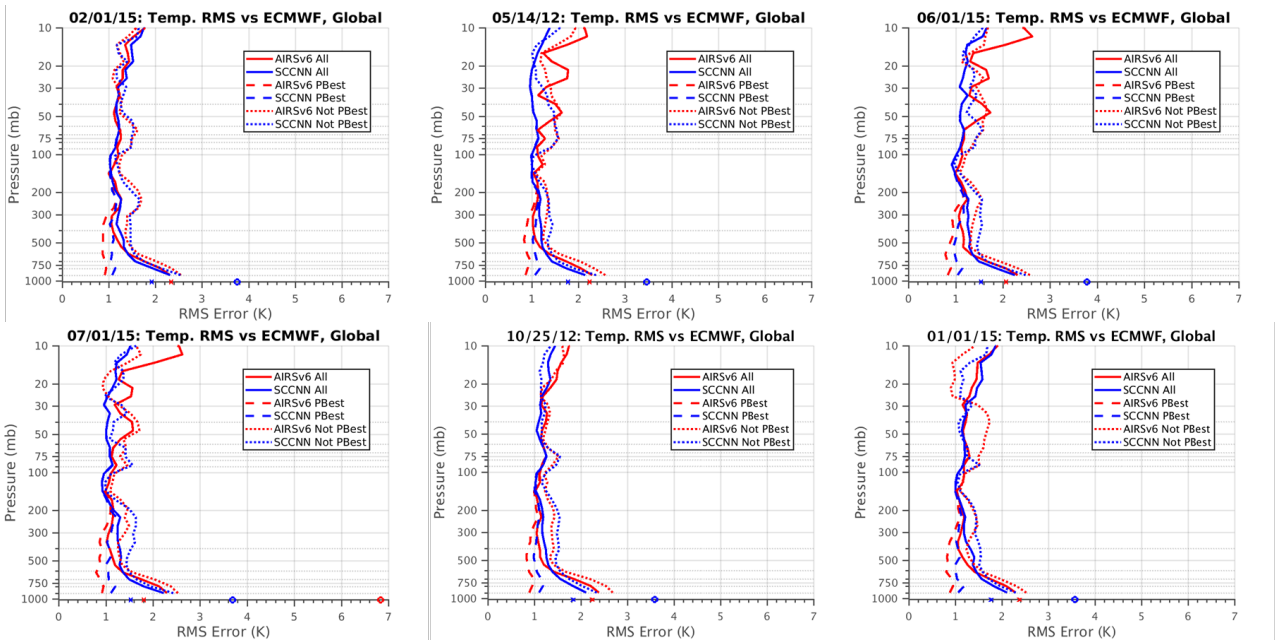


Figure 4.3.3: Global temperature RMS error vs ECMWF for newer AIRS+A1version of V7 candidate (v6.54 NN) versus v6 AIRS+AMSU product.

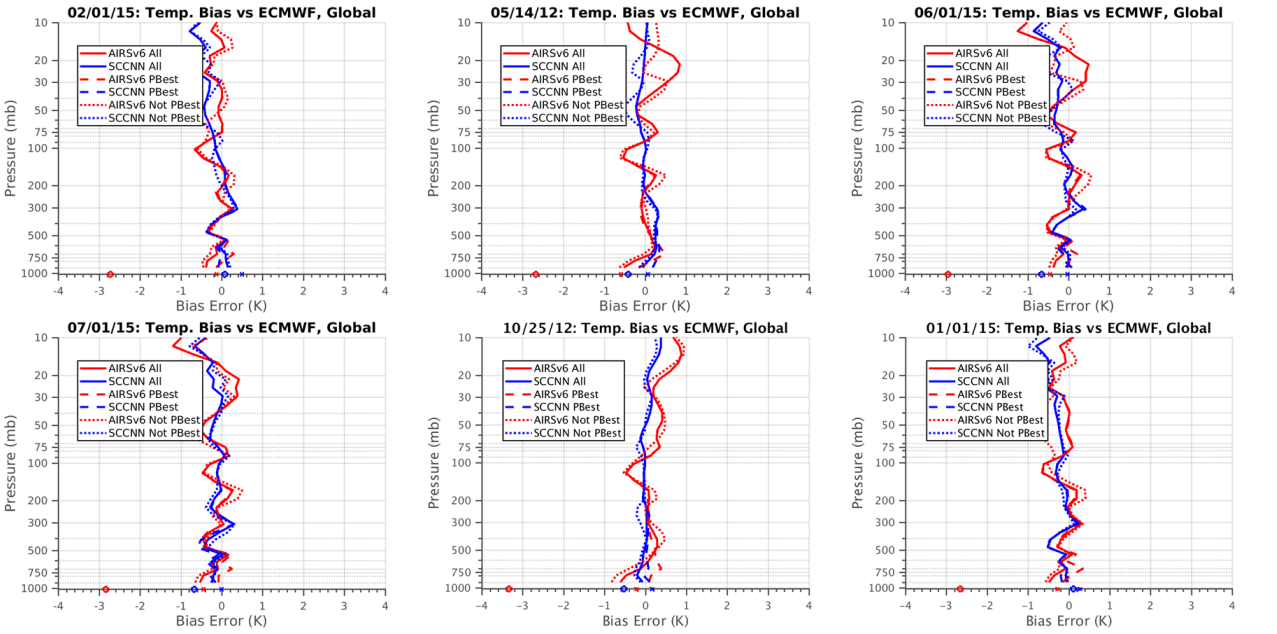


Figure 4.3.4: Global temperature bias error vs ECMWF for newer AIRS+A1version of V7 candidate (v6.54 NN) versus v6 AIRS+AMSU product.

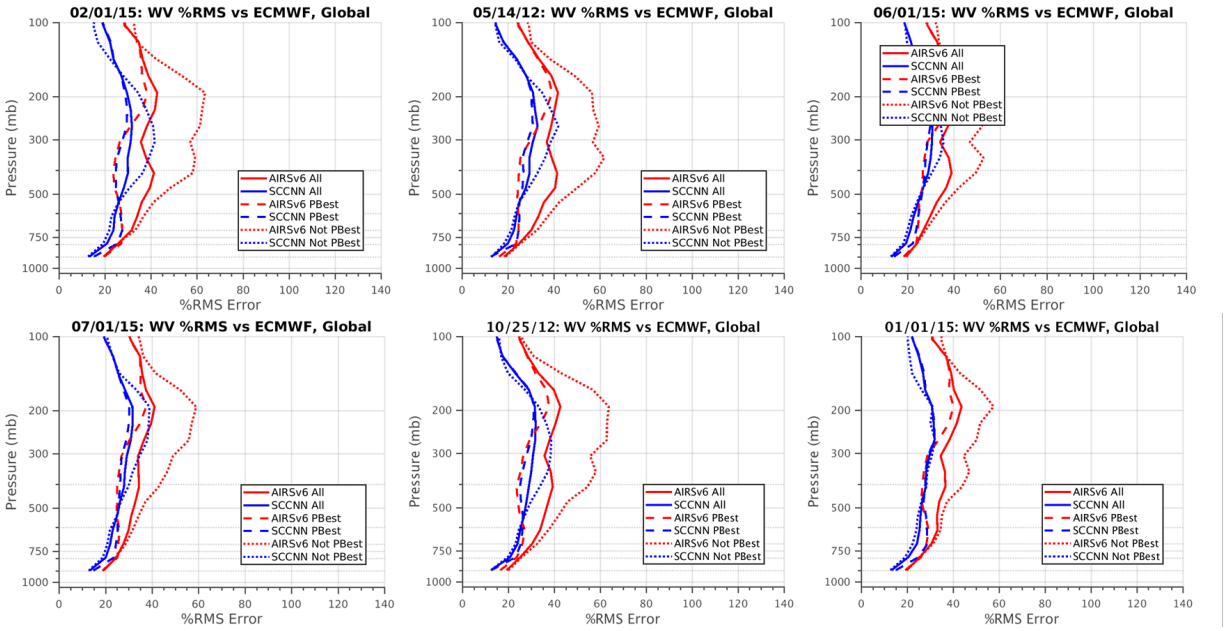


Figure 4.3.5: Global water vapor %RMS error vs ECMWF for newer AIRS+A1version of V7 candidate (v6.54 NN) versus v6 AIRS+AMSU product

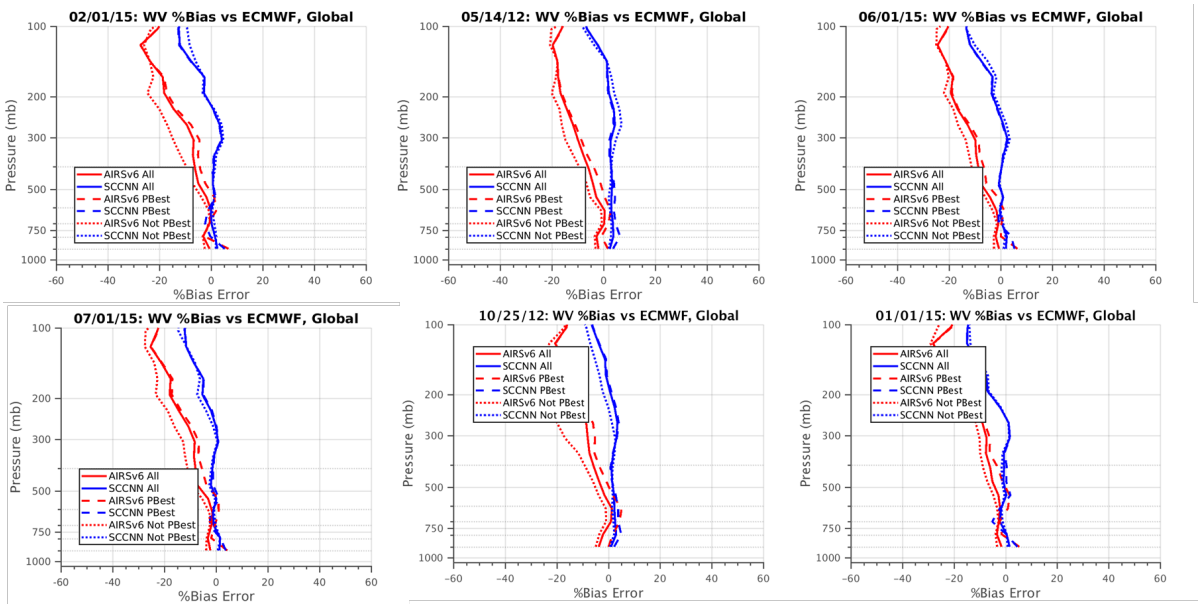


Figure 4.3.6: Global water vapor %bias error vs ECMWF for newer AIRS+A1version of V7 candidate (v6.54 NN) versus v6 AIRS+AMSU product

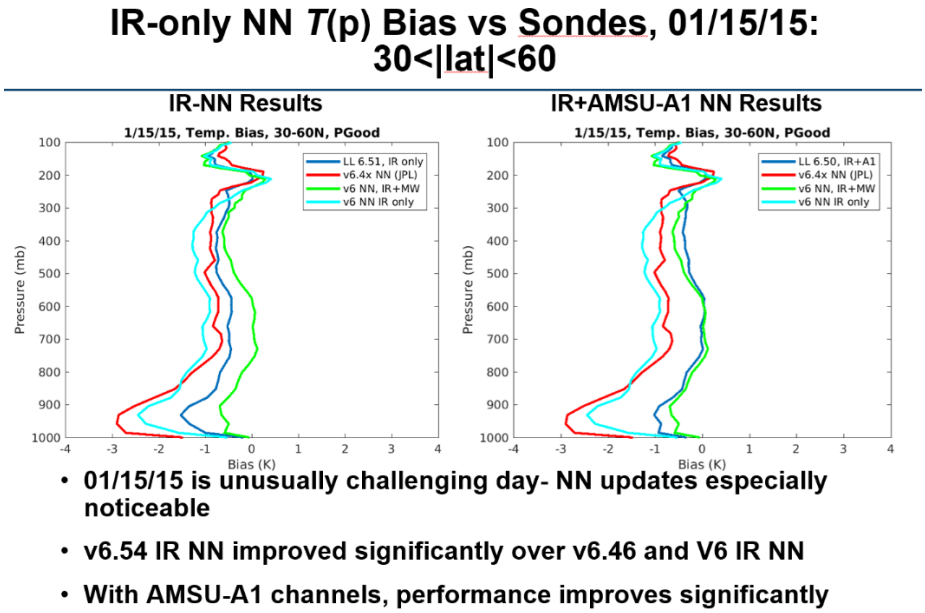


Figure 4.3.7: Improvements in PBL temperature bias versus sondes for 01/15/15, a challenging test

4.4 Other Major Improvements in the V7 Retrieval Products

Comparisons with the independent field campaign data indicate that the V7 produces smaller biases and Root Mean Square Errors (RMSEs) in PBL over non-frozen ocean. Figure 1.1 of this report and detailed results in Section 5.1.1 and 5.1.3 show the analyses focusing on this point. Over non-frozen land, the comparisons with collocated ECMWF data (Section 5.1.1) show that the bias in the V7 water vapor profiles is smaller than for V6, which is mainly due to an improved first guess from SCCNN.

For ozone products, Version 7 shows significant improvements in vertical resolution and agreement with ozonesonde in the lower stratosphere (See Section 5.4.3). The bias in total column ozone is greatly reduced in V7 than in V6 by comparison with OMI and OMPS products (See Sections 5.4.2 and 5.4.4).

4.5 Performance Limitations of Challenging Scenes in V7

Despite the overall improvements of the V7 retrieval products, several performance limitations have been identified. These issues are related to highly challenging scenes such as surface temperature over Antarctic sea ice, near surface temperature and water vapor profiles over snow and/or ice covered land, and cloud properties in cases with small effective cloud fractions (ECF < 0.05), all of which are highly regional in nature. These analyses are presented in detail in this section. More thorough studies are needed to examine how these known limitations impact science investigations using AIRS data in these conditions. It is also noted that the values of degree of freedom (DOF) and verticality of the water vapor and temperature profiles reported in the V7 Level 2 product are smaller than those in V6. This is related to how the current algorithm (both V6 and V7) calculates the temperature and water vapor averaging kernel matrices and error estimate and users should be cautious using these variables interpreting the sensitivity and information content of the AIRS spectra. Detailed explanations and recommendations regarding the temperature and water vapor profile averaging kernel matrices in the AIRS L2 product are presented in Section 6.

4.5.1 Larger Surface Temperature Error over Antarctic Sea Ice in V7 SCCNN

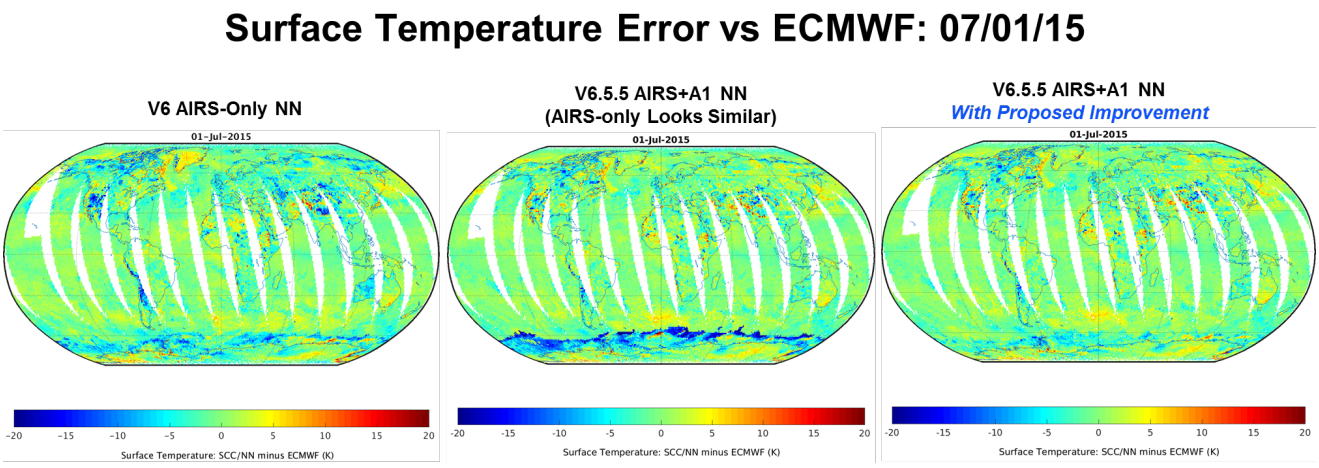


Figure. 4.3.8 The surface temperature difference between AIRS SCCNN and ECMWF on July 1, 2015. Two operational versions are shown on the left and middle panels for V6 IR-only and V7 IR-only, respectively. The right panel shows the result with proposed improvements on SCCNN.

V6.5.5 AIRS-A1 and AIRS-only NN surface temperature appears to have larger error for Antarctic sea ice in certain cases than V6 AIRS-only NN as shown in Figure 4.3.8, despite similar processing and training. What follows is an explanation and a proposed resolution to this issue.

In V6 NN, AMSU channel 1 was nominally used to identify the presence of "sea ice" at the poles. Retrievals were then done using a coefficient set trained on sea ice. The V6 AIRS-only NN required a different approach. In the AIRS-only case, early builds of V6 NN initially could not flag "sea ice". Hence, polar ocean was all treated the same. JPL recommended at the time that sea ice be flagged in the AIRS-only case using "Tsurf_Forecast", with retrievals then done using a coefficient set trained on "land". This recommendation was implemented in the final version

The V6.5.5 AIRS-only NN and AIRS+A1 NN currently use the same approach as V6 AIRS-only NN. By all indications, this is executing properly (without apparent implementation bugs or training set anomalies). The retrieval code handling this is unchanged from the previous version, and expected outputs are generated. However, surface temperature over sea ice near Antarctica seems to have larger errors (cold bias) in some summer cases with heavy, low cloud cover than V6 did. This issue is traceable to this basic assumption that treating sea ice as "land" will work well for sea ice. This (questionable) assumption worked better for V6 vs V6.5.5 sea ice, for reasons that follow.

A review of the training code shows a subtle change in V6.5.5 approach vs V6 introduced when code was recently updated: The V6 "land" stratification screens for profiles with land fraction >= 0.01, while the V6.5.5 "land" stratification screens for profiles with land fraction >=0.9 . Most of the time this variation would have small or very small impact. Interpolation logic in the retrieval code is really meant to be how the system blends boundary cases between "land" versus "ocean", rather inclusion of overlapping boundary cases in the training set. (For example, if land fraction is only 0.01, then the "ocean" NN result is given 0.99 weight and the "land" NN result is given only 0.01 result by the NN retrieval code.) However, sea ice for AIRS-only NN could be an exception, because we are treating frozen ocean as 100% land. The V6 polar "land" neural network would generalize well to sea ice due to inclusion of profiles with

very small land fraction that are up to 99% sea ice. The V6.5.5 polar "land" neural network may not generalize as well with its stricter landfrac cutoff.

The proposed method to fix this issue is very easy: Use already available "sea ice" coefficients rather than "land" coefficients for profiles flagged as sea ice. The AIRS-only and AIRS+A1 coefficients *already* include a "sea ice" stratification that we haven't been using, since we had left the existing V6 logic in place. By changing only a couple of lines of code in the NN retrieval, we can actually use the "sea ice" training set for its intended purpose – and get better results. No retraining would be required. The example result earlier in this section uses this approach. Although this algorithm fix is not included in V7, it will be addressed in the future data versions.

We have also presented quality control flag – a retrieval error covariance predictor based on mixture density networks. This flag is working as a prototype at GSFC and its outputs could be included in a final product to the users in the future.

4.5.2 Water Vapor Profiles Near Surface Over Land

As shown in Section 5.2, the comparison of AIRS L2 water vapor profiles with the Integrated Global Radiosonde Archive (IGRA) radiosondes shows that in the lower troposphere (below 700 hPa), the V7 algorithm applies a smaller weight to the water vapor profiles (See Section 6). As a result, water vapor profiles in the lower troposphere stay close to the SCCNN profiles. Although this change statistically shows a better agreement with reanalysis data and radiosonde observations (See Section 5.1, Figs 5.1.3 and 5.1.4), it limits the capability of physical retrieval to correct the water vapor retrievals near surface from the a priori (SCCNN) compared to in V6 or V6IR. It is possible that the difficulty of SCCNN to capture the climate extremes may result in larger errors in the retrieval. Our study shows that during the 2003 heatwave over Europe, V7 retrievals have a larger moist bias in lower atmosphere with respect to the radiosonde measurements than V6. More detailed studies are needed to quantify this impact in the V7 water vapor retrievals, especially the profiles near the surface.

4.5.3 V7 IR-only Retrieval Near the Surface Over Frozen Surface Classes

By comparing with field campaign in-situ observations in the Central Arctic (Section 5.1.3), we have noticed that over the frozen surfaces V7 IR-only retrievals show a larger cold (dry) bias on temperature (specific humidity) profiles near the surface (below 600 hPa). The comparison with ECMWF model analysis (Section 5.1.1) also draws similar conclusions although ECMWF data suffers a larger uncertainty in the Arctic regions over ice and snow-covered land surfaces. However, the impact of this bias on the long-term climate data record over this region from AIRS V7 retrievals has not been fully assessed. Further algorithm developments are currently undergoing to improve the IR-only algorithm including SCCNN.

4.5.4 Underestimation of Observation Uncertainty by Error Estimate Reported in the AIRS L2 Product

For both V6 and V7, each retrieved physical quantity has its own matching error estimate reported in the L2 retrieval product. The name of the parameter with "Err" appended gives the error estimate for this parameter. For details on how error estimate in the AIRS retrieval is derived, readers can refer to the following document: AIRS/AMSU/HSB Version 6 Level 2 Quality Control and Error Estimation (Olsen et al. 2017) and AIRS/AMSU/HSB Version 7 Level 2 Quality Control and Error Estimation (Thrastarson et al. 2020). It has been noticed by the AIRS team that the L2 reported error estimate often underestimates the uncertainty of the

retrieval, which is shown by a much smaller magnitude compared to bias and RMSE calculated against radiosonde measurements (see Section 5.1). Instead, error estimate is used to derive the quality indicators (QC) and only represents general quality of the retrieval. Therefore, caution is needed when using the "Err" variables in research.

4.5.5 Thin and Broken Clouds with Effective Cloud Fraction < 0.05 in the AIRS Cloud Products

As discussed within the section regarding cloud products evaluation, the thinnest and most tenuous clouds are not entirely reliable. This problem was greatly reduced from early versions of the AIRS algorithm to an "acceptable" level in v6. On occasion striping in clouds can appear in the scan direction of the swath and are associated with clouds of very low effective cloud fraction (ECF) less than 0.05, and frequently even lower than 0.01 (see Level 2 and Level 3 examples in the clouds section). These problem clouds can appear in specific cloud regimes including in the cold sectors of midlatitude cyclones, and less often in the tropics. In previous studies, we have highlighted our approach to filtering out these clouds. In Kahn et al. (2008, https://doi.org/10.5194/acp-8-1231-2008), we show that AIRS cloud detection using the ECF variable is quite reliable down to an ECF of 0.03 or so. Typically we will set a threshold of ECF to filter out clouds and not use any values of ECF below that threshold. If the threshold is on the high side, say ECF=0.1, many thin cirrus clouds or shallow and broken trade cumulus clouds will be removed, as a result, a threshold of that magnitude will lead to a significant reduction in midlatitude cloud amount and subtropical shallow cumulus. If a lower threshold, for instance ECF=0.01 is used, much more of the cirrus and shallow cumulus will be retained but more of the problematic clouds will be included. Ultimately, there is no single value of ECF in which we recommend to remove and we have not studied exhaustively this issue in v7. As the AIRS v7 algorithm is based on cloud clearing, as previous algorithms, which assumes that the temperature and water vapor are uniform in a 3x3 array of AIRS footprints, any variability in temperature and water vapor might be aliased into spurious clouds within the 3x3 field of regard.

5. Analyses and Results

5.1 Analyses on L2 Temperature (T) and water vapor (Q) vertical profiles

In this section we present the detailed results for the L2 T and Q vertical profiles from V7, which form the core products of the AIRS retrieval system. The differences in the L2 T and Q profiles between V7 and V6 are assessed against ECMWF model analysis (5.1.1) and four different sets of *in situ* observations: IGRA radiosondes (5.1.2), radiosondes during the Marine ARM GPCI Investigation of Clouds (MAGIC) campaign (5.1.3), the radiosondes obtained from multiple field campaigns in the Central Arctic (5.1.4), and the PREPQC radiosondes (5.1.5 and 6.2). Initial science assessment is performed by quantification of the retrieval yield, retrieval bias, and the sampling biases in different regions and cloud conditions. These analyses aim to 1) evaluate the general quality of the retrieval, 2) quantify the differences between the first guess and the physical retrieval on the key products (T and Q) from the retrieval system, and 3) explore the possible connections between the differences and other physical parameters such as cloud properties and surface classes. The trend (drift) of temperature and water vapor biases at each level is assessed against the PREPQC measurements as shown in Sec. 5.1.5. The vertical structure and variabilities of the L2 T and Q profiles is examined against PREPQC radiosonde measurements using the empirical averaging kernel method can be found in Sec. 6.2.

The retrievals are evaluated using four metrics: the retrieval yield, retrieval bias, root mean square differences (RMS), and the sampling bias, which will be elaborated below. The L2 quality control indicators for temperature and water vapor profiles are used. The name of each QC indicator is the same as the name of the corresponding parameter with "_QC" appended to the name.

- Retrieval yield profiles are calculated as the percent of successful retrievals (passing the QC) at each level/layer.
- Retrieval biases and RMSE are calculated using equation below where both the truth and the retrievals are filtered with the L2 QC indicators.
  For temperature:

$$bias \equiv mean[x_{data}(QC) - x_{truth}(QC)]$$
$$RMSE \equiv \sqrt{mean\{[x_{data}(QC) - x_{truth}(QC)]^2\}} \quad (5.1.1)$$

  For water vapor:

$$bias \equiv mean[x_{data}(QC) - x_{truth}(QC)]/mean[x_{truth}(QC)] \times 100\%$$
$$RMSE \equiv \sqrt{mean\{[x_{data}(QC) - x_{truth}(QC)]^2\}}/mean[x_{truth}(QC)] \times 100\% \quad (5.1.2)$$

- Sampling biases are caused by the cloud-state-dependent sampling of the infrared sounding instrument (Yue et al. 2013) and the sensitivity to the surface conditions (Yue and Lambrigtsen 2017). It is defined as the following equations.
  For temperature:

$$Sampling_bias \equiv mean[x_{truth}(QC)] - mean(x_{truth}) \quad (5.1.3)$$

  For water vapor:

$$Sampling_bias \equiv \{mean[x_{truth}(QC)] - mean(x_{truth})\}/mean(x_{truth}) \times 100\% \quad (5.1.4)$$

5.1.1 Pixel-Scale Evaluations Against ECMWF Model Analyses: Yield, Retrieval Bias, and Sampling Bias Characteristics

Contributor: Qing Yue, Evan Fishbein, Bjorn Lambrigtsen

Data

The water vapor and temperature vertical profiles from AIRS V6 and V7 (labeled as V6.6) Level 2 (L2) support products are studied for both the IR-only and IR+MW retrievals. Data from January and July of 2003 and 2011 from AIRS retrievals are analyzed by compositing retrievals based on surface types and cloud conditions. Retrieval yield profiles are calculated using the quality control indicators (QC) reported in the L2 product. The differences of both final retrieval and the first guess against the collocated European Centre for Medium-Range Weather Forecasts (ECMWF) model analysis are calculated for various ensembles and regions. Note that ECMWF model analysis is used in this study as the reference data. For simplicity of presentation, we use the term *bias* when referring to the difference between retrieval and ECMWF data.

Both the first guess and the final retrievals of the temperature profiles are reported at 100 pressure levels. The water vapor profiles are reported in the L2 product as the specific humidity on pressure levels (mass fraction of water vapor in moist air in units of kg/kg) and the layer mass density (in units of water molecules per square meter per layer). The water vapor first guess is a layer mass density reported in pressure layers. The AIRS V6 and V6.6 algorithms use a Stochastic Cloud Clearing/Neural Network (SCCNN) based first guess, which derives the training coefficients from matched observed infrared and microwave radiances with selected collocated ECMWF analyses. For more details, readers may refer to Milstein and Blackwell (2015).

Methodology

The retrievals are evaluated using four matrices: the retrieval yield, observation bias, root mean square differences, and the sampling bias, which will be elaborated below.

First the ECMWF model analysis is collocated with AIRS data at pixel scale (Fishbein, 2018). It serves as a benchmark for this study, although the ECMWF model analysis is itself subject to various uncertainties, especially over areas with sparse observations. The pixel-scale collocation between ECMWF and the satellite Field of View (FOV)/pixel is important because it removes biases that arise from satellite orbital sampling in the presence of a diurnal cycle, and the inherent "smoothing out" of pixel-scale variability in a gridded data set. Moreover, QC flags in the AIRS L2 data are used to subsample the collocated ECMWF model analysis to provide exactly the same samples from AIRS and ECMWF.

The QCs in the AIRS L2 product are applied individually to temperature and water vapor retrievals profile by profile. As a result, three data sets are obtained: (a) AIRS retrievals including both the physical retrieval and FG, (b) collocated ECMWF data subsampled to the retrieval QC (EC-S), and (c) collocated ECMWF data not subject to retrieval QC (EC). The retrievals and EC-S datasets have the same sample sizes taken from the same times and locations, thus any difference between them suggests either an observational bias in the retrieval, a model bias in the ECMWF data, or some combination of both. Differences in the EC-S and EC datasets are the result of sub-sampling by the AIRS retrieval QC, and are due to sampling effects that are a function of cloud state within the AIRS FOV, surface properties, etc (Yue et al. 2013; Wong et al. 2015). Therefore, statistics for ensembles based on effective cloud fraction (ECF) ranges, and surface classifications are also derived.

The ECMWF specific humidity is interpolated to the 100 AIRS support product pressure levels. The layer quantities of both SCCNN and physical retrieval of AIRS water vapor layer density profiles are converted from pressure layer to level by a linear interpolation of log(layer quantity) on the log(pressure) followed by a unit conversion from molecules/m^2 to g/kg. The level

quantities in the L2 products (H2OMMRLevSup) are also evaluated, which produces a similar result when comparing with layer-to-level conversion in this study except near the tropopause where sharp vertical gradients occur. Since SCCNN water vapor is also reported on layers instead of levels, results for humidity profiles in this report are based on layer quantities unless mentioned otherwise.

The surface classification (SurfClass) in the AIRS L2 IR+MW product is used to composite profiles (Table 4.1.1). It is part of the microwave retrieval step developed by Phil Rosenkranz. Four types of surfaces are studied: non-frozen land (SurfClass =1), non-frozen ocean (SurfClass =2), frozen land (SurfClass =5, 6, or 7), frozen ocean (SurfClass =3 or 4).

Results

1) Retrieval Yield

1a) Global mean vertical profiles of yield

Figure 5.1.1 shows the retrieval yield profiles over the four surface types for different ECF ranges. The ECF from AIRS V6.6 IR+MW retrievals is used to create the plot; however, statistics presented here are not sensitive to the choice of AIRS ECF data versions. The top panels are yield with QC=0 and 1 (good), and the bottom ones are QC=0 (best). The name of each QC indicator is the same as the name of the corresponding parameter with "_QC" appended to the name. QC=0 indicates the highest quality retrieval; QC=1 indicates good quality retrievals, where data may be used for studies with temporal and/or spatial averaging; and QC=2 indicates the use of such data is not recommended.

The yield decreases as ECF increases because of the limited information content from the infrared spectra on atmospheric states under cloudy conditions (ECF>70%). The AIRS retrieval yield has a clear dependence on height, reflecting partial profile retrievals above cloud top. The temperature and water vapor profiles have consistent QCs and only temperature retrieval yield is shown in Figure 5.1.1. Near surface, the new versions have lower QC=0&1yield than V6 but higher QC=0 yield.

AIRS yield has a clear dependence on surface types. Over frozen surfaces, the yield over clear sky (ECF < 0.2) is lower than bins with 0.2<ECF<0.6, which is related to clouds that were not detected and the larger uncertainty in cloud clearing over the frozen surfaces (See Figure 5.1.11 and related discussions).

1b) Spatial distributions of yield

The maps of yield at different pressure levels are shown in Fig. 5.1.2 using the data in January 2003. ECF is also shown to highlight the correspondence between the drop of yield and increase of cloud fraction. A large reduction of yield below 700hPa (near the surface) over the high latitude land area in the cold season is also clear.

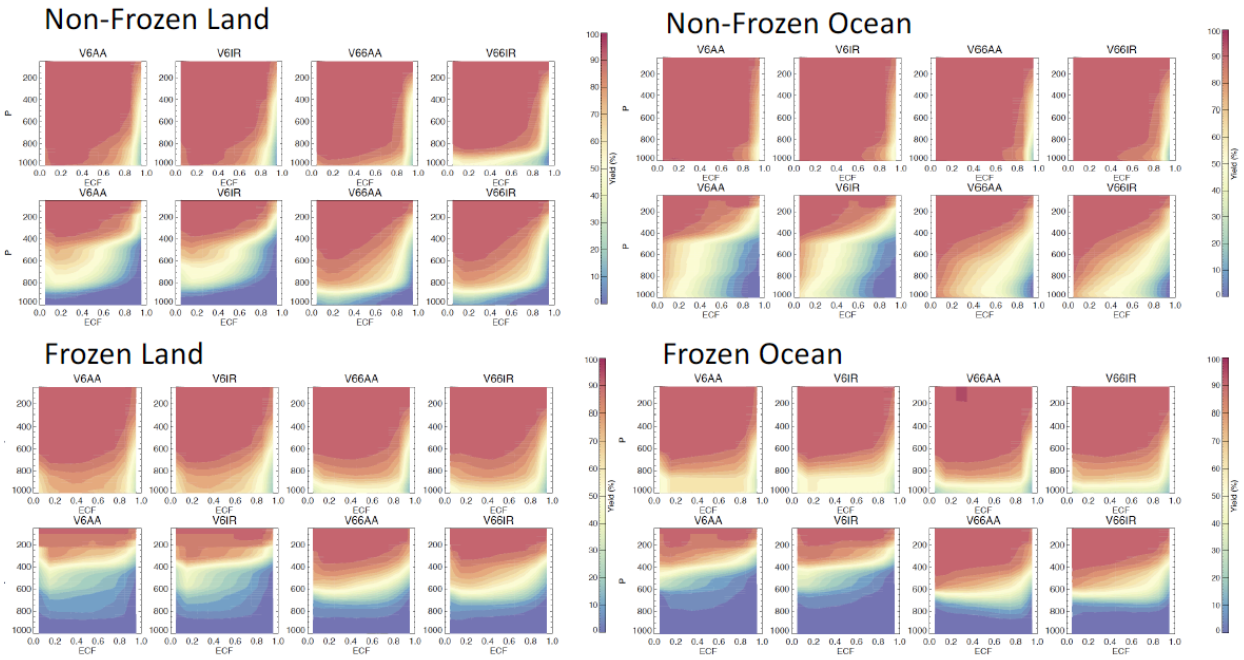


Figure 5.1.1. Retrieval yield over non-frozen and frozen ocean and land surfaces as a function of effective cloud fraction (ECF) bins. ECF from AIRS V6.6 IR+MW retrieval is used. In each plate, the top panels are for QC=0 and 1, and the bottom panels are for QC=0.

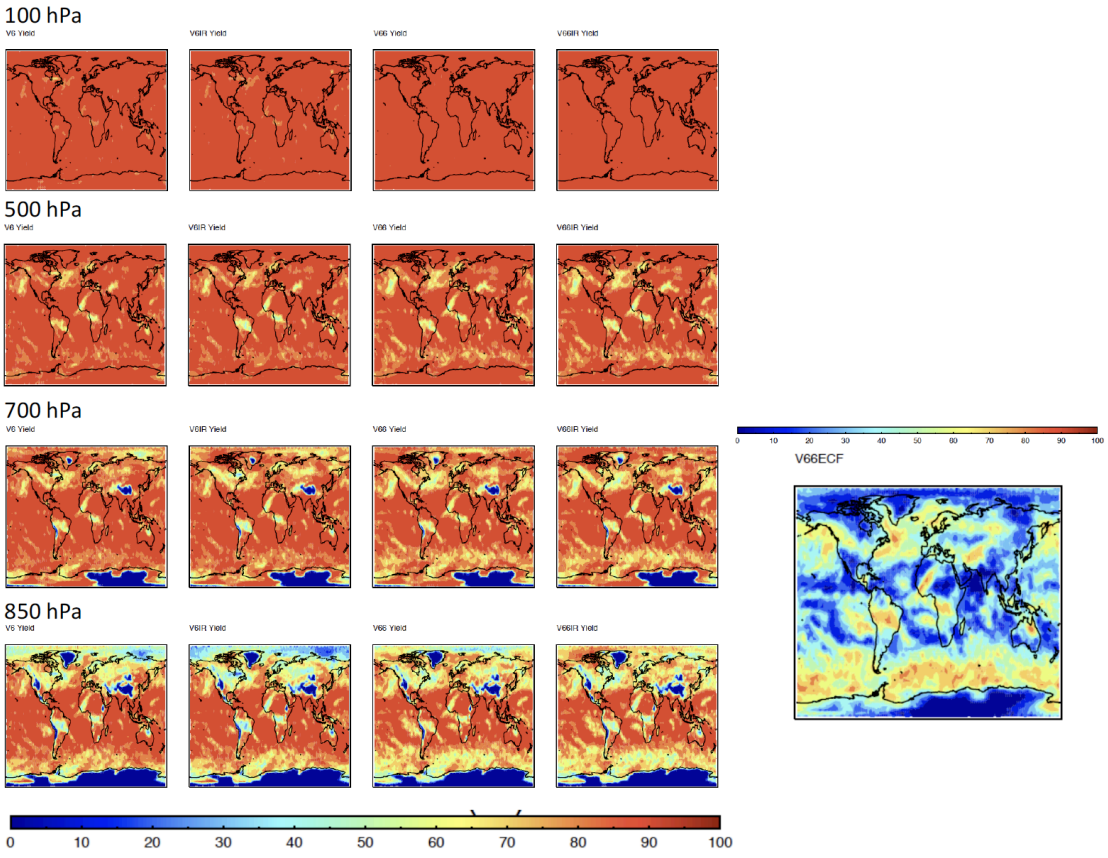


Figure 5.1.2. Spatial distributions of retrieval yield for the four AIRS algorithms at four pressure levels: 100, 500, 700, 850 hPa. Maps are produced using data from January 2003. ECF (%) from V6.6 IR+MW retrieval is also shown.

2) Mean vertical profiles of observation and sampling biases calculated against ECMWF

As discussed in Yue et al. (2013), total retrieval bias is the summation of the observational bias and the sampling bias. In order to track the changes of SCCNN in different retrievals and the differences between the retrieval and the first guess from SCCNN, Figures 5.1.3-5.1.6 show the mean profiles of the biases (observational and sampling) calculated against ECMWF model analyses for the four surface types binned by ECF, respectively. For temperature, the unit of the results is Kelvin (K). For water vapor profiles, relative differences are calculated and the unit is %. Table 5.1.1 shows the frequency of occurrence for the four surface types plus the coastal and undetermined surface classes from AIRS 07/2003 retrievals. Note that ~75% of the data are from non-frozen surfaces. Moreover, ECMWF model analyses may have higher uncertainties in the high latitude regions due to lack of observational data assimilated and model limitations. Therefore, *the results over frozen surfaces presented here focus on tracking the changes among the retrievals instead of the absolute magnitude of biases in AIRS products*.

The observational biases and sampling biases calculated against ECMWF are largely dependent on the cloud state within the field of view over all four surface types. Over non-frozen surfaces, the mean bias magnitude is generally very small (magnitude <1K for temperature and < 15% for water vapor), but increases to -1K and -20% when ECF>0.8. The reduction of the bias magnitudes in new versions is apparent, especially in the V6.6 IR+MW products at larger ECF.

Table 5.1.1. Frequency of occurrence for different surface types used to categorize the data. The statistics are based on the 2003 July data from the AIRS V6.6 IR+MW algorithm.

Surface Type	Non-frozen Land	Non-Frozen Ocean	Frozen Land	Sea Ice	Coastal	Undetermined
Frequency %	22.96	52.02	10.69	11.42	2.91	~0

Over frozen surfaces, the magnitude of the bias in all four versions is larger than that over the non-frozen surfaces. Compared to ECMWF, the new AIRS retrievals produce colder temperatures near surface and much higher humidity profiles (20% wetter than ECMWF) than V6, which are significantly larger at smaller ECFs. These features are also seen in the SCCNN humidity profiles, although the physical retrieval reduces the magnitude of biases above 500 hPa compared to SCCNN. These characteristics on cloud dependence indicate potential issues with cloud clearing over frozen surfaces. Figure 5.1.11 shows a comparison between the AIRS V6 ECF and cloud fractions determined from CloudSat and CALIPSO data (2B-CLDCLASS-Lidar) that is collocated with AIRS data in the Arctic region. Results are based on four years of observations from January 2003 to December 2010 (Peterson et al. 2020). Although infrared ECF is defined differently from cloud fraction determined from active measurements, these results can be used to infer the cloud detection difference between AIRS and CloudSat/CALIPSO. It is clear that over sea ice surfaces, AIRS detects a high percentage of clear conditions (ECF = 0) whereas CloudSat/CALIPSO data indicate a high frequency of overcast conditions. It is possible that the underestimation of AIRS ECF over the frozen surfaces may cause the overestimation of the atmospheric humidity profile by AIRS, which warrants further comparison with in-situ measurements in these regions.

Figures 5.1.7-5.1.10 show the RMS differences between AIRS and ECMWF for the four surface types binned by ECF, respectively. For temperature, the unit of the results is Kelvin (K). For water vapor profiles, relative differences are calculated and the unit is %. Only results using January 2003 data are shown. The temperature RMS increases significantly below 800hPa which is related to larger temperature variability within the boundary layer. The water vapor RMS increases with altitude over non-frozen surfaces from <10% near the surface to larger than 30% above 600 hPa due to a decreasing magnitude of water vapor concentration with altitude. However, a reduction of RMS by the physical retrieval comparing to the SCCNN profiles is apparent in the free troposphere for both temperature and water vapor when ECF is less than 0.7. The V6.6 IR+MW retrievals produce the smallest RMS among all four AIRS retrievals except for the humidity profiles over the frozen surfaces, where similar magnitude of humidity RMS is seen in V6.6 or even larger in V6.6 IR-only than previous versions.

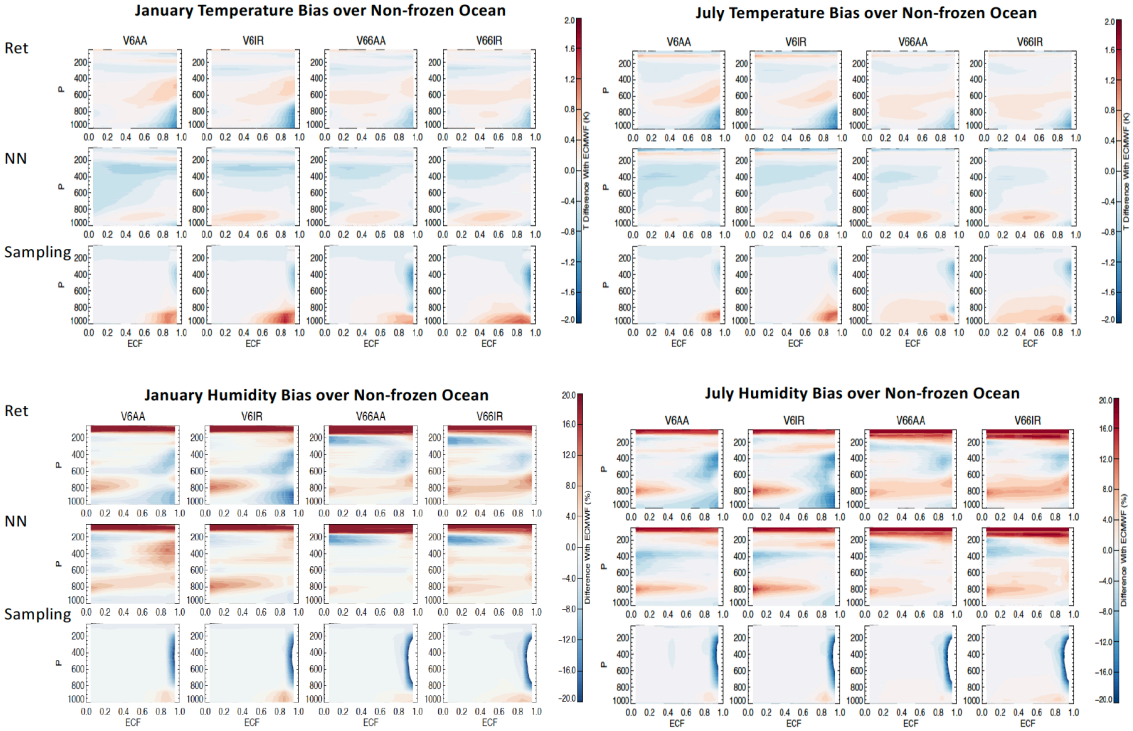


Figure 5.1.3. The observation bias (first row), NN bias (second row), and the sampling bias (third row) for temperature and humidity profiles over non-frozen ocean. Each column represents results from different retrievals. Results using data in January 2003 are on the left and July 2003 data on the right. The top plate is for temperature (K) and the bottom plate is for water vapor (%). The water vapor layer quantities in the L2 product are converted to level specific humidity by simple log(pressure) interpolation. Differences in humidity are calculated as mean(satellite – ECMWF)/mean(ECMWF)x100%.

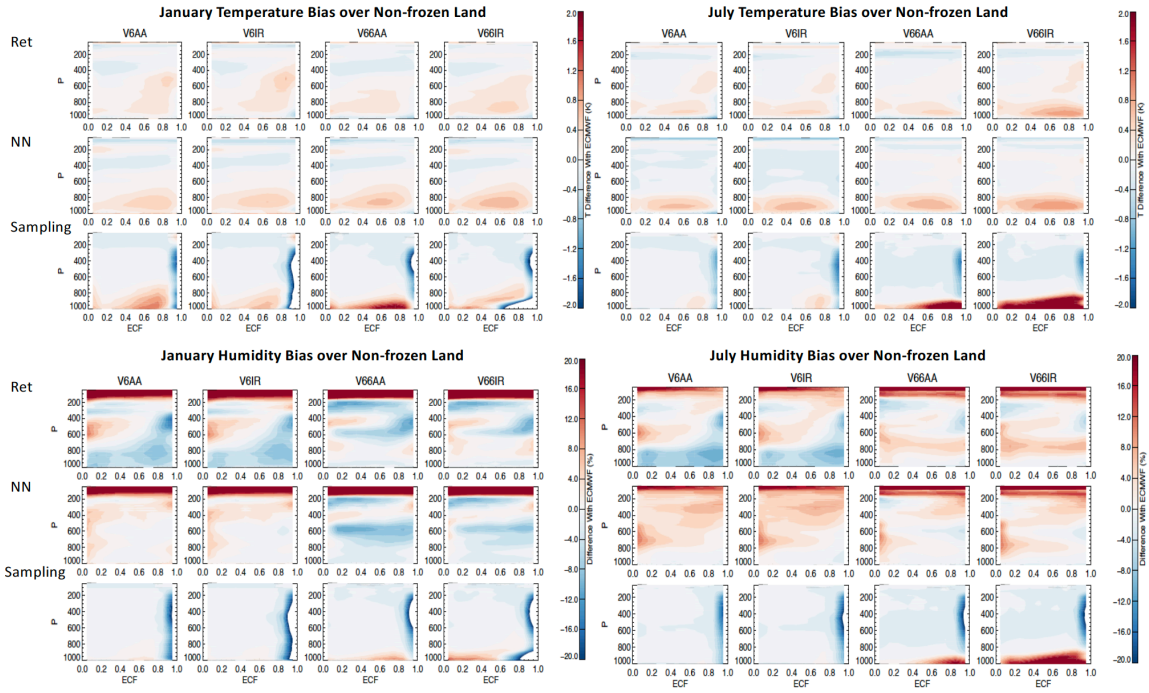


Figure 5.1.4. Similar to Figure 5.1.3, but showing results over non-frozen land instead.

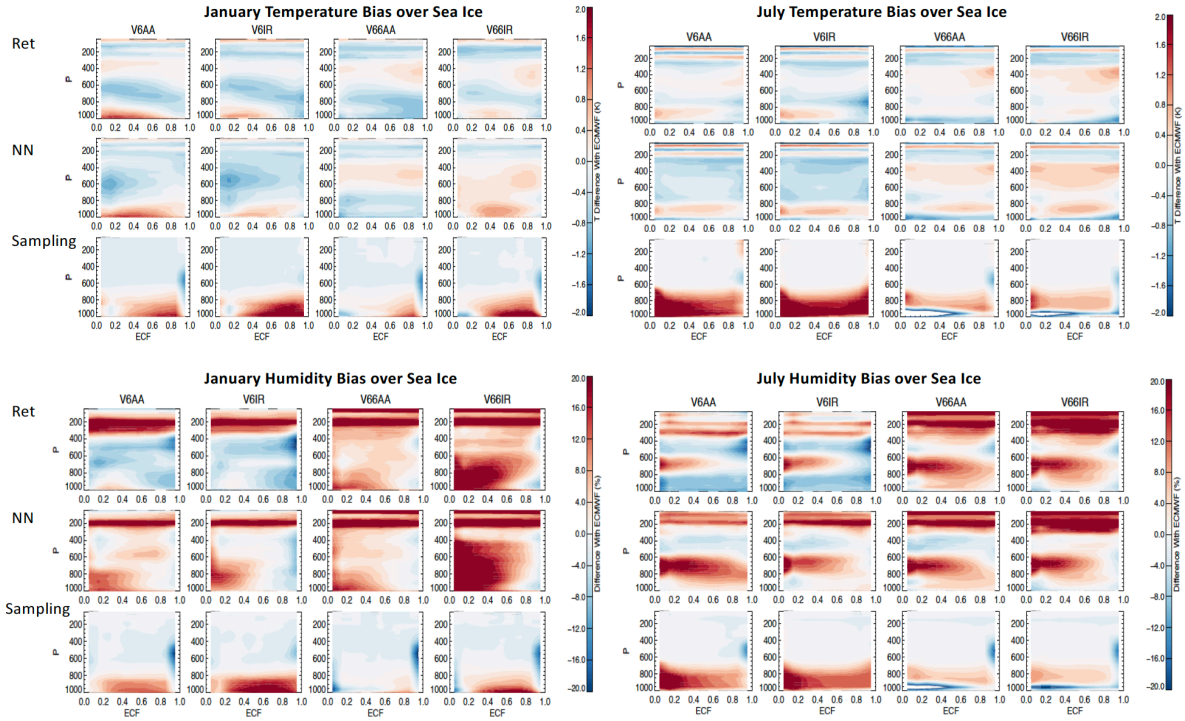


Figure 5.1.5. Similar to Figure 5.1.3, but showing results over sea ice (frozen ocean) surfaces instead.

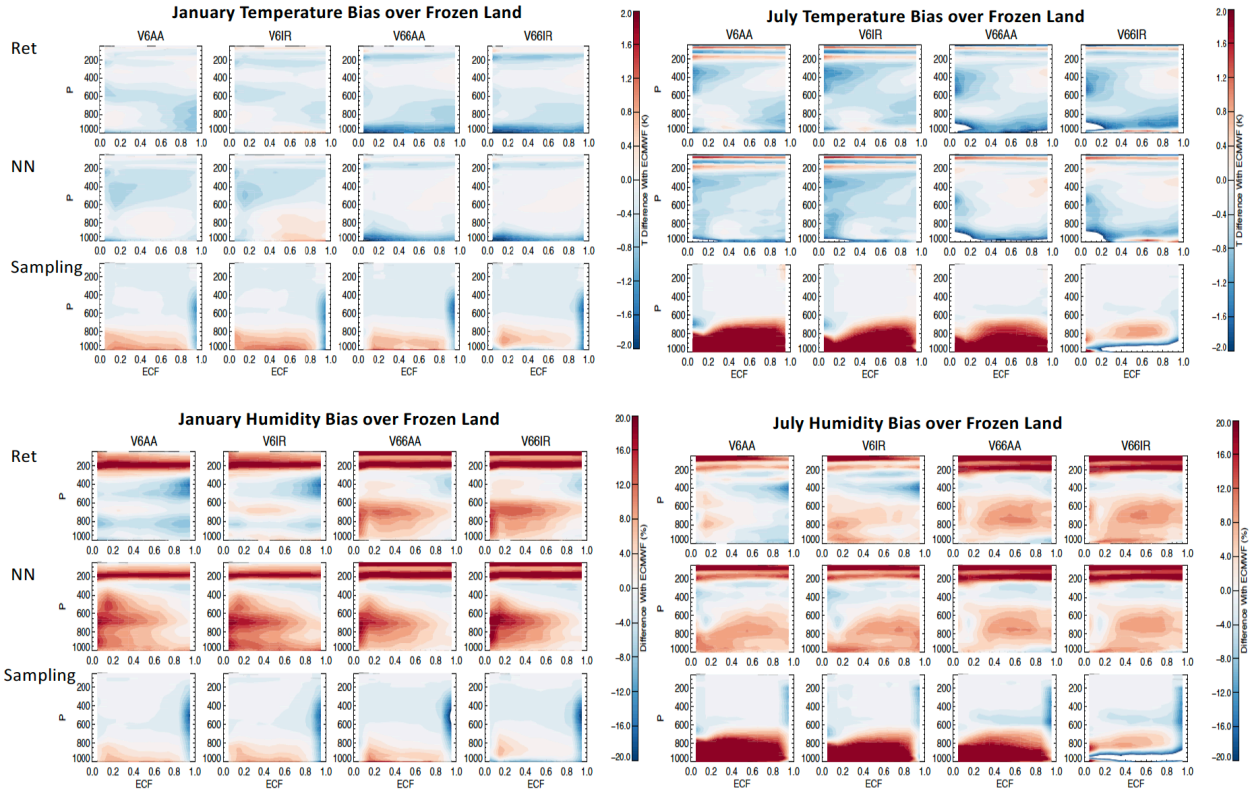


Figure 5.1.6. Similar to Figure 5.1.3, but showing results over frozen land surfaces instead.

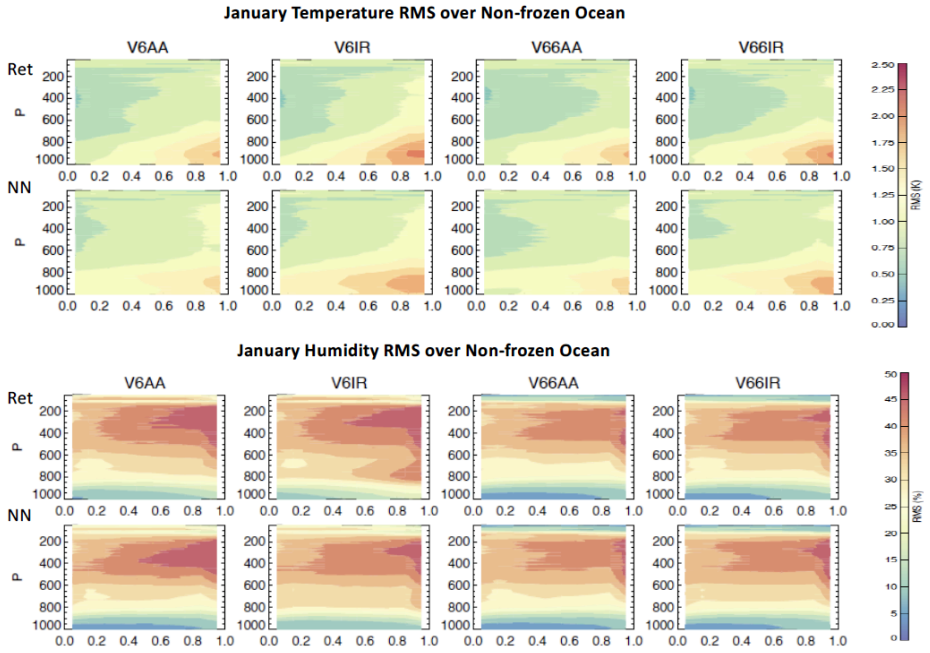


Figure 5.1.7. The RMS error of AIRS temperature and humidity profiles in January 2003 from different AIRS retrievals over non-frozen ocean. Results for final retrieval is shown in the first row and NN (initial guess) in the second row. Each column represents results from different retrievals. The top plate is for temperature (K) and the bottom plate is for water vapor (%). The water vapor layer quantities in the L2 product are

converted to level specific humidity by simple log(pressure) interpolation. RMS in humidity are calculated as RMS(satellite – ECMWF)/mean(ECMWF)x100%.

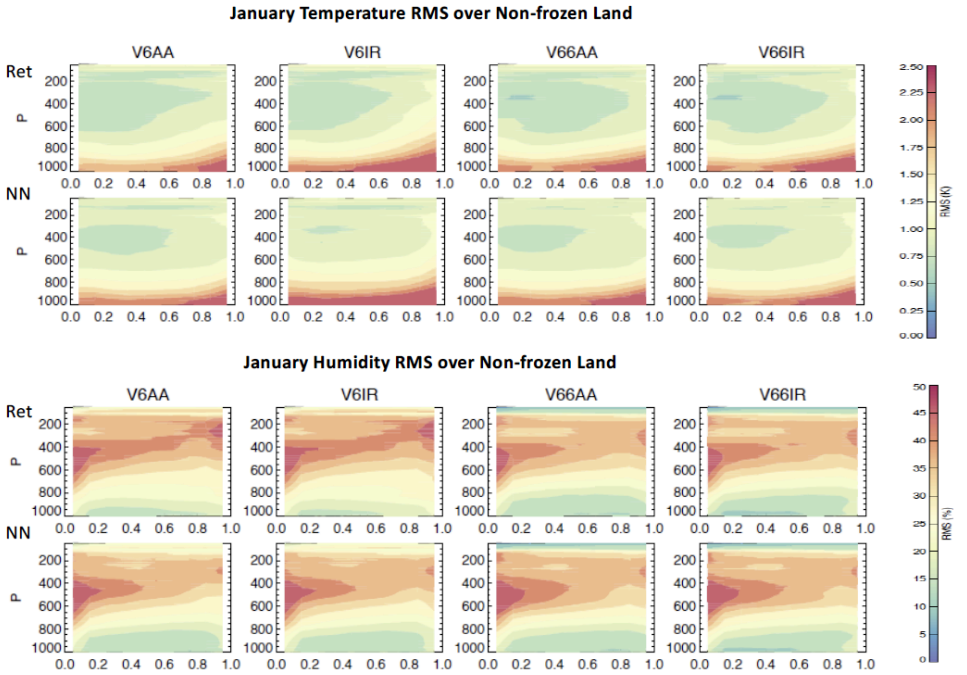


Figure 5.1.8. Similar to Figure 5.1.7, but showing results over non-frozen land surfaces instead.

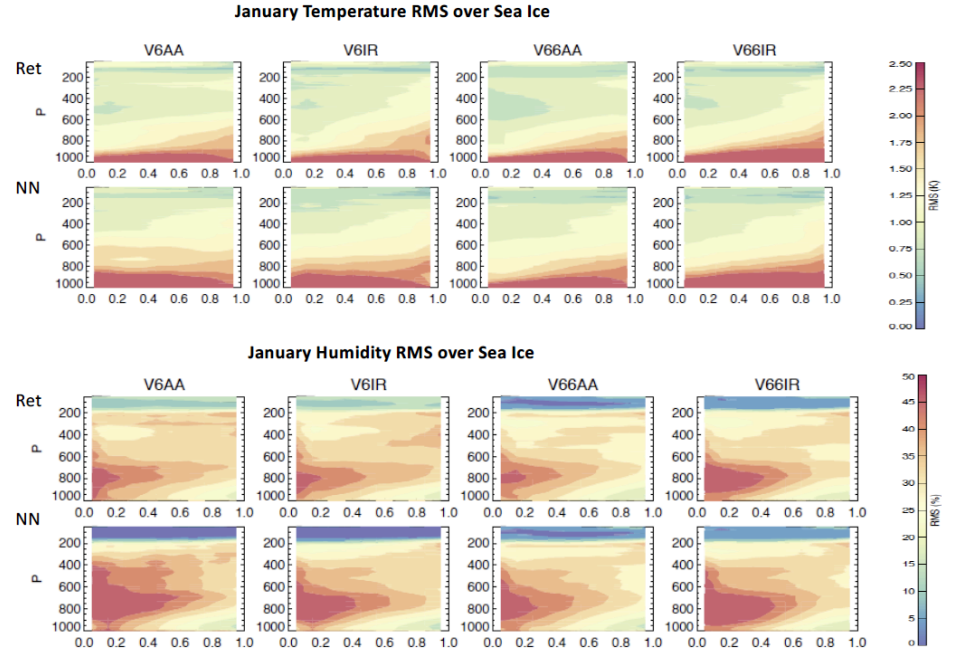


Figure 5.1.9. Similar to Figure 5.1.7, but showing results over sea ice surfaces instead.

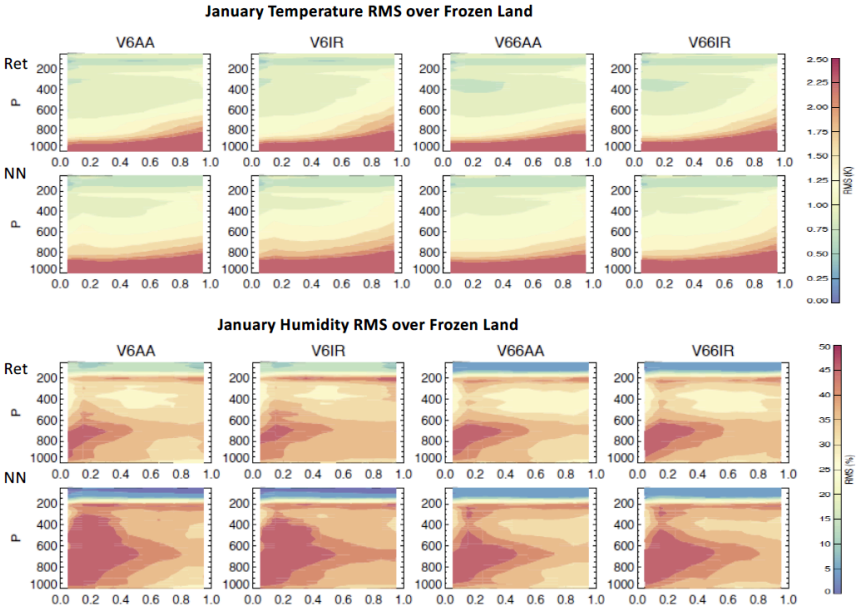


Figure 5.1.10. Similar to Figure 5.1.7, but showing results over frozen land surfaces instead.

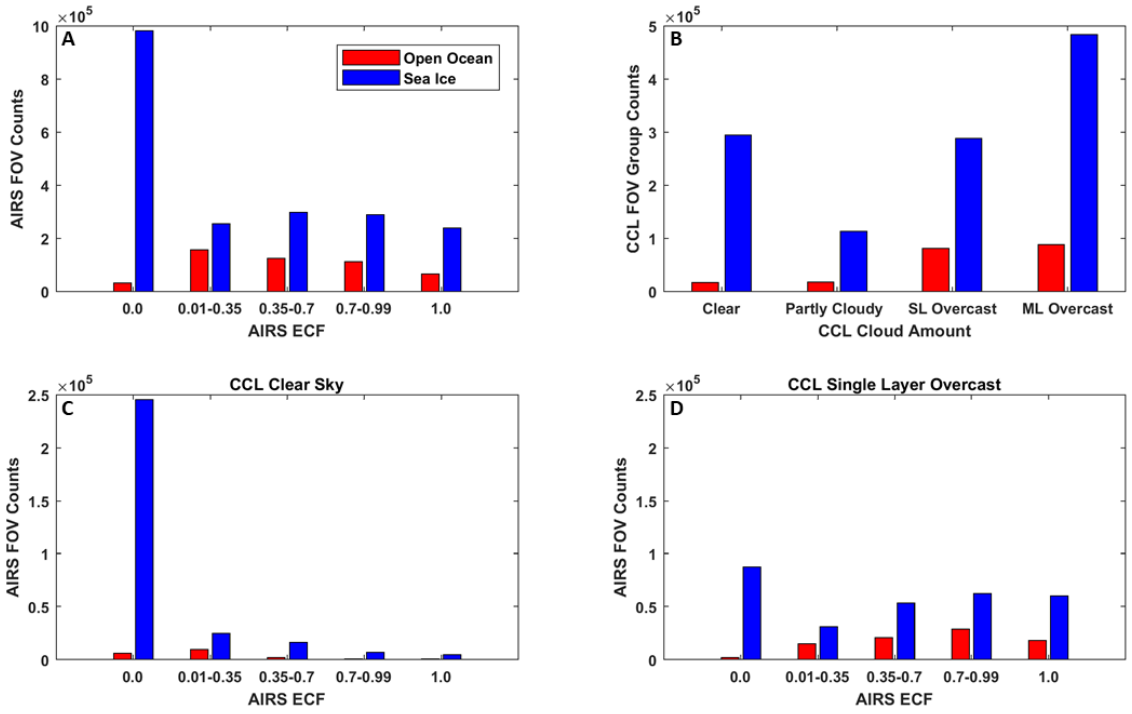


Figure 5.1.11. Number of pixels in the Arctic (poleward of 60N) binned by V6 IR+MW AIRS ECF data (A) and cloud fraction by CloudSat and CALIPSO 2B-CLDCLASS-Lidar products collocated with AIRS footprints (CCL: B). Results are generated using data from 01/2006 and 12/2010. The three bins used for the CCL cloud amounts are defined as below: clear if >90% of CCL pixels within the collocated AIRS footprint are clear; partly cloudy if 10~90% of CCL pixels within the collocated AIRS footprint are cloudy; Overcast if >90% of C-C pixels within the collocated AIRS footprint are cloudy. Results over nonfrozen ocean (open ocean) and sea ice are shown separately with different colors. (C) Same as (A) except that

only AIRS FOVs corresponding to CCL FOV clear-sky groups are included. (D) Same as (C) expect for the CCL FOV group being single layer overcast conditions by CCL. More details see Peterson et al. (2020).

5.1.2 Pixel-Scale Evaluations Using IGRA Radiosondes

Contributor: Sun Wong

Data

V6 and V7 retrievals of temperature (T) and specific humidity (q) from both the IR+MW and IR-only retrieval algorithms are validated against the radiosonde measurements from Integrated Global Radiosonde Archive (IGRA). In the following discussions, *the V7 retrievals with AMSU microwave channels are named V6.X, while the one with IR-only channels are named V6.XIR*. The retrievals of January and July of 2003 and 2011 are analyzed. This report will focus on the comparison between 2003 July and 2011 July retrievals over the European continent.

The IGRA data is used as reference because of their wide area coverage over the European continent (see Fig. 5.1.12). Caution needs to be taken to interpret specific humidity data above freezing point because the archive includes those radiosondes that may fail in freezing condition.

In late July of 2003 a heat wave occurred over the European continent while the climate in the July of 2011. Comparing the T and q soundings of these two years provides an opportunity for testing AIRS retrievals against extreme weather conditions associated with the heat wave. Biases and root-mean-squared errors (RMSEs) of 4 versions of AIRS T and q retrievals (V6, V6IR, V6.X, and V6.XIR) from collocated radiosonde measurements are stratified against near-surface temperatures (measured by the radiosondes).

Methodology

Most IGRA radiosondes are operational and launched at 0 and 12 UTC; therefore, they will be matched within a few hours when Aqua passes Greenwich Median or International Date Line. The temporal tolerance of our data matching is 2 hours, and the spatial tolerance is 100 km.

Biases and RMSEs are computed for the AIRS retrievals and NN temperature and specific humidity profiles against measurements of the matched radiosondes. For q, we report the relative RMSEs, computed as the RMSEs divided by the mean radiosonde measured q. Then the biases and RMSEs as well as the actual soundings are binned according to the near-surface T measured by the radiosondes and averaged in 10°-40°E and 30°-50°N (the box shown in Fig. 5.1.12) for July in 2003 and 2011.

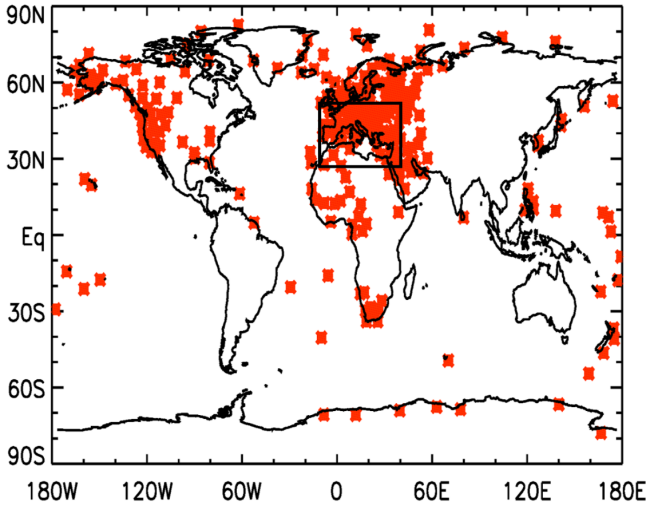


Figure 5.1.12 Global distribution of the IGRA radiosonde launch sites. The black box encircles the area influenced by the 2003 July heat wave, a focus of the study in this section.

Results

Profiles of q from the radiosonde measurements and the 4 different version of AIRS retrievals are shown in Fig. 5.1.13, binned in different ranges of near-surface temperatures for July 2003 and 2011. In a normal summer like 2011, the retrieved q are very close to the radiosonde measured values, while in the summer influenced by a heat wave (2003 when near-surface T > 300 K, or the right 3 panels in the top of Fig. 5.1.13), all versions of q retrievals tend to yield a moister condition than the radiosonde measurements in the lower troposphere (below ~700 hPa). The q RMSEs below 800 hPa for near-surface T range of 300-310K changes from 25-35% in 2011 to 30-40% in 2003 (Fig. 5.1.14). The newer retrievals (V6.X and V6.XIR) tend to have smaller q RMSEs except during the July of 2003 when near-surface T is greater than 280 K (when the heatwave was about to kick in). In the middle troposphere (500-600 hPa), newer retrievals always have smaller RMSEs. In the heatwave condition (left 3 panels in the top of Fig. 5.1.14), newer retrievals have larger RMSEs than older retrievals in the boundary layer (below 700 hPa).

Figure 5.1.15 shows how retrievals correct RMSEs from the NN for the specific case when near-surface T is in the range of 300-310 K. In the middle troposphere (500-600 hPa), older versions of retrievals (V6 and V6IR) run away from the NN and result in higher RMSEs. The newer versions improve this behavior and reduce RMSEs from the NN. However, in the lower troposphere (below 700 hPa), older retrievals tend to correct the large RMSEs introduced by the NN during the heatwave (top panel), while the newer versions keep sticking to the NN. Further investigation indicates that newer versions have much smaller verticality in the lower troposphere.

Profiles of T retrievals do not show influences by the heatwave as seen in Fig. 5.1.16, in which the behaviors of retrievals are similar in 2003 and 2011. Therefore, in Fig. 5.1.16, we plot biases and RMSEs with combined samples from both July in 2003 and 2011. The 4 versions of retrievals are very close to each other in the middle troposphere. In general, the biases are less than 1 K and the RMSEs are about 1 K. In the lower troposphere, newer versions (V6.X and V6.XIR) have RMSEs smaller than those of V6IR, but larger than those of V6. When the near-surface T is too cold (below 285 K, or the two left columns in Fig. 5.1.16), cold biases (and hence RMSEs) increase from their warmer counterparts. From Fig. 5.1.15, we can see that the cause of these cold biases arise from 800-900 hPa, where the lapse rate is too steep in the retrievals compared to the radiosonde measurements.

Conclusions

V6.X and V6.XIR improve the T and q retrievals in the middle troposphere (300-600 hPa) by reducing errors in the NN and have smaller errors than the older versions of retrievals (V6 or V6IR). In the lower troposphere (below 700 hPa), smaller verticality in these newer retrievals makes T and q profiles very close to the NN profiles. This avoids the retrievals running away from the NN as in V6 or V6IR. However, if extreme events occur, the failure of NN to capture the climate extremes may result in larger biases and RMSEs of the newer retrievals. During the 2003 heatwave over Europe, while the T retrievals are fine in all versions, the newer retrievals have much moister lower atmosphere than the radiosonde measurements and V6 or V6IR. The corrections to such q biases introduced by the NN during this event are more significant in V6 and V6IR than in V6.55 and V6.55IR, because the newer versions tend to limit the corrections with their lower verticality. Therefore, caution needs to be taken for using the newer versions in some application (e.g., using vapor pressure depression in prediction of draught).

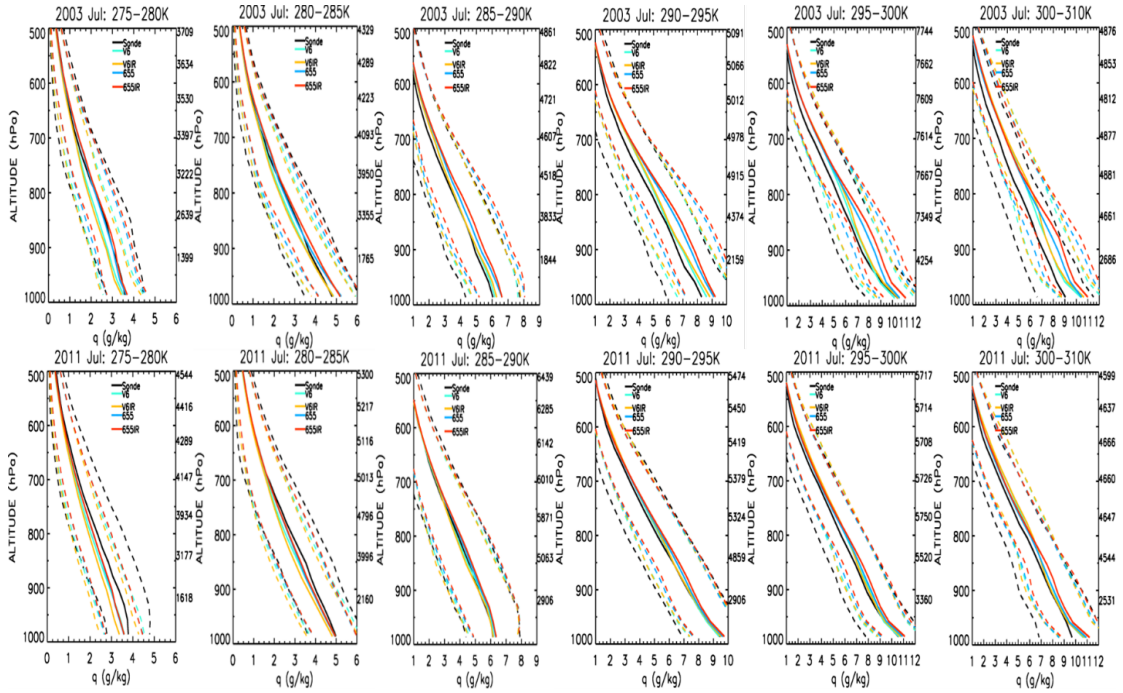


Figure 5.1.13 Specific humidity (q, in g/kg) profiles averaged over European continent (10°-40°E, 30°-50°N) for July of 2003 (top) and 2011 (bottom). The profiles are plotted in different ranges of near-surface temperatures as shown in the titles of each panel. The profiles from IGRA radiosonde measurements are in black, V6 retrievals in turquoise, V6IR retrievals in yellow, V6.X retrievals in blue, and V6.XIR in red. The corresponding 1-σ variations are in dashed lines. Numbers on the left axes are sample sizes. The V7 retrievals with AMSU microwave channels are named V6.X, while the one with IR-only channels are named V6.XIR.

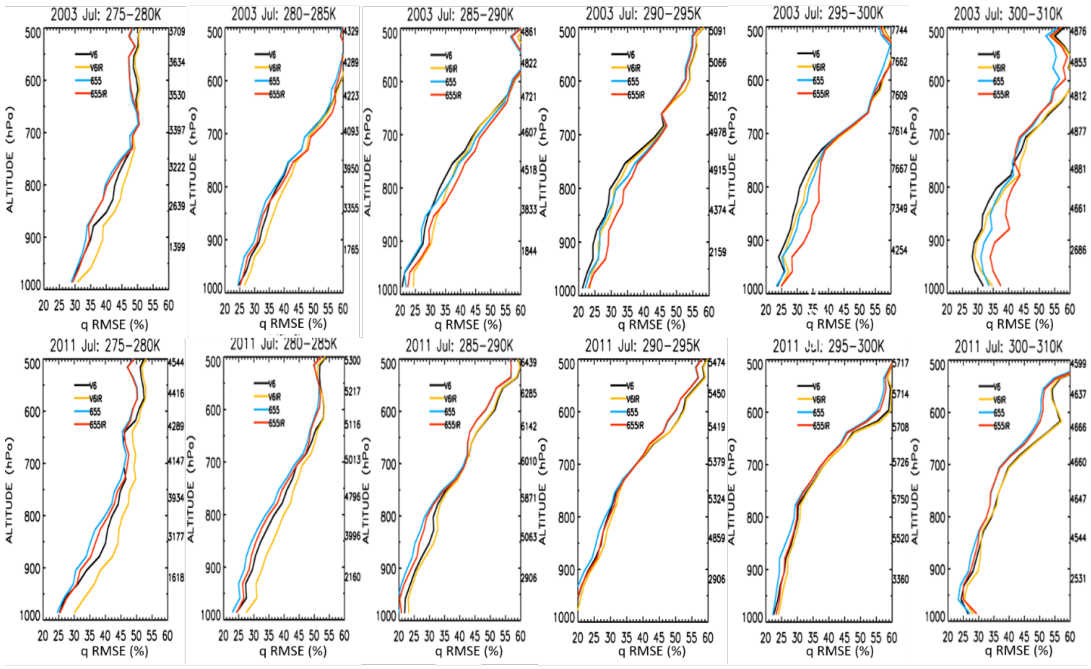


Figure 5.1.14 Specific humidity (q, in g/kg) RMSE profiles averaged over European continent (10°-40°E, 30°-50°N) for July of 2003 (top) and 2011 (bottom). The profiles are plotted in different ranges of near-surface temperatures as shown in the titles of each panel. The profiles for V6 retrievals in black, V6IR retrievals in yellow, V6.X retrievals in blue, and V6.XIR in red. Numbers on the left axes are sample sizes. The V7 retrievals with AMSU microwave channels are named V6.X, while the one with IR-only channels are named V6.XIR.

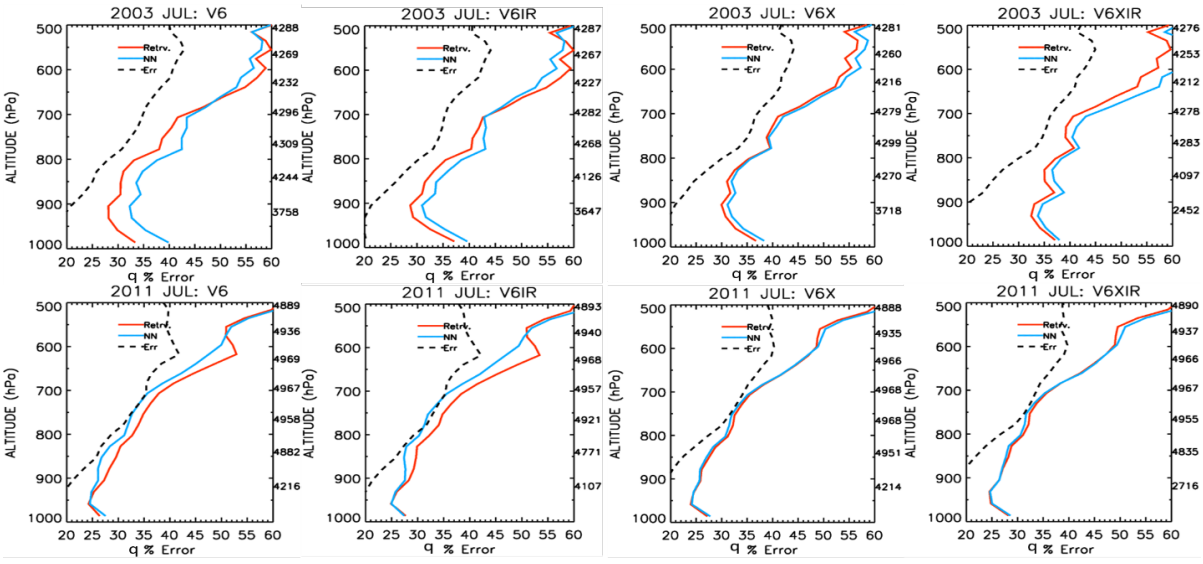


Figure 5.1.15 For the near-surface T in 300-310 K, q RMSEs from different versions (different columns) of retrievals (red) and NN (blue). The dashed lines are the error estimates reported in the AIRS archive. The top panel is for July of 2003 and the bottom for July of 2011. Numbers on the left axes are sample sizes. The V7 retrievals with AMSU microwave channels are named V6.X, while the one with IR-only channels are named V6.XIR.

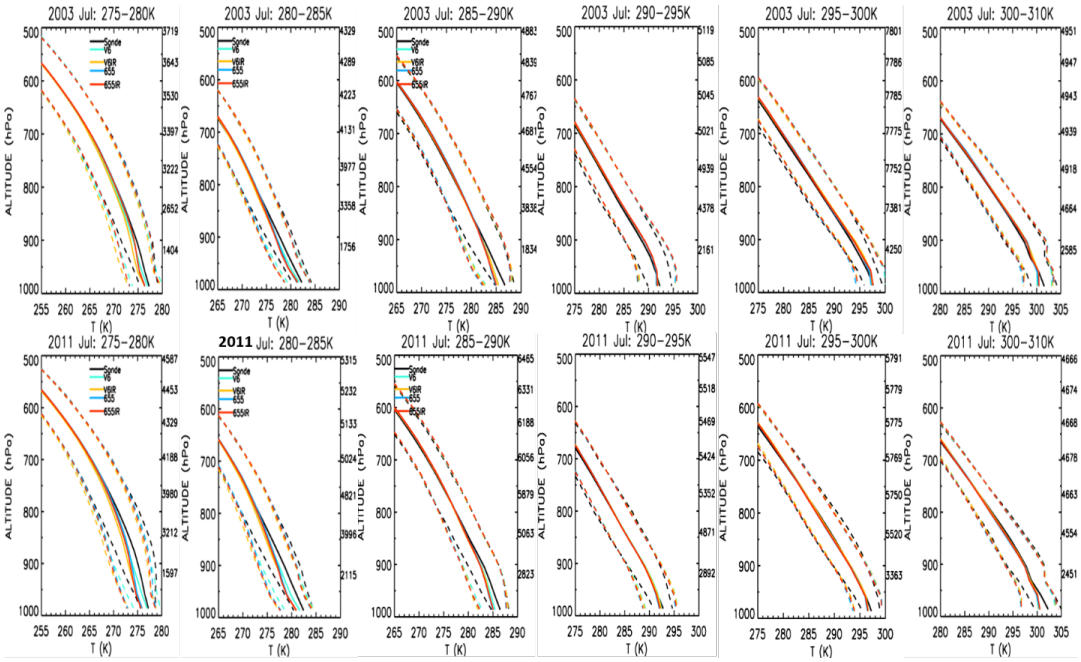


Figure 5.1.16 Similar to Fig. 5.1.13, but for T profiles (in K).

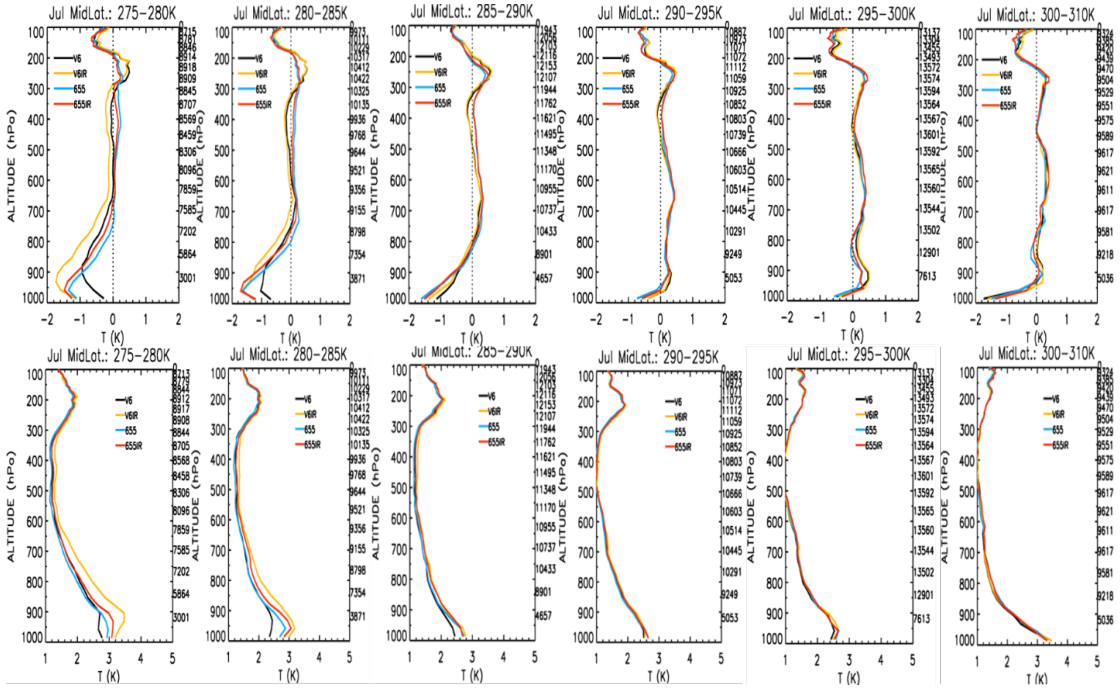


Figure 5.1.17 Biases (top) and RMSEs (bottom) of T retrievals binned according to ranges of near-surface T (shown in the title of each panel). The profiles for V6 retrievals in black, V6IR retrievals in yellow, V6.X retrievals in blue, and V6.XIR in red. Numbers on the left axes are sample sizes. The V7 retrievals with AMSU microwave channels are named V6.X, while the one with IR-only channels are named V6.XIR.

5.1.3 Comparisons with Radiosonde Measurements Taken During the MAGIC Field Campaign

Contributor: Peter Kalmus

Data

L2 TAirSup (henceforth **T**) and L2 H2OCDSup (which is converted to specific humidity in units of g/kg, henceforth **q**) from AIRS/AMSU (IR+MW) and AIRS (IR) are examined by comparing with the collocated ECMWF reanalysis and the radiosonde measurements from the Marine ARM GPCI Investigation of Clouds (MAGIC) campaign (Fig. 5.1.18).

We analyze approximately 550 vertical profiles for each variable (T and q) and each AIRS data type (airx2sup and airs2sup). Radiosondes from the ship-borne MAGIC campaign provide "truth" profiles for temperature and relative humidity (we convert the latter to specific humidity). The radiosonde manufacturer (Vaisala, model RS92-SGP) quotes two-sigma uncertainties of ±0.5 K and 5% for T and RH, respectively. The radiosondes provide T and RH to typical pressures of 15–30 hPa, at typical resolutions of 0.3–0.4 hPa.

The MAGIC campaign (9/2012–10/2013) included 19 round trips between Los Angeles and Honolulu. There is a data hiatus between January 2013 and May 2013 due to a ship refit, so most MAGIC data were taken in the months between June and December with some data in January and May. Radiosonde launches were attempted four times per day, near 5:30, 11:30, 17:30, and 23:30 UTC, except for Legs 14a and 14b which featured eight launch attempts per day. ECMWF data is also used, and is included in figures.

Methodology

The AIRS L2 retrievals are collocated with the MAGIC data with following criteria: temporally not more than 6 h before or after each MAGIC radiosonde launch time and spatially centered not more than 200 km from the radiosonde launch location. If more than one retrieval satisfies these constraints, the spatially nearest one is chosen. We find an AIRS L2 retrieval corresponding to the MAGIC radiosonde under these constraints for all but about 80 of the radiosonde launches.

We discard any matchups in which the L2 data quality is marked 2 for *either* data set (i.e. we exclude the matchup if the airs2sup QC equals 2 *or* if the airx2sup QC equals 2: dual QC). Results QCed by the corresponding QC flags in each product are also shown for comparison purpose. Comparison is carried out by:

a. Calculating profiles of mean bias and RMSE of all truth-AIRS pairs;
b. Calculating longitudinally-binned profiles of bias and RMSE, along the transect, of truth-AIRS pairs. In each bin, a mean bias and RMSE is calculated. This creates 2-dimensional longitude-pressure slices of bias and RMSE along the transect.

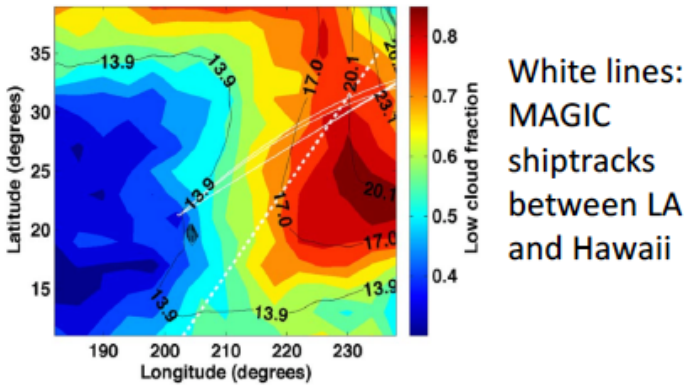


Figure 5.1.18. MAGIC data used in this analysis. Background contours are MODIS mean cloud fraction.

Conclusions

Over the MAGIC region, the differences between V6.0 and V6.55 AIRS-only retrievals are small, but v6.55 shows improvement in T and q mean bias at almost every pressure level, and in T and q RMSE in the boundary layer. T and q RMSE shows no change above the boundary level. The v6.55 retrievals produce a smaller boundary level T RMSE by ~0.5 K near the surface and decreasing gradually until the two versions show no RMSE difference at the top of the boundary layer, and a smaller boundary level q RMSE by ~0.5 g/kg near the surface and decreasing gradually until the two versions show no RMSE difference at the top of the boundary layer (Fig. 5.1.20). Improvements in the AIRS+MW products between the two versions are similar, but the RMSE improvements near the surface in v6.55 are about twice as large (Fig. 5.1.19). These improvements are also seen in mean plots binned by latitude across the transect (Figures 5.1.21-5.1.23). Indeed, v6.55 AIRS-only T and q retrievals show smaller biases around the inversion layer than even v6.0 AIRS+MW retrievals.

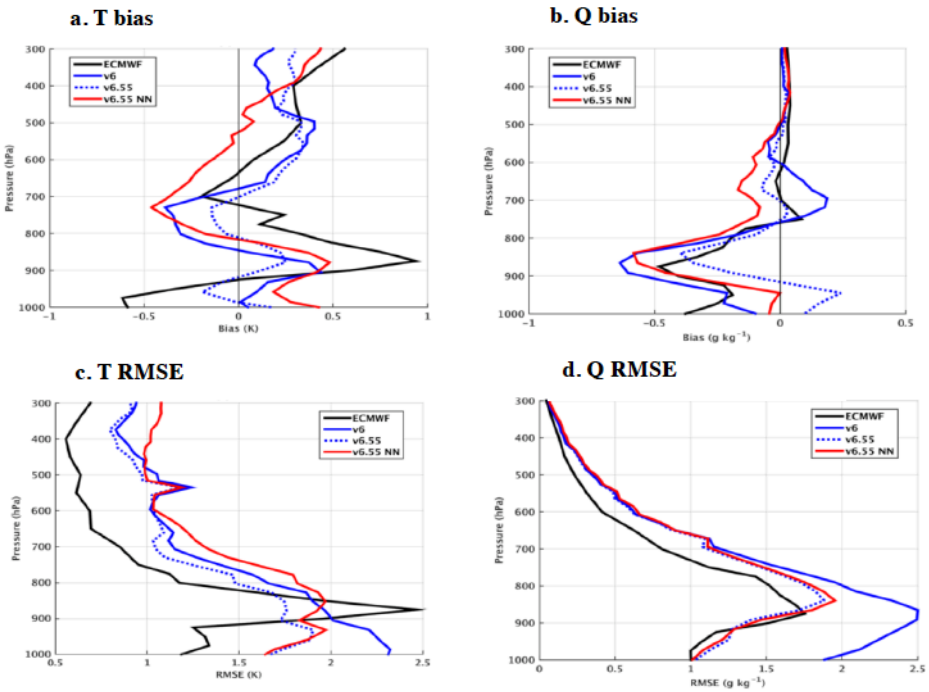


Figure 5.1.19. Comparisons with MAGIC sondes and ECMWF (Fig. 5.1.18 for location) for temperature bias (a) and RMSE (c); and water vapor bias (b) and RMSE (d**), for AIRS+MW retrievals**. Black lines are for ECMWF reanalysis. Blue lines are for V6 (solid) and V6.55 (dotted); red lines are for v6.55 SCCNN. Note that absolute difference in K and g/kg is shown.

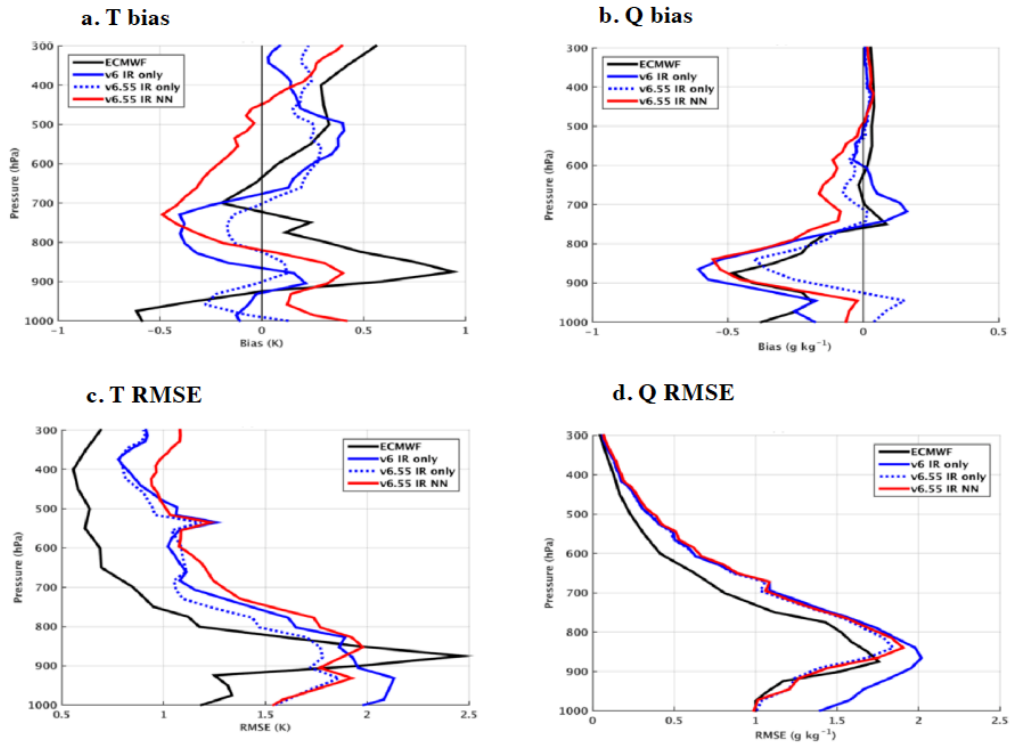


Figure 5.1.20. Comparisons with MAGIC sondes and ECMWF (Fig. 5.1.18 for location) for temperature bias (a) and RMSE (c); and water vapor bias (b) and RMSE (d), **for AIRS only retrievals**. Black lines are for ECMWF reanalysis. Blue lines are for V6 (solid) and V6.55 (dotted); red lines are for v6.55 SCCNN. Note that absolute difference in K and g/kg is shown.

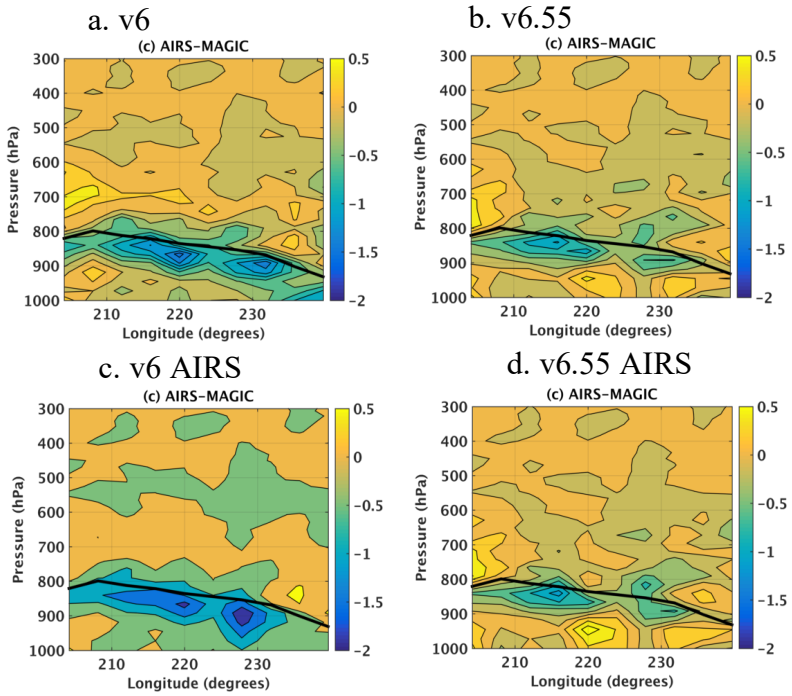


Figure 5.1.21. Comparisons with MAGIC sondes (Fig. 5.1.18 for location) for q bias for (a) v6 AIRS+MW; (b) v6.55 AIRS+MW (c) v6 AIRS only; (d) v6.55 AIRS only. Black lines show boundary layer top estimated from radiosonde temperature inversions. Note that absolute difference in g/kg is shown.

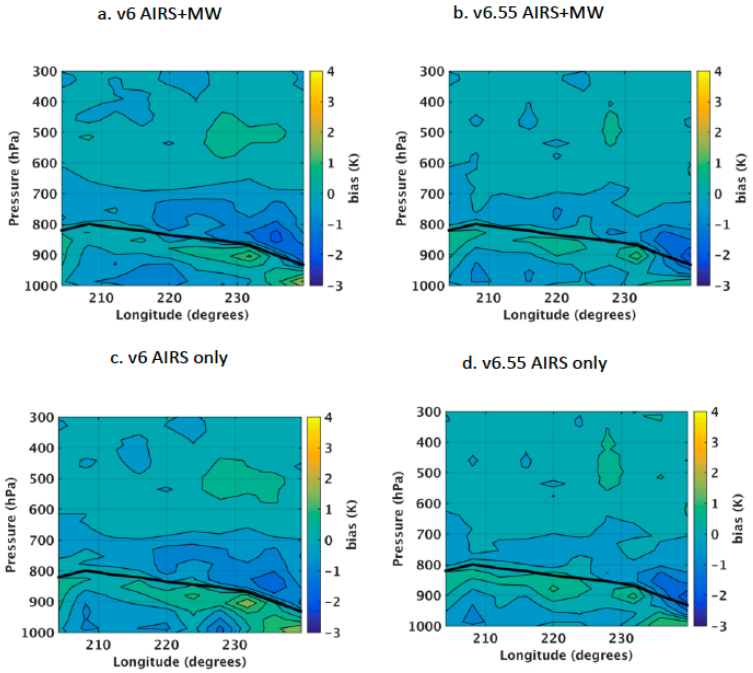


Figure 5.1.22. Comparisons with MAGIC sondes (Fig. 5.1.18 for location) for T bias for (a) v6 AIRS+MW; (b) v6.55 AIRS+MW (c) v6 AIRS only; (d) v6.55 AIRS only. Black lines show boundary layer top estimated from radiosonde temperature inversions. Note that absolute difference in K is shown.

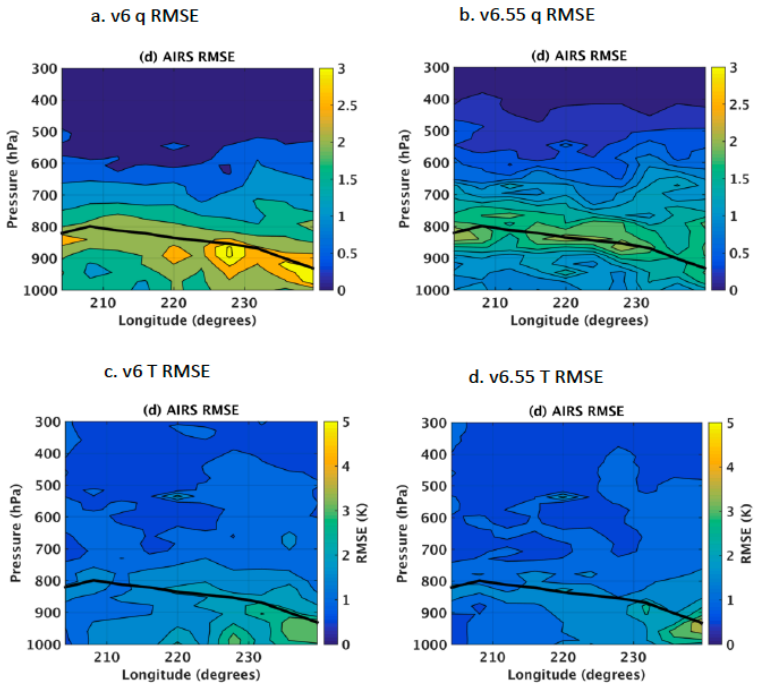


Figure 5.1.23. Comparisons with MAGIC sondes (Fig. 5.1.18 for location) for T and q RMSE for the AIRS-only product. (a) v6 q RMSE; (b) v6.55 q RMSE (c) v6 T RMSE; (d) v6.55 T RMSE. Black lines show boundary layer top estimated from radiosonde temperature inversions. Note that absolute difference in g/kg or K is shown.

5.1.4 Pixel-scale Comparisons with Field Campaign Measurements in the Central Arctic

Contributor: Qing Yue, Evan Manning, Bjorn Lambrigtsen

Data

Validation of AIRS L2 temperature and water vapor profiles in the high latitude region has been scarce especially in the central Arctic, where in-situ measurements from ground-based stations are highly limited. We have obtained the thermodynamic profiles collected during three intensive field campaigns over the central Arctic: Arctic Summer Cloud Ocean Study (ASCOS; Tjernström et al., 2014); Arctic Clouds during Summer Experiment (ACSE; Sotiropoulou et al., 2016); Norwegian Young Sea Ice Cruise (N-ICE2015 or N-ICE; Granskog et al., 2016). During these campaigns, radiosondes were launched periodically to provide vertical profiles of atmospheric temperature and water vapor profiles along the ship tracks in the Arctic Ocean. Table 5.1.2 provides the time duration, geographical regions, type of radiosondes launched, and number of profiles that are collocated with AIRS in this study. Altogether there are 393 profiles included in this study. Figure 5.1.24 shows the location of these sondes specified by different classification indicators: surface class, month in which observations were made, and the effective cloud fraction (ECF) reported by AIRS. Both the surface class and ECF are obtained from the collocated AIRS pixels. Here V6.6 IR+MW ECF and SurfClass products are used although statistics presented here do not vary with the versions of AIRS products where indicators are obtained.

Table 5.1.2: The three Central Arctic field campaigns used in this study.

Campaign Name	Geographical Region	Time Period	Sonde Type	Launch Freq.	# of Sondes Collocated with AIRS
ASCOS	77.9°-87.5°N 11.1°W-9.6°E	08/03/2008-09/07/2008	Vaisala RS92	4/Day	75
ACSE	71.4°N-85.2°N 25.7°E-178.1°W	07/06/2014-10/01/2014	Vaisala RS92	4/Day	191
N-ICE2015	79.2°N-83.3°N 3.4°E-29.8°E	01/12/2015-06/22/2015	Vaisala RS92	2/Day	127

Methodology

The AIRS L2 temperature and water vapor profiles are collocated with the radiosonde measurements from the field campaigns using the nearest neighbor method. Only one AIRS pixel that is closest in both time and space to each radiosonde is taken within collocation thresholds of 50 km and 1 hour. The ECMWF model analysis used in Sec 5.1.1 is also collocated to the in-situ data and AIRS data (Fishbein 2018). The retrieval yield profile, bias, RMS, and sampling bias in different versions of AIRS products and ECMWF are calculated against the radiosonde measurements.

Three classification indicators are used to dissect AIRS data using different methods. For the three campaigns, only three SurfClass values are found: 2 – nonfrozen ocean; 3 – New Sea Ice (high microwave emissivity); and 4 – Old Sea Ice (low microwave emissivity). Table 4.1.1 shows complete information on SurfClass from the AIRS IR+MW product. In the high latitude region, the surface types greatly impact the quality of the retrieval and the difference between the IR-only and IR+MW products (Yue and Lambrigtsen 2017). Binning profiles based on the ECF from AIRS gives a first-order dependence of retrievals on the cloud state within the pixel (Yue et al. 2013, Wong et al. 2015). The thermodynamic structure in the central Arctic shows distinct seasonal variability and hence the classification based on the month when observation was collected.

Conclusions

1) By SurfClass

Figure 5.1.25 shows the results for three different surface classes of the observations. Due to the quality of radiosonde humidity profile in the upper level, only profiles between 400hPa and surface level are shown for humidity. As approaching the surface, the QC=0&1 yield in the new versions are smaller than V6 but with a larger QC=0 yield. In this region, the spurious temperature bias above 200hPa in V6 has been removed in the new versions. Over all three surface types, the V6.6 IR-only produces the largest cold bias (<2K) below 800hPa. Over the nonfrozen ocean, the V6.6 IR+MW produces a smaller bias in T (<0.5K) and Q (<10%) throughout the profile than the previous versions and ECMWF however. Over the old sea ice surfaces, the two new versions show an improved humidity retrieval but with a larger cold bias below 700 hPa. Improvements in the SCCNN in the new versions are clear, where a much smaller bias against sondes are seen except the V6.6 IR-only temperature below 800hPa. The magnitude of RMS increases as surface types varying from ocean, new sea ice, and old sea ice. The new versions show smaller RMS in both T and Q than previous versions. However, the ErrEst in the retrieval products always underestimates the errors of the profiles.

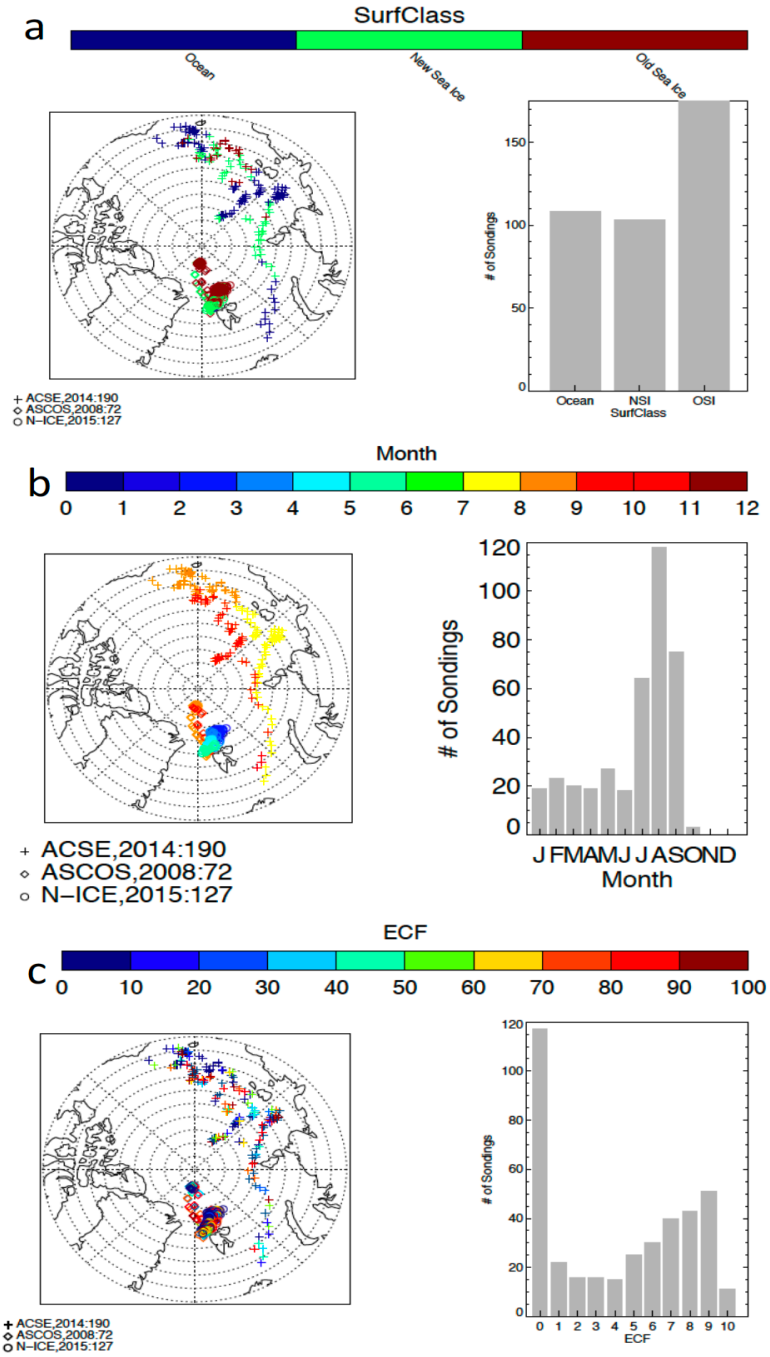


Figure 5.1.24: the location of the collocated sondes during the three field campaigns (indicated by different symbols) specified by classification indicators: the surface class (a), month when the observation was conducted (b), and the effective cloud fraction (ECF) from AIRS (c). The histogram shows the number of observations by different values of classification indicators.

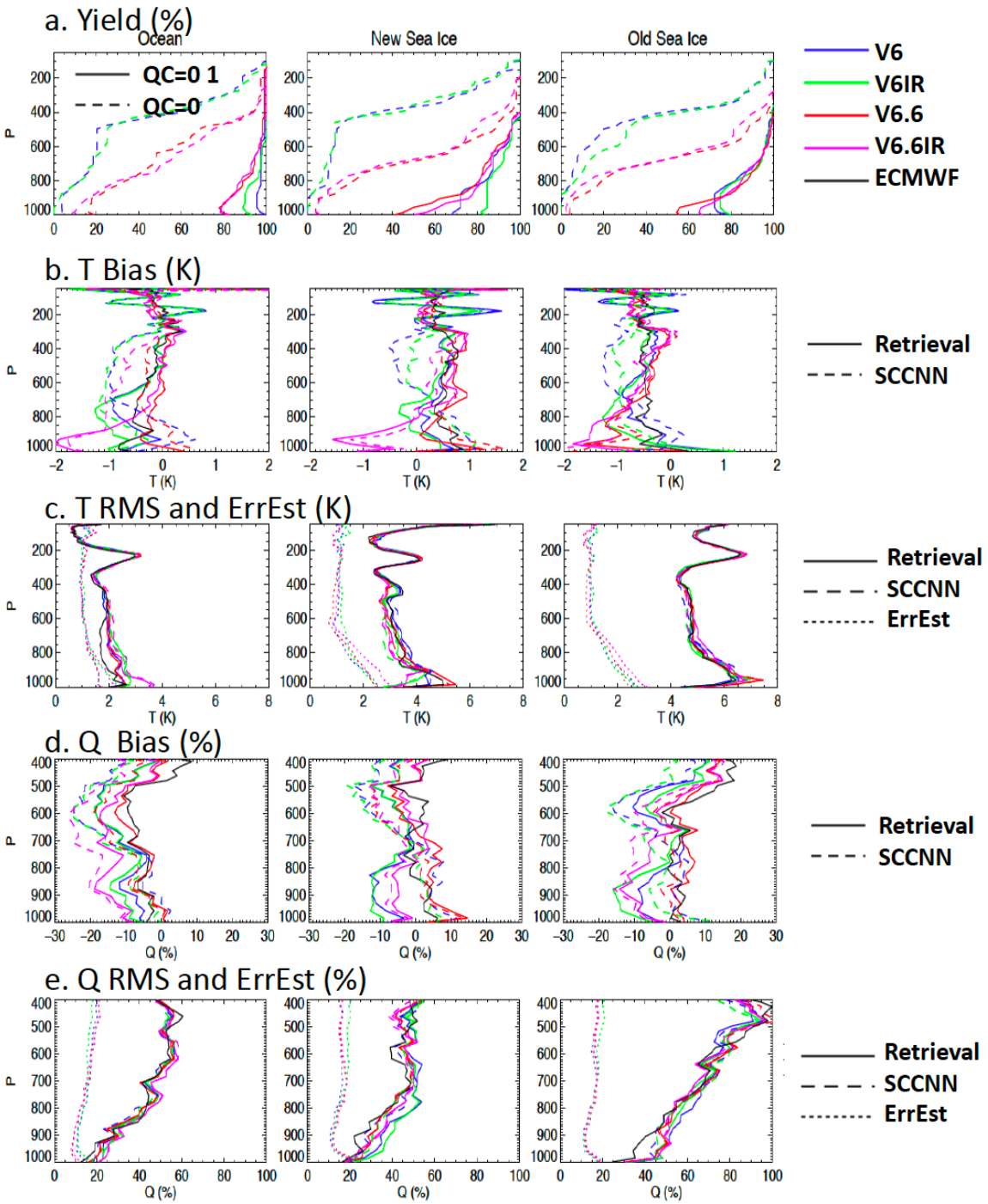


Figure 5.1.25. Yield (a), bias (b and d), and RMS (c and e) for temperature (T) and humidity (Q) profiles of various versions of AIRS products and ECMWF model analyses as indicated by different colors. ErrEst stands for the reported error estimate in the AIRS L2 products. The solid and dashed lines in Panel a) correspond with QC = 0 or 1 and QC=0. They correspond with biases of AIRS retrieval and AIRS SCCNN in Panels b and d. The sold, dashed, and dotted lines in Panels c and e correspond with RMS of AIRS retrieval and AIRS SCCNN, and ErrEst reported in the AIRS L2 products.

2) By ECF

Retrieval yield as a function of ECF is shown in Figure 5.1.26 by combining all three surface types together. All four retrievals show a dip in yield when cloud fraction is 0.3~0.4.

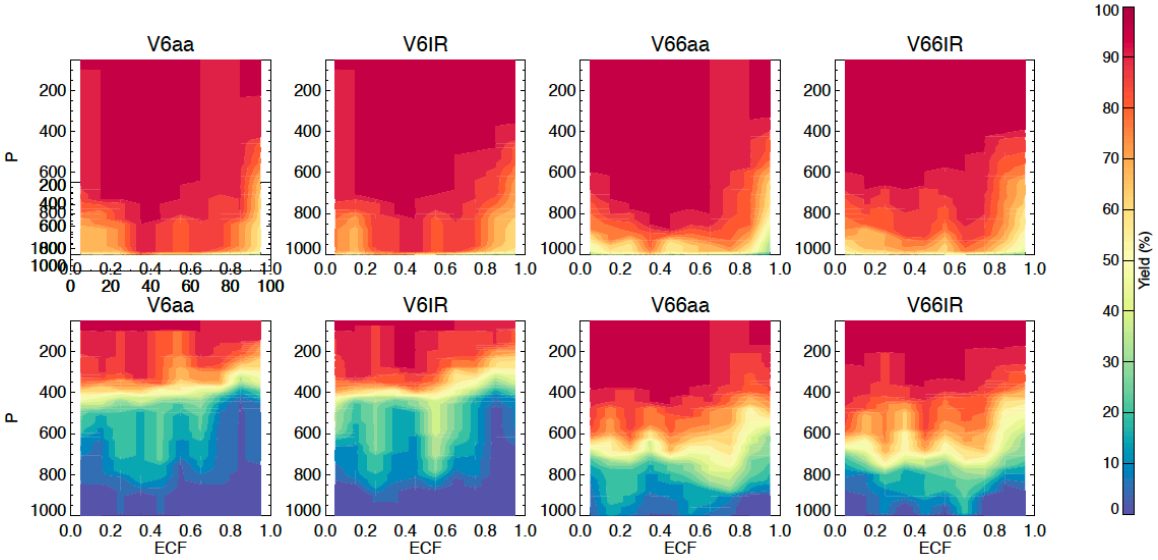


Figure 5.1.26 Retrieval yield profiles as a function of AIRS ECF for various versions of AIRS products. Top row shows results for QC=0 & 1 and bottom shows QC=0.

Figure 5.1.27 shows the bias and RMS profiles binned by ECF bins. Five bins are used: 0~0.2, 0.2~0.4, 0.4~0.6, 0.6~0.8, 0.8~0.9. The new versions generally have smaller near surface retrieval yields than V6. In clear conditions (ECF between 0 and 0.2), a larger cold bias (~ -2K) and larger RMS below 800 hPa is noticed for temperature profiles in the new versions than in ECMWF and V6, while the humidity bias is greatly reduced. Reduction in RMS and bias is clear in V6.6 IR+MW for ECF between 0.2 and 0.8 especially for humidity profiles; however, the V6.6 IR-only product produces cold bias and dry bias below 850 hPa with magnitudes larger than other versions. Similarly, the reported ErrEst underestimates the retrieval error in all ECF bins. At all ECF bins, it is noticed that SCCNN in the V6.6 (both IR-only and IR+MW) produces biases with smaller magnitudes than those in V6. However, the differences between the final retrievals from different versions are much smaller than those from SCCNN above 700hPa where the AIRS instrument has higher information content.

3) By Month

In order to increase the sample size, results in this section are presented by binning the data into three bins: January-February-March (JFM), April-May-June (AMJ), and July-August-September (JAS). Figure 5.1.28 presents the yield profiles as a function of the center point of each bin. Below 600 hPa, the lower yield in winter seasons and the higher yield in summer seasons is clear in Figure 5.1.28.

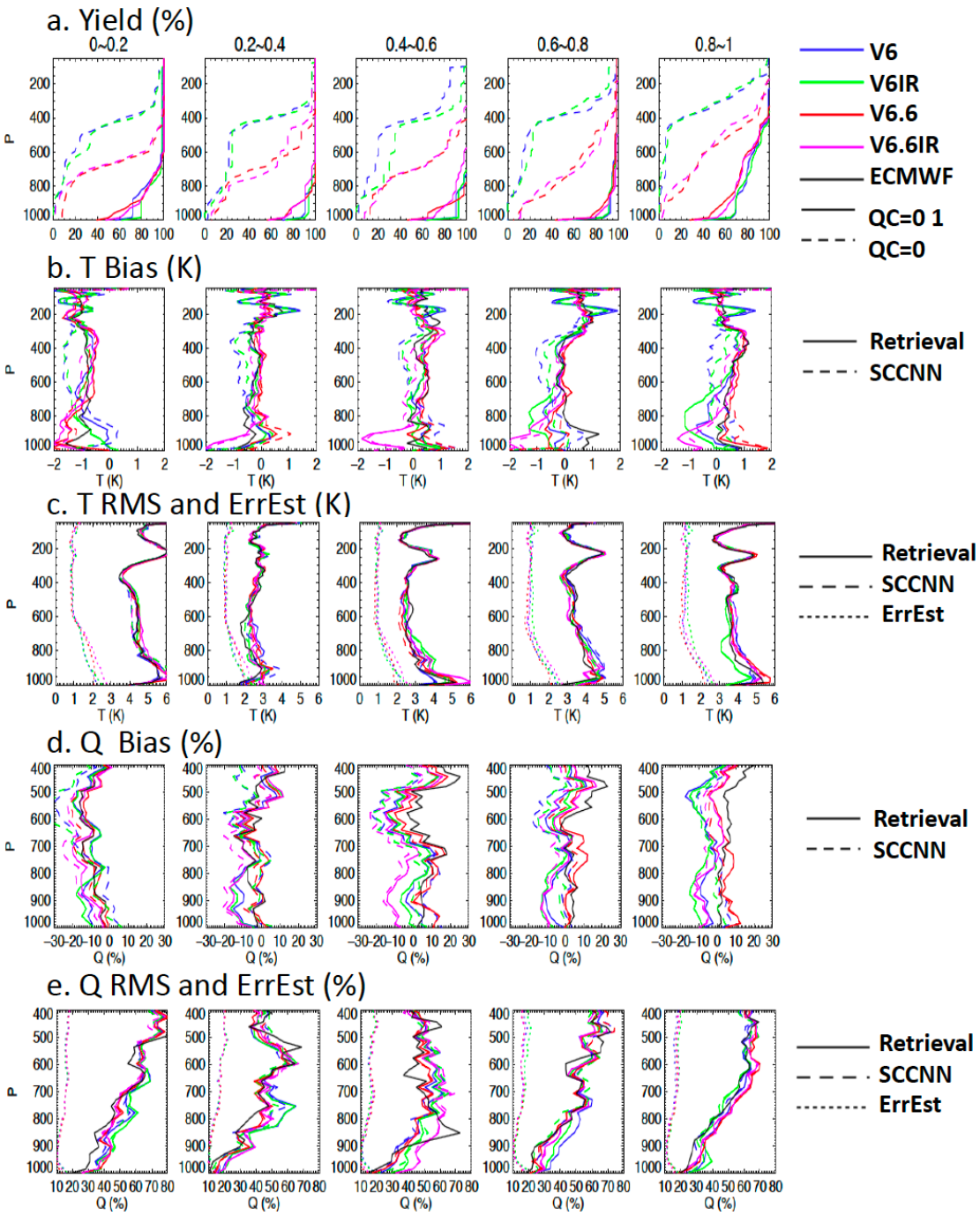


Figure 5.1.27. Similar to Figure 5.1.25, except by five ECF bins: 0~0.2, 0.2~0.4, 0.4~0.6, 0.6~0.8, 0.8~0.9.

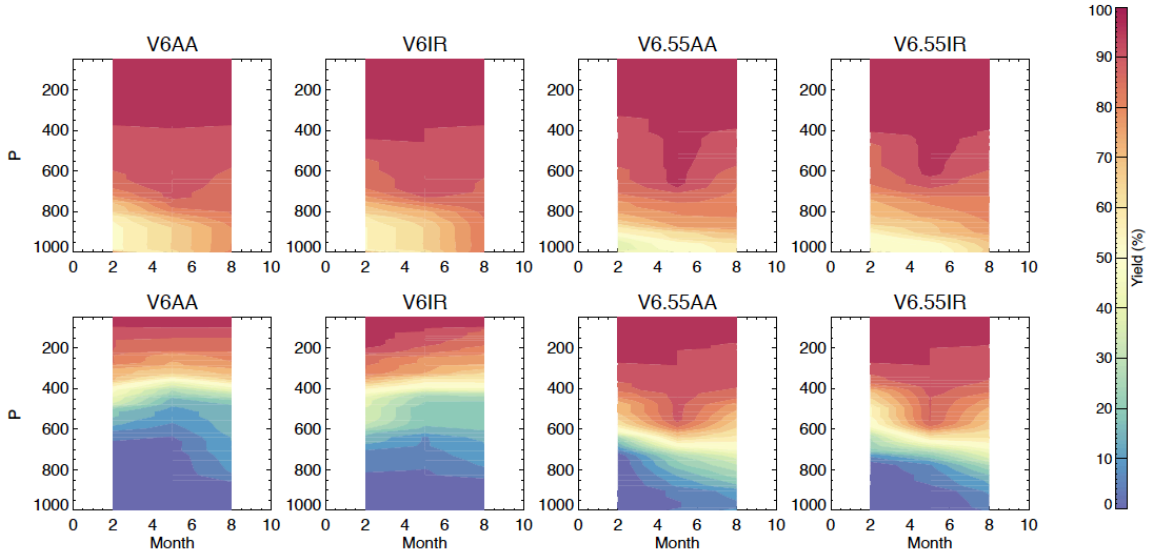


Figure 5.1.28. Retrieval yield profiles as a function of month for various versions of AIRS products. Top row shows results for QC=0 & 1 and bottom shows QC=0. Data are grouped into three bins: January-February-March (JFM), April-May-June (AMJ), and July-August-September (JAS). Results are presented as a function of the center point of each bin.

Figure 5.1.29 shows the mean bias and RMS of different AIRS versions and collocated ECMWF analyses as a function of month when the observation was taken. The largest bias and RMS in both temperature and humidity profiles from AIRS are seen in JFM, the coldest months in the Arctic. The retrievals from all four algorithms and ECMWF model analyses all show significant cold and dry bias compared with the sondes with smaller RMS from AIRS data than ECMWF. New versions from AIRS do not show significant improvements compared with V6 in JFM. During AMJ, a warm and wet bias is seem in all AIRS and the ECMWF data. The new versions produce much smaller temperature biases below 800 hPa than V6 and ECMWF due to the improvements in the SCCNN, however, with a much larger wet bias. In JAS all datasets produce the smallest bias and RMS among all three bins. This is also the time period when the V6.6 IR+MW retrievals significantly reduce the bias and RMS compared to V6; however, V6.6 IR+MW has the largest cold bias (-2K) near surface and dry bias (-15%) at 900 hPa among all AIRS versions.

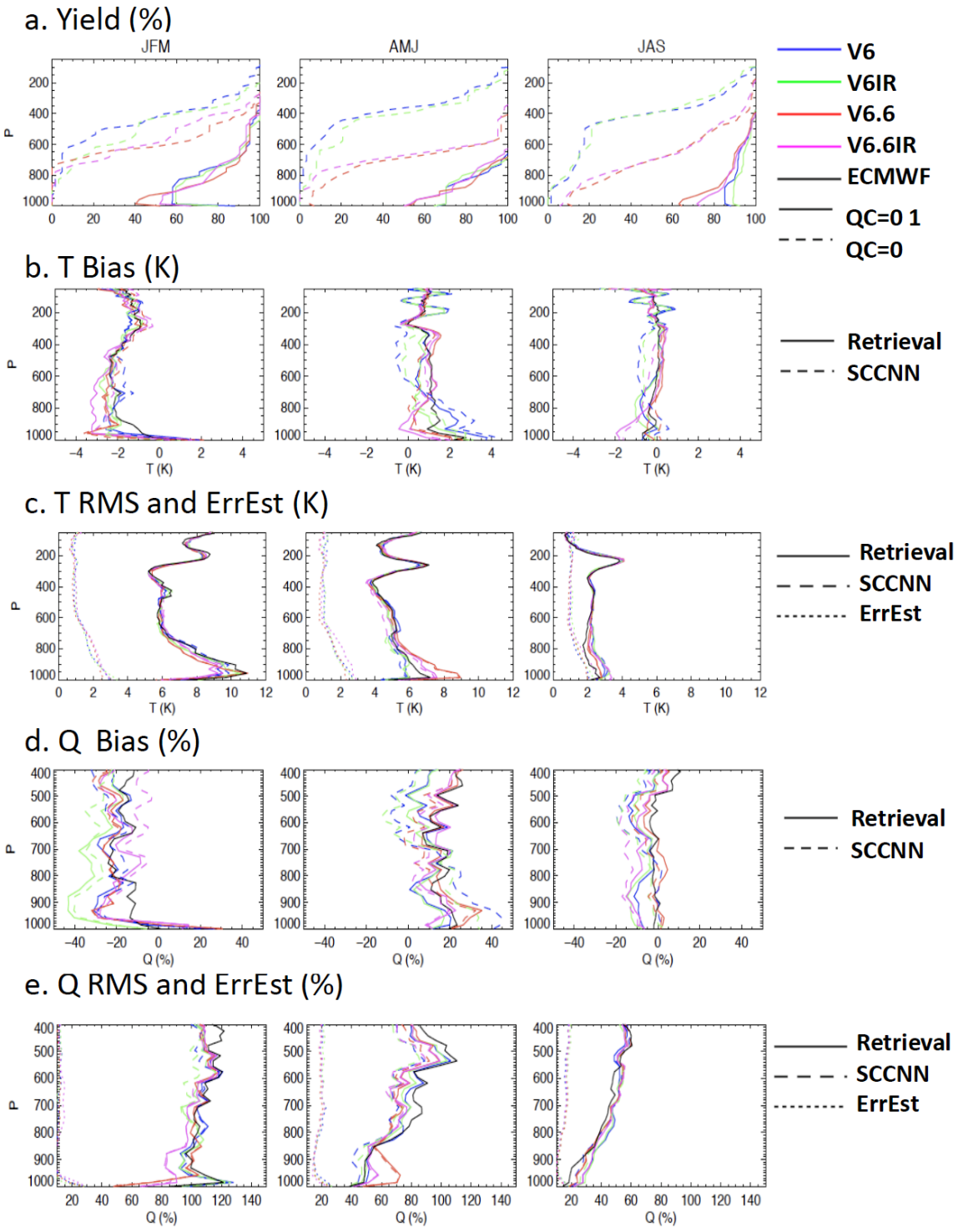


Figure 5.1.29. Similar to Figure 5.1.25, except by month of observation taken. Data are divided into three bins: January-February-March (JFM), April-May-June (AMJ), and July-August-September (JAS).

5.1.5 Temperature and Water Vapor Profile Bias and Bias Trends Compared with PREPQC Radiosondes

Contributor: Fredrick Irion

Data

Versions 6.0.11 (hereafter V6) and 6.5.5 AIRS retrieved temperatures and water vapor were examined for temporal trends in bias against radiosonde observations in the National Centers for Environmental Prediction (NCEP) quality-controlled final observation (PREPQC) data files. The latter are from operational radiosondes mostly launched near 00 and 12 hrs GMT, with some launched near 6 and 18 hrs GMT. In addition to temperature, pressure and (often) specific humidity observations, the data records contain codes relating to quality control, correction methodology, radiosonde type etc., which were used in filtering the data before comparison to AIRS retrievals. There are significant differences in methodology for bias trend testing between temperatures and water vapor as described in the subsections below.

Methodology and Conclusions

1) Temperature bias trends

To test and compare Version 6 and 6.5.5 AIRS temperature retrievals, we matched radiosonde data to AIRS observations and calculated zonally-averaged differences as a function of pressure. Radiosonde data were filtered by the following criteria:

1. Sonde launch was within 1 hour and 100 km of an AIRS observation over land. (Note that this can lead to multiple AIRS observations matched to the same sonde profile.)
2. For each pressure/temperature observation within a sonde profile:
   (a) If the "Temperature Program Code" (TPC) was 8, indicating that virtual temperature was given (T_V), the sensible temperature (T_s) was calculated from the formula:

$$T_s = \frac{T_V}{1+0.608Q} \tag{5.1.5}$$

   where T_V, T_S are in Kelvin, and Q is the specific humidity using the same mass units (e.g., kg H_2O/kg dry air, not mg H_2O/kg dry air).
   (b) A single-level observation (but not necessarily the whole profile) would be rejected if either the "Temperature Quality Marker" (TQM), or the "Pressure Quality Marker" (PQM) was greater than 2.
   (c) An observation would be kept if the "Temperature Program Code" (TPC) and "Temperature Reason Code" (TRC) were matched as follows:
      (TPC = 8 and TRC = 1) (Virtual temperature reported calculated from sensible temperature and specific humidity), or
      (TPC = 1 and TRC = 100) (Temperature and quality marker unchanged from original values read into program "prepdata"), or
      (TPC = 6 and TRC = 1) (Rawinsonde has had a radiation bias correction with temperature and/or height recalculated),
   otherwise the observation would be rejected.
3. An entire sonde profile would be rejected if
   (a) the instrument type (ITP) was
      2 through 6, 8, or 9 (not radiosondes), or
      90 (radiosonde type unknown), or

100 through 109, or 190 through 225 (vacant, "not vacant", or reserved fields), or

(b) Accepted sonde temperature levels did not reach altitudes above the 90 mb level, or altitudes below the 700 mb level, or

(c) A height gap in the accepted temperature observations was greater than 5 km (calculated from pressures assuming a 7km scale height).

Both daytime and nighttime observations are included, however comparisons for temperature are mostly at 0° in longitude and points eastward over Europe due to launch of sondes near noon GMT and the (ascending) 1:30PM equator crossing time of AQUA satellite (see Figure 5.1.30).

Radiosonde data were interpolated (by log pressure) to the AIRS 100-level vertical pressure grid prior to comparison, but we did not extrapolate radiosonde data beyond their original range. Only those AIRS temperature levels that had quality flags of 0 (best) or 1 (good) were used in analyses, and an AIRS retrieval had to be successful for both V6 and V6.5.5 to be included.

To calculate and compare any temperature bias drifts in Versions 6 and 6.5.5 AIRS retrievals, we used all available matchups from the beginning of 2004 through the end of 2016. (Dates after 2003 were used to avoid any 'jumps' resulting from calibration changes after the AIRS on-orbit shutdown of October 29 to November 14, 2003.) To avoid seasonal effects skewing the result, data were binned by calendar month, with the date of each matched observation set to the year+fraction-of-year. (For example, Feb 1, 2009 at 00:00 GMT would have the date 2009.084931.) Data were then binned first by seven latitude bands, then by the AIRS pressure levels. For each level, a linear least-squares fit of the bias vs. the fractional date was calculated. The slopes from each calendar month were then averaged, and these averages are presented (in mK) in Figures 5.1.31 through 5.1.33 for each latitude bin and pressure. AIRS comparisons against the neural-net *a priori* are on the left panels, and the final AIRS retrieval against sondes are on the right. Error bars show the standard deviation of the monthly averaged slopes on each pressure level.

From Figures 5.1.31 through 5.1.33, temperature drifts are comparable between V6 and V6.5.5. The lowest temperature drifts are seen in the 30°N – 60°N zonal average (~ ±20mK/yr), where the most sondes are available for comparison. The highest drifts (by absolute measure) are in the boundary layers for the 15°S – 15°N zone (~ -60mK / yr) and 60°S – 30°S (~ 50mK / yr) zone, although we note that comparatively fewer sondes are available in these latitude ranges. Generally, temperature drifts tend to be improvements on the a priori, but an exception to this is seen in the lower stratosphere of the southern polar region (90°S – 60°S) at pressures less than ~200 mb.

2). Water bias trends

AIRS-radiosonde matchups used for water vapor analyses were more selective than for temperature as the reliability of radiosonde water vapor observations (reported as specific humidity) can vary significantly, particularly under freezing conditions (see, for example, Miloshevich et al., 2006). We therefore restricted radiosonde results to variants of the Vaisala RS-92 sensor and its follow-on, the Vaisala RS-41, which have been well-studied (e.g., Edwards et al., 2014). Only sonde data from sites for which RS-92s or -41s were repeatedly and regularly used (with at least 6000 sonde launches from 2004 through 2016) were included, and only those sondes launched at 00:00 or 12:00 GMT.

Acceptance of a radiosonde profile for water vapor used the same criteria as temperature (as described above) with the additional provision that a single-level observation (but not necessarily

the whole profile) would be rejected if either the "Specific Humidity (Quality) Marker" (QQM), or the "Pressure Quality Marker" (PQM) were greater than 1. Figure 6 shows radiosonde launch sites, colored by number of successful AIRS matchups, indicating that results are valid mostly over central and eastern Europe. Note that unlike for temperature, we have not binned the data by latitude.

For comparison with AIRS results, radiosonde profiles were converted from specific humidity to "slab columns," that is, the integrated number of water vapor molecules per square cm between two pressure levels on the AIRS forward-model grid. (This is the same data format as the AIRS support product for water vapor and other trace gases.) Results were compared using relative bias.

Figure 5.1.34 compares average drifts in the relative water vapor bias [(AIRS – sonde) / sonde (%)] for the *a priori* and final retrievals, and for V6 and V6.5.5. The size of the error bars, which are the standard deviations of the trends calculated over calendar months individually, suggests a wide variation in the relative bias by season. Both the *a priori* and final retrieval drifts for both versions tend to be within ±1%/yr at pressures greater than ~650 mb. In the middle-to-upper troposphere (~600 mb to 300 mb), negative biases of ~ -0.75%/yr are seen for both V6 and V6.5.5, however these appear to be an improvement on the *a priori*.

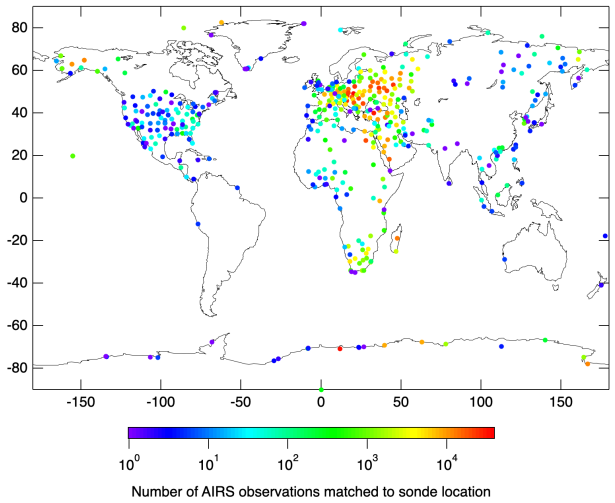


Figure 5.1.30: Number of AIRS retrievals successfully matched to radiosonde temperature observations by radiosonde launch site, 2004 – 2016 inclusive.

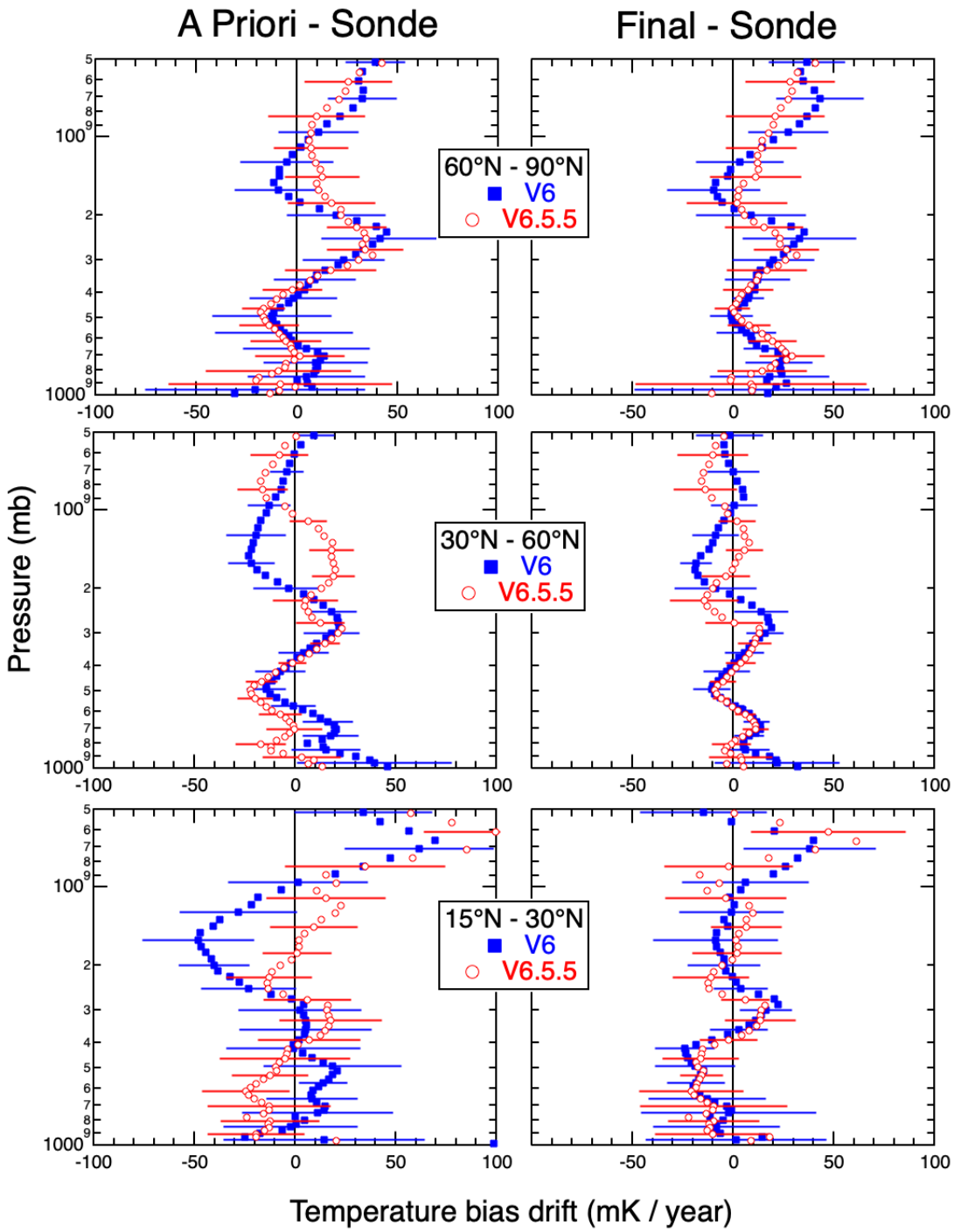


Figure 5.1.31: Averaged AIRS V6 and V6.5.5 temperature bias drift against coincident radiosondes as a function of pressure for AIRS *a priori* profiles (left panels) and final retrievals (right panels). Rows show binning of results by latitude bands, 60°N – 90°N, 30°N – 60°N, and 15°N – 30°N.

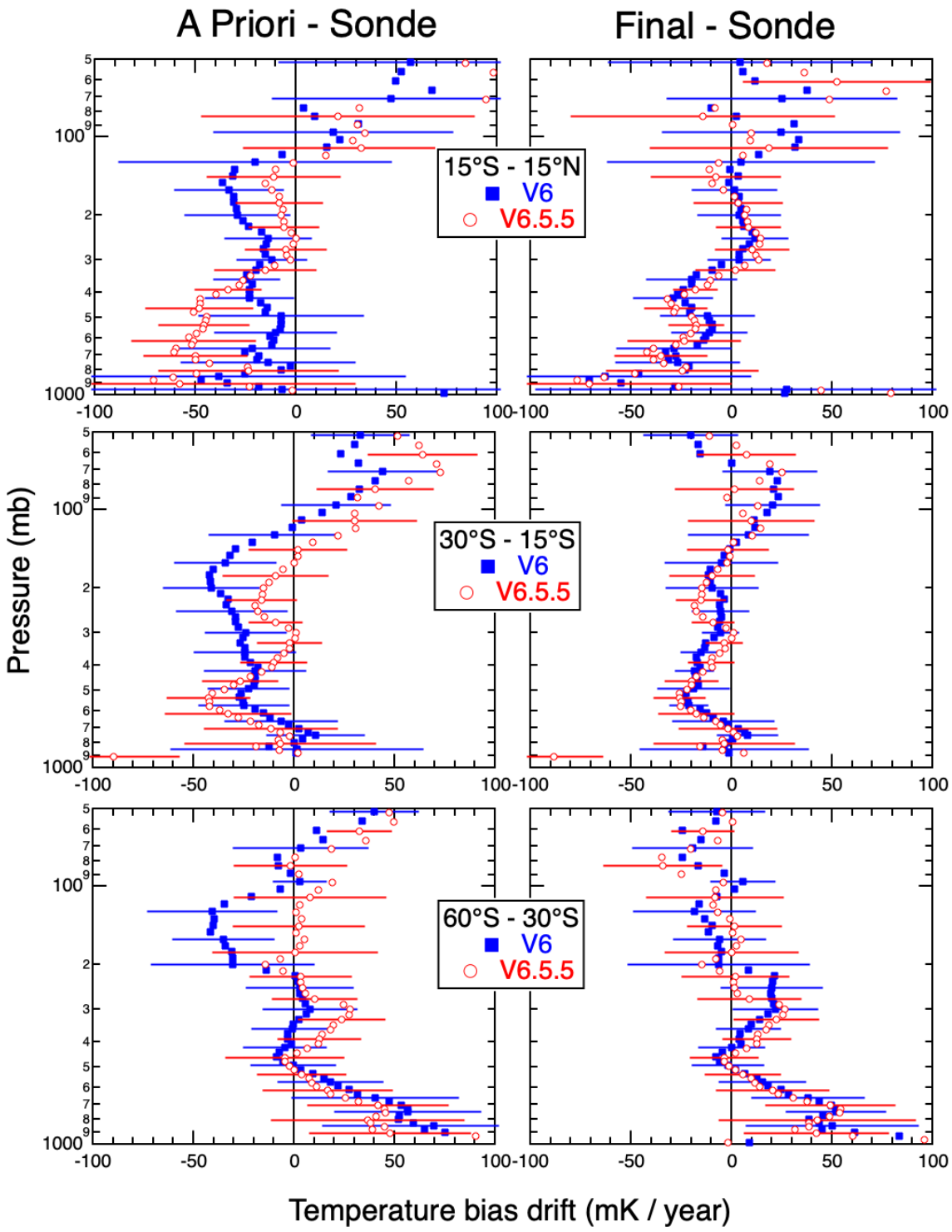


Figure 5.1.32: Averaged AIRS V6 and V6.5.5 temperature bias drift against coincident radiosondes as a function of pressure for AIRS *a priori* profiles (left panels) and final retrievals (right panels). Rows show binning of results by latitude bands, 15°S – 15°N, 30°S – 15°S, and 60°S – 30°S.

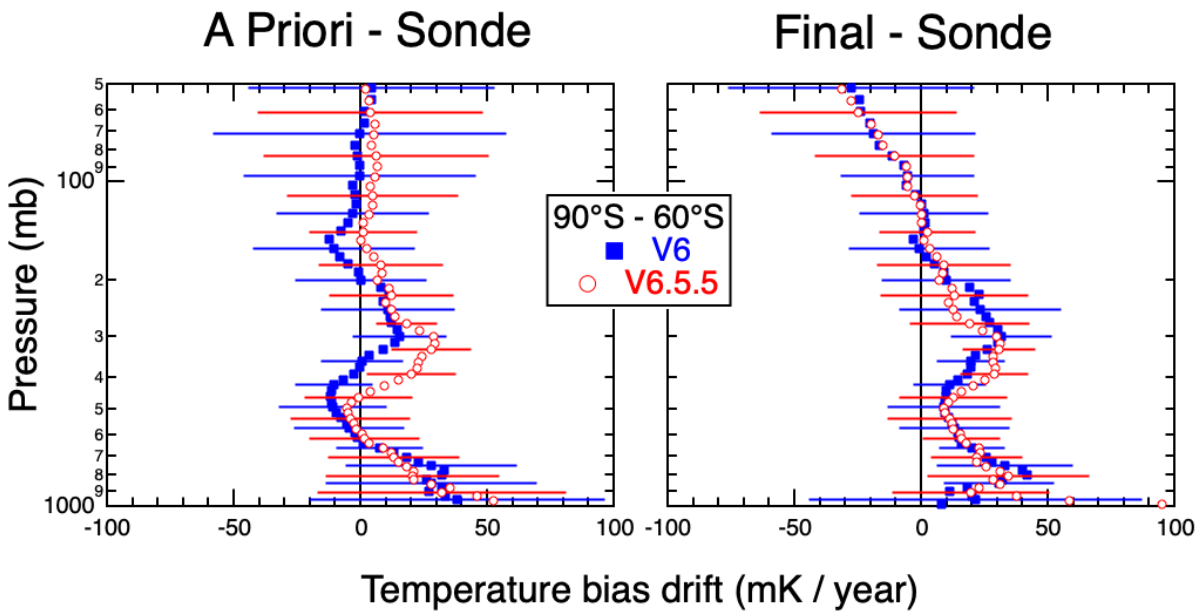


Figure 5.1.33: Averaged AIRS V6 and V6.5.5 temperature bias drift against coincident radiosondes as a function of pressure for AIRS a priori profiles (left panels) and final retrievals (right panels) for latitude band, 90°S – 60°S.

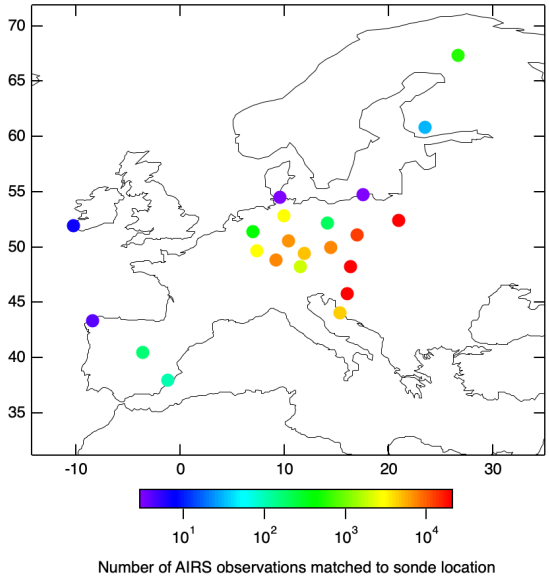


Figure 5.1.34: Map of number of AIRS observations matched to Vaisala RS92 or RS41 radiosondes as used in the determining AIRS water vapor bias trend, 2004 through 2016. (See text for selection criteria.)

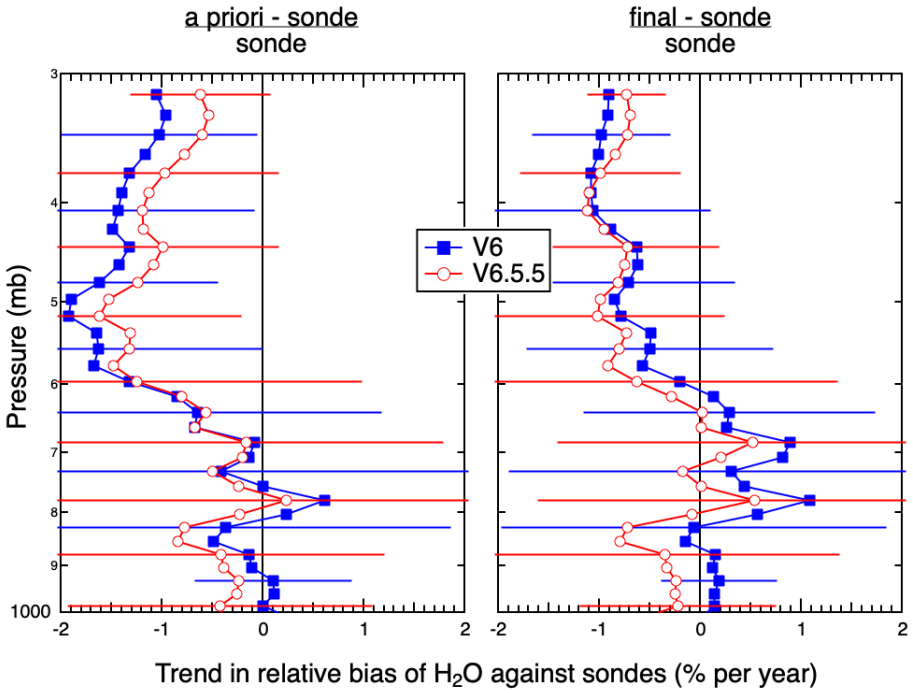


Figure 5.1.35, left panel: Trend in relative bias of AIRS *a priori* water vapor vs coincident radiosondes. Right panel: Trend in relative bias of AIRS final retrieval. Points are averages of individual monthly trends, with standard deviations of such monthly trends shown as error bars (see text). For clarity, only every fourth error bar is shown.

5.2 Analyses on AIRS Cloud Product

5.2.1 Granule-scale Level 2 Comparisons of Cloud Products Between V6 and V7

Contributor: Brian H. Kahn

Data

This is a straightforward head-to-head comparison of v6 versus v6.55 granule-scale cloud products. All data are obtained from the Standard and Support L2 hdf files, both for the combined IR/MW (AIRX2RET and AIRX2SUP) and IR only (AIRS2RET and AIRX2SUP). While a couple of granules are shown in this brief evaluation, a series of spot checks were done over dozens of granules during January 1, 2011 and July 1, 2011. The variables evaluated were the following: (1) the two-layer CldFrcStd variable at 90x135 spatial resolution from AIRX2RET and AIRS2RET, (2) the two-layer TCldTop variable at 90x135 spatial resolution from AIRX2RET and AIRS2RET, and cloud_phase_3x3 from AIRX2SUP an AIRS2SUP. No independent truth data was used in this study. The ice cloud microphysical variables are not evaluated.

Methodology

No quality control was applied as it is of strong interest to understand all of the potential artifacts in the cloud products within selected AIRS granules. These cloud products are used in a wide variety of geophysical states, not just those associated with temperature and specific

humidity soundings associated with full profiles of PGood and PBest QC. A second version of all plots are shown removing values of total effective cloud fraction < 0.05 (sum of both layers in ClFrcStd) in order to highlight some of the issues in the most tenuous clouds reported.

Conclusions

1. Adverse change in v6.55: While v6 has some occasional striping and speckling behavior at the granule scale, there appears to be additional areas and increased severity of this behavior in v6.55. These speckling and striping issues show up as thin cloud artifacts within what is likely to be clear sky. These artifacts appear to be concentrated in the upper layer rather than the lower layer.
2. Adverse change in v6.55: Following on point (1) above, there is some additional striping/speckling behavior in other likely cloudy scenes in v6.55 compared to v6. As with point (1) above, these artifacts appear to be concentrated in the upper layer rather than the lower layer.
3. Adverse change in v6.55: The previous v6 was not without its own problems. Fairly frequently, very cold and thin clouds would be retrieved in scenes that are likely to be clear. These often show up at the edges of more robust clouds. These artifacts appear to be more frequent in v6.55 compared to v6 although we have not done a careful analysis on these types of clouds. In the past we would throw out clouds with ECF < 0.01 and these would be eliminated nearly all of the time.
4. Neutral result for v6.55: There is little difference in cloud_phase_3x3 as the spatial locations of ECF values greater than or less than 0.01 change very little between v6 and v6.55. Thus, the sample set of ice clouds used in subsequent ice cloud property retrievals appears to be very similar between v6 and v6.55.
5. Other neutral results for v6.55: As a whole, other than the highlighted issues pointed out in this report, there are rather small changes in the magnitude and distributions of these cloud properties. We speculate that most of these problem cases could be eliminated with a careful filter of the thinnest clouds using ECF thresholds determined by the demands of a given analysis.
6. Mean values and standard deviations are reported on the figure panels to follow and demonstrate the impacts of artifacts or shifts in the overall average values.

Discussion

Figure 5.2.1 shows granule 014 on July 1, 2011 and is situated along the west African coast. This particular granule highlights some of the issues summarized above. The upper row of Figure 5.2.1 shows cloud_phase_3x3 for v6 (left), v6.55 (middle), and brightness temperature at 1231 cm^{-1}. While some of the "clear sky" pixels change between v6 and v6.55, these slight changes suggest overall that the ECF hasn't changed substantially between v6 and v6.55. Thus we can conclude that the baseline cloud_phase_3x3 product has changed very little at the granule scale between v6 and v6.55.

The middle row of Figure 5.2.1 shows the upper layer TCldTop for v6 (left) and v6.55 (middle) with the difference of v6-v6.55 (right). Note the cold areas of convection near the African coastline as well as over land and that there are very few differences between v6 and v6.55 where the changes are smallest appears to be associated with larger values of upper layer ECF. The problematic areas are located in the southern portion of this granule. There are areas of cold cloud at the bottom edge of the granule that shows up in both v6 and v6.55, but some of

the details on which pixel is cloudy does change slightly between versions. The other area of concern shows up in the difference plot with differences in TCldTop of 10-20K or more in a few pixels. These TCldTop values are warm and low and appear to be associated with the warmest brightness temperatures found in the scene and are likely mostly clear (although very thin cirrus or broken near surface shallow cumulus fields cannot be ruled out). In other granules, these types of regions in the brightness temperature imagery that look clear upon first inspection can also exhibit more severe speckling and striping (some other examples to be shown in Figure 5.2.5).

The bottom row of Figure 5.2.1 shows the lower layer of TCldTop for v6 (left) and v6.55 (middle) with the difference of v6-v6.55 (right). The largest differences in the lower layer between v6 and v6.55 appear to be in the more robust and thicker clouds rather than the thin cloud artifacts that show up strongly in the upper layer. These differences are fairly random in appearance as evidenced by the mean difference of -0.1 K. In the more problematic areas in the upper layer, for those clouds with a lower layer, the differences are very slight. The cold clouds at the bottom of the granule appear to have a lower layer in a majority of pixels. However, the other problematic region with warmer upper layers have very few pixels with a second layer.

We now filter Figure 5.2.1 to remove all clouds with a total ECF < 0.05 to isolate which clouds are the most problematic, and those filtered clouds are shown in Figure 5.2.2. The two most important take away messages from Figure 5.2.2 are the following. (1) The problematic upper layer cold clouds at the bottom of the granule are now gone but the lower layer remains. This suggests that these cold clouds in the upper layer have very small values of ECF, while the lower layer has larger values of ECF. Upon examination of the brightness temperature imagery in the upper right of Figure 5.2.2, one can see an extensive low stratocumulus cloud. It is possible that these particular cold and thin upper layer artifacts are associated with problems in cloud retrievals over low stratocumulus. (2) The warmer upper lever cloud is virtually eliminated with the ECF < 0.05 filter. Note that the standard deviations in the differences for both cloud layers are significantly reduced in Figure 5.2.2.

Figures 5.2.3 and 5.2.4 are the equivalent plots for the total (upper and lower layers summed), the upper layer, and lower layer ECF values for v6 (left column), v6.55 (middle column), and v6-v6.55 differences (right column). Interestingly, the cold cloud artifacts in the southern edge of the granule (Figure 5.2.3) show very little difference between v6 and v6.55. However, the ECF < 0.05 filtering is able to remove these clouds in the upper layer, an as suspected, they remain in the lower layer. The second problematic region with warmer TCldTop shows differences in ECF for both layers although those differences appear to be larger in the upper layer. The filtering approach is able to remove some of these values but many problematic pixels remain.

We have also investigated AIRS IR only retrievals for the same granule shown in Figures 5.2.1-5.2.4 The results are very similar to the combined IR/MW approach (not shown).

Lastly, in Figure 5.2.5 we show three more example granules on July 1, 2011 for (i) granule 004 (upper row), (ii) granule 005 (middle row), and (iii) granule 008 (lower row) that also exhibit the same types or related issues in the cloud retrievals. We show the 1231 cm^{-1} brightness temperature and the v6 and v6.55 upper layer TcldTop for the combined AIRS IR/MW retrieval. Starting with granule 004, note the striping and speckling apparent in v6 that is significantly worse in v6.55. These problematic retrievals are located in the cold sector of a midlatitude cyclone where clouds are shallow and convective. However, there appears to be much more severe striping in v6.55 compared to v6 further south when cloud tops are colder and

more numerous. In granule 005, the northern part of the granule is very consistent between v6 and v6.55 which is located over a tropical part of the ocean. However, to the south of the granule, striping appears in both versions but is more severe in v6.55. Furthermore, there appears to be some very cold cloud retrievals in the areas of thin cirrus passing through the middle of the granule. It is somewhat questionable whether these very cold retrievals within slightly warmer and lower high cloud top temperatures are realistic. This problem is exacerbated in granule 008 in the lower row of Figure 5.2.5. Near the edges of more robust clouds, there is a tendency to fill in with very thin and cold cloud retrievals in likely clear skies.

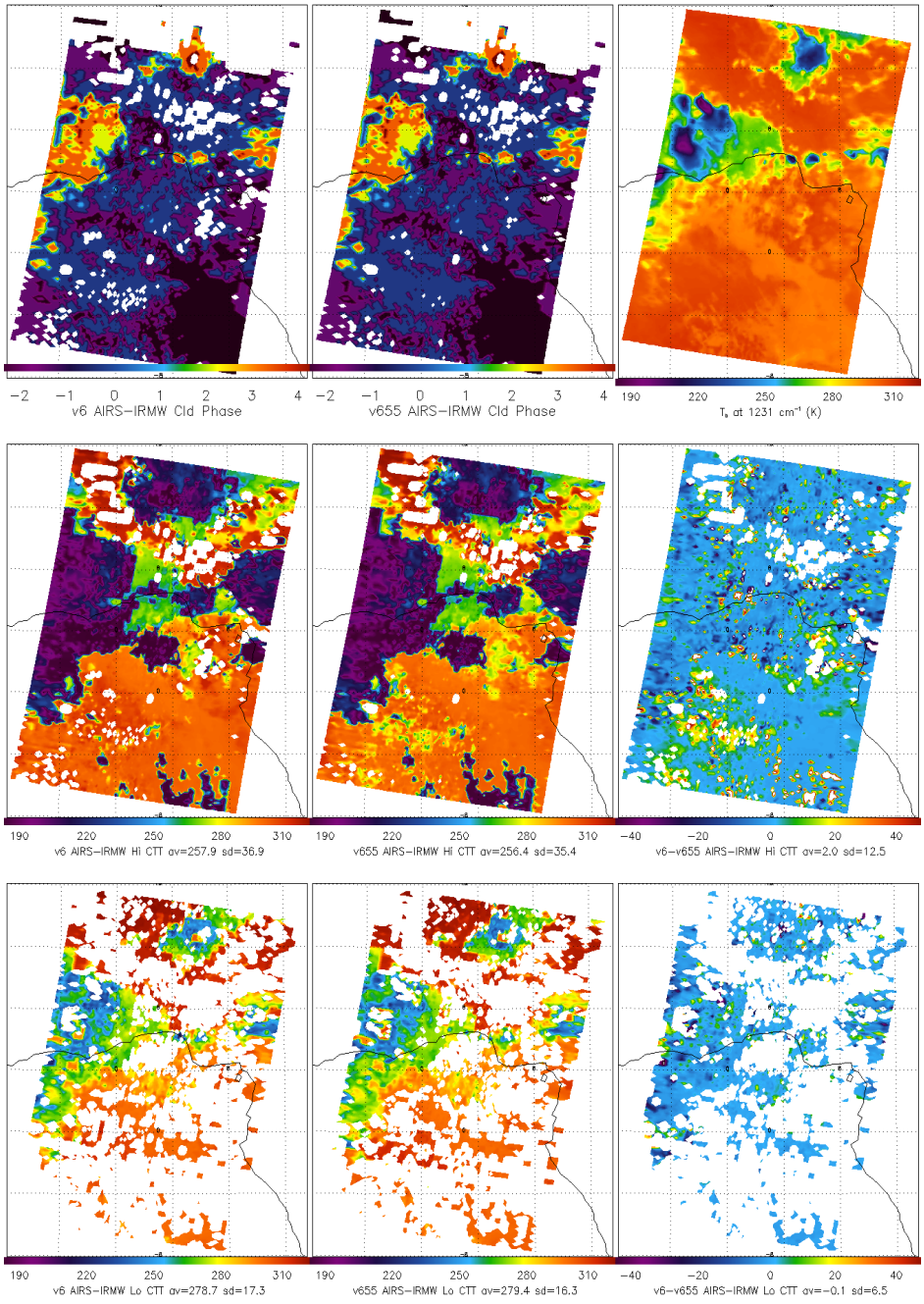


Figure 5.2.1 Upper row: cloud_phase_3x3 for v6 (left), v6.55 (middle), and brightness temperature at 1231 cm^{-1}. Middle row: upper layer TCldTop for v6 (left) and v6.55 (middle) with the difference of v6-v6.55

(right). Bottom row: lower layer of TCldTop for v6 (left) and v6.55 (middle) with the difference of v6-v6.55 (right). The averages and standard deviations of CTT (K) for the full granule are reported below each sub-panel in the left and center columns, while the differences between v6-v655 average and standard deviations of CTT (K) are reported below each sub-panel in the right column.

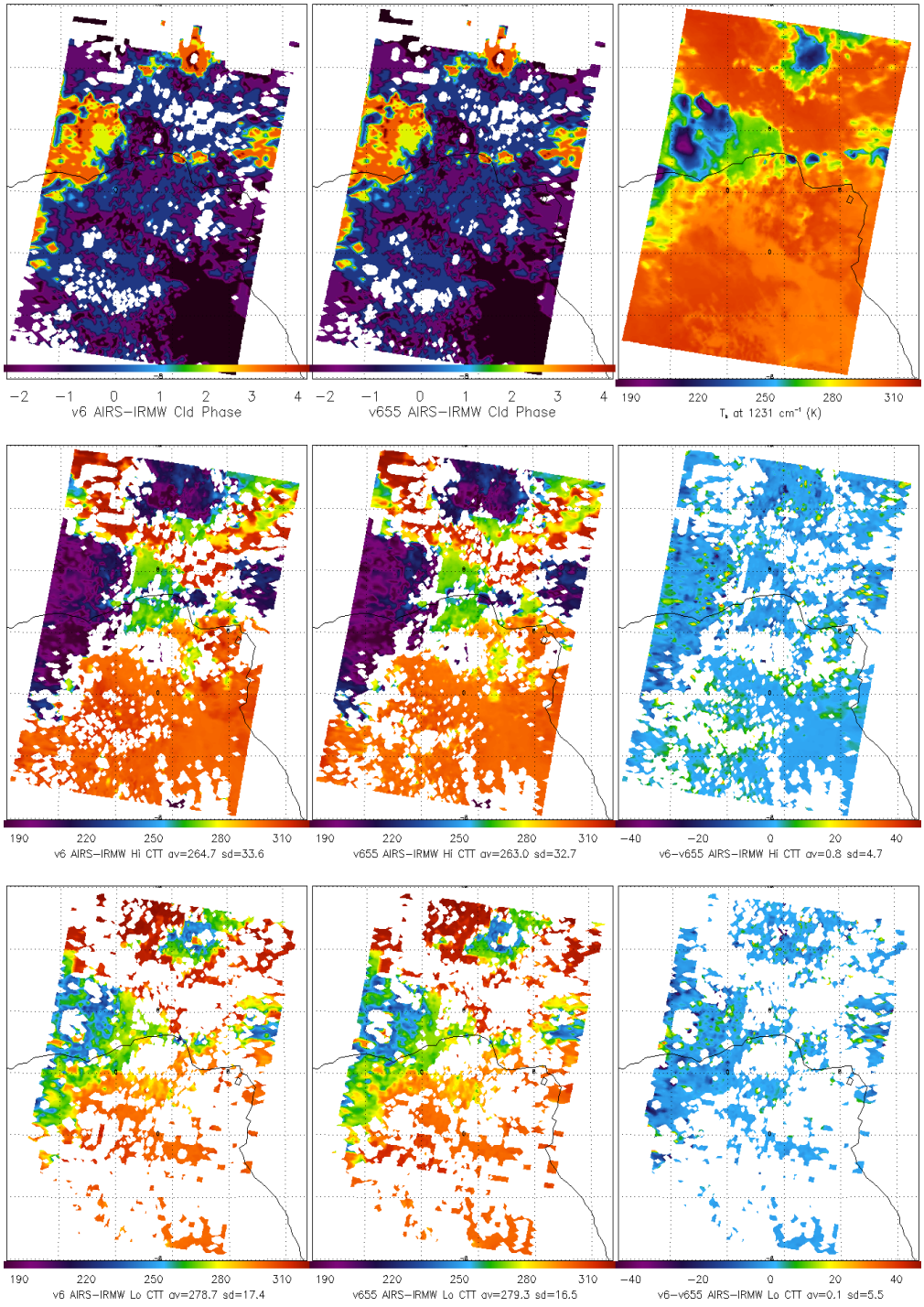


Figure 5.2.2. Same as Figure 5.2.1 except for all clouds with ECF < 0.05 filtered out. The upper row is a repeat of the upper row in Figure 5.2.1.

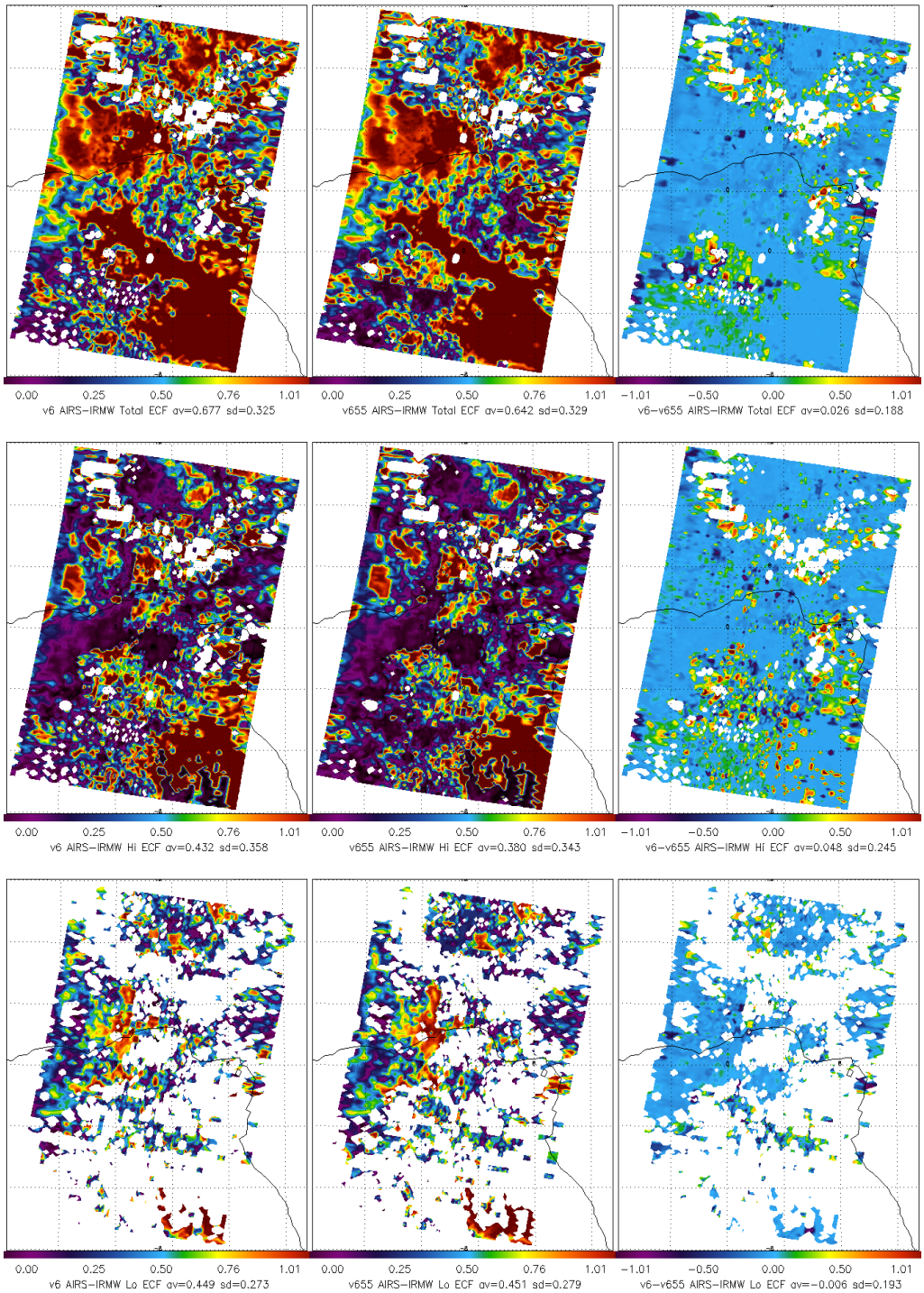


Figure 5.2.3. Upper row: total effective cloud fraction (ECF) which is a sum of the two layers of ECF (if a second layer exists within a given AIRS pixel), shown are v6 (left), v6.55 (middle), and v6-v655 (right). Middle row: upper layer ECF for v6 (left), v6.55 (middle), and v6-v6.55 (right). Bottom row: lower layer of ECF for v6 (left), v6.55 (middle), and v6-v6.55 (right). The averages and standard deviations of ECF for the full granule are reported below each sub-panel.

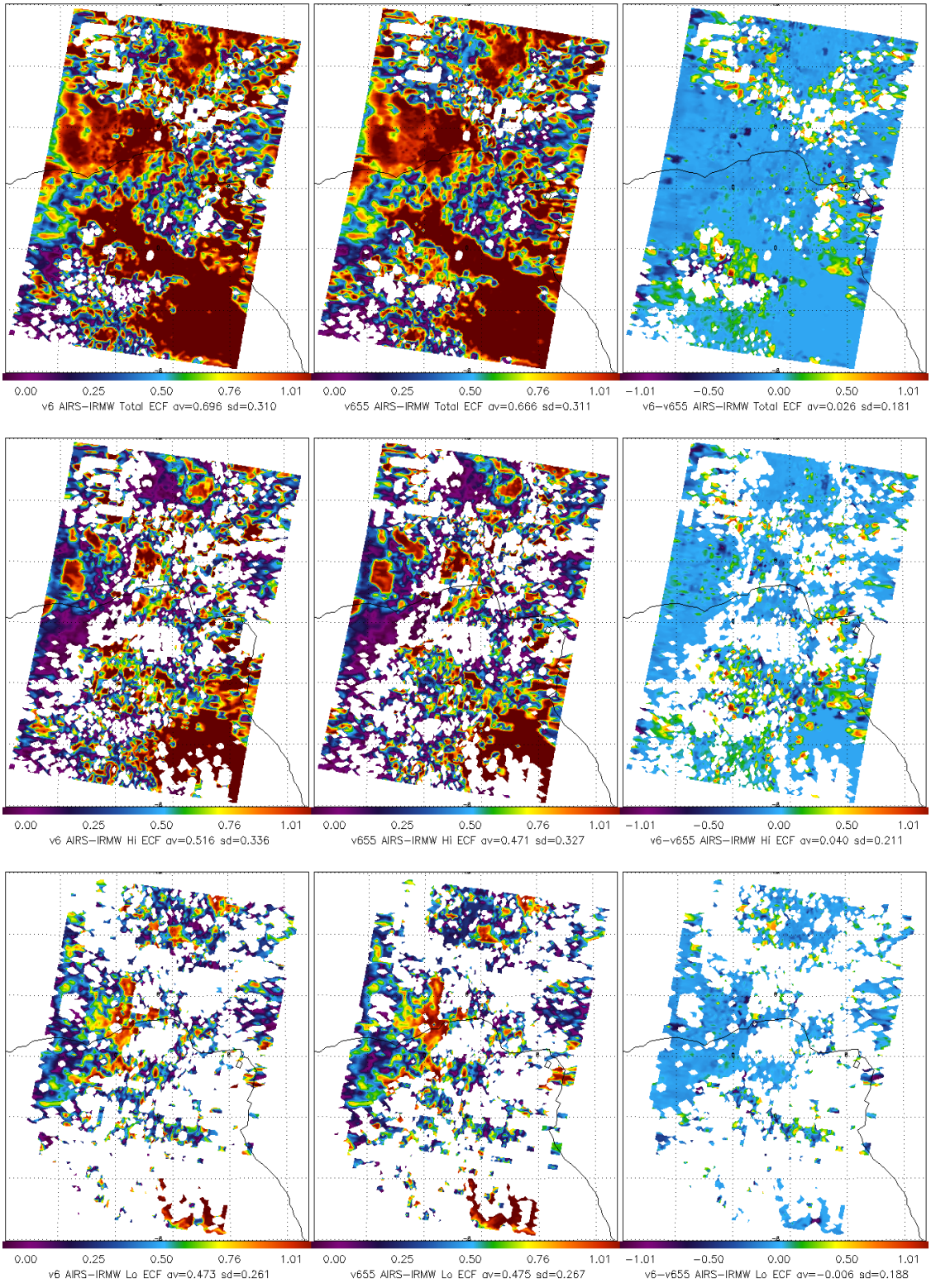


Figure 5.2.4. Same as Figure 5.2.3 except for all clouds with ECF < 0.05 filtered out.

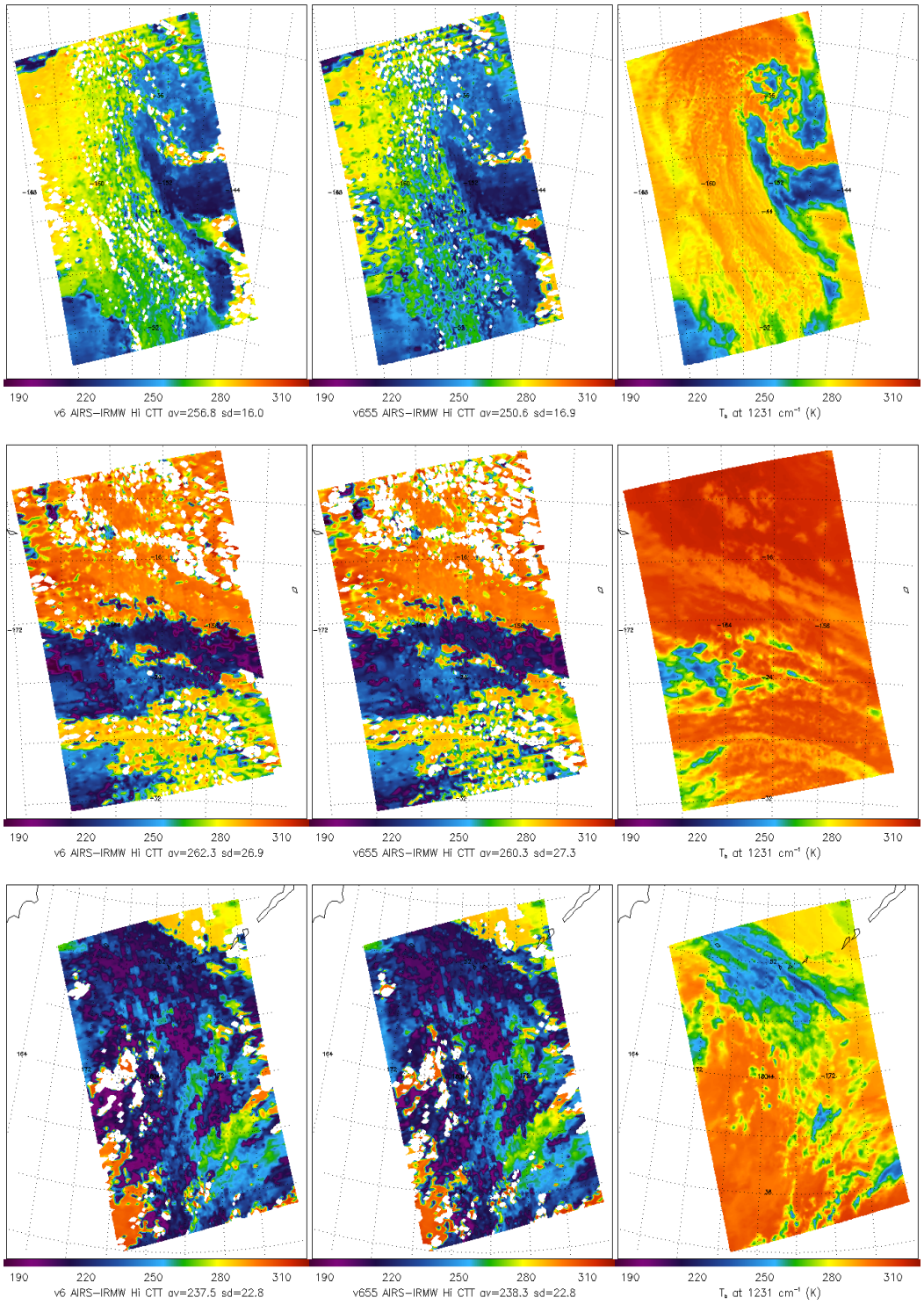


Figure 5.2.5 Three more example granules on July 1, 2011 for (i) granule 004 (upper row), (ii) granule 005 (middle row), and (iii) granule 008 (lower row) for v6 upper level CTT (left column), v655 upper level CTT (center column), and the 1231 cm^{-1} Tb (right column). As previous figures show, the average and standard deviations for the full granule are reported for upper level CTT. All values of ECF are shown with no filtering in order to illustrate the problematic cloud retrievals. If a filter with ECF < 0.05 is applied (not shown), these issues are largely eliminated.

5.2.2 Global-scale Monthly Level 3 Comparisons of Cloud Products Between V6 and V7

Contributor: Brian H. Kahn

Data

Below is a head-to-head comparison of v6 versus v6.55 global-scale L3 monthly cloud products. All data are obtained from the Standard and Support L3 hdf files using the combined IR/MW (AIRX3STM and AIRX3SPM). While only one month is shown in this brief evaluation, a series of spot checks were done over a few other months (including January and July 2011) and the results obtained for other time periods are generally similar in behavior. The variables evaluated are the following: (1) the total CloudFrc variable for separate ascending and descending orbits (CloudFrc_A and CloudFrc_D), (2) the effective cloud fraction weighted CloudTopTemp variable for separate ascending and descending orbits (CloudTopTemp_A and CloudTopTemp_D), (3) the fine cloud fraction variables for separate ascending and descending orbits (FineCloudFrc_A and FineCloudFrc_D), and (4) cloud_phase_3x3 for separate ascending and descending orbits (cloud_phase_3x3_A and cloud_phase_3x3_D). No independent truth data was used in this study. The ice cloud microphysical variables are not evaluated.

Methodology

No additional quality control was applied to the monthly L3 fields and were used in the form provided in the files. The conclusions of the Level-3 results are best understood after reading the Level-2 daily snapshot evaluation presented in a separate section. The reader is referred to that section for a fuller set of details regarding the degradation in some of the cloud products from v6 to v6.55 at the pixel scale.

Conclusions

1. Adverse changes in v6.55: There are substantial differences in the CloudFrc variable from v6 to v6.55 that strongly depend on latitude/longitude, land versus ocean, and ascending versus descending. In particular the CloudFrc magnitudes increase by 0.05-0.1 in much of the extratropical storm track regions in both hemispheres in v6.55. In contrast, there is a slight reduction in CloudFrc in the tropical latitudes. The differences between v6 and v6.55 are most exaggerated during the ascending overpass over tropical land masses and the extratropical storm tracks. There are reduced differences between v6 and v6.55 over tropical land during the descending overpass, however. Furthermore, the changes in the storm tracks from ascending to descending are smaller and more regionally dependent. The differences between v6 and v6.55 are entirely consistent with the pixel-scale problems illustrated in the L2 daily evaluation for these same cloud products.
2. Change in v6.55 that is more difficult to assess whether adverse, neutral, or beneficial: There are differences in the cloud fraction weighted cloud top temperature product (CloudTopTemp) between v6 and v6.55. In the L2 granule scale comparisons, it is abundantly clear that most of this change is caused by the upper level cloud top temperature and/or cloud fraction given the results of the L2 evaluation. The changes are larger in the ascending orbits and are largest over subtropical land surfaces. The descending orbit differences are roughly half the magnitude of the ascending orbit differences with most of the differences on the order of 5K or less.

3. Change in v6.55 that is more difficult to assess whether adverse, neutral, or beneficial: The changes in the fine cloud fraction show very large dependences in the vertical. In the three example pressure levels (346, 648, and 887 hPa) that follow, the amount of CloudFrc can drastically increase or decrease such that other levels have compensating changes such that the total CloudFrc change is not as large. The amount of cloud that is shuffled around vertically is larger than one would infer from the total CloudFrc variable. It is difficult to assess whether this variable is improved or not and will take a great deal of additional effort, and ultimately would require independent data sets to assess whether the changes are beneficial, neutral, or adverse. The bottom line is that these changes are very significant.
4. Adverse change in v6.55: The changes in the cloud_phase_3x3 variable, both for ascending and descending, are larger than expected for the liquid tests (-2 or -1) or the unknown category (0). In the case of unknown, the changes can exceed 5-10% of the total number of pixel counts especially in the subtropics and midlatitude storm tracks. There is an increase in unknown cases in v6.55 in the subtropics during the ascending orbit, while the opposite is true in descending orbits, except for a few regions. The changes for the ice cloud categories are much less, typically less than 1%, except for some increases in a few to several % within high latitude NH land masses and over a few of the more prominent mountains ranges in the mid-latitudes. Both of these increases are unrealistic and are related to additional fictitious clouds being added over land surfaces with poorly characterized emissivity or in the presence of winter/diurnal thermal inversions that can cause some of the phase tests to fail. The problems with total CloudFrc detailed above and in the L2 evaluation are cascading to cloud_phase_3x3.

Discussion

Figure 5.2.6 shows monthly maps for January 2003. The total CloudFrc variable in the upper two rows exhibit substantial differences in the CloudFrc variable from v6 to v6.55 that strongly depend on latitude/longitude, land versus ocean, and ascending versus descending node. In particular the CloudFrc magnitudes increase by 0.05-0.1 in much of the extratropical storm track regions in both hemispheres in v6.55. The pattern of this increase does change from ascending to descending but overall an increase in v6.55 CloudFrc in midlatitude storm tracks is consistent with the L2 evaluation. In contrast, there is a slight reduction in CloudFrc in the tropical latitudes during the ascending orbit, with a much smaller difference during the descending orbit although there are regional variations and exceptions. The differences between v6 and v6.55 are also exaggerated during the ascending overpass over tropical land masses, but unlike the extratropical storm tracks, v6.55 shows strong decreases in CloudFrc. The differences are reduced between v6 and v6.55 over tropical land during the descending overpass, however. In general, the changes in the storm tracks from ascending to descending are smaller and more regionally dependent than the tropical land areas. It appears that the differences between v6 and v6.55 are consistent with the pixel-scale problems illustrated in the L2 daily evaluation for these same cloud products.

The lower two rows of Figure 5.2.6 show the differences in the cloud fraction weighted cloud top temperature product (CloudTopTemp) between v6 and v6.55. In the L2 granule scale comparisons, it is abundantly clear that most of this change is caused by the upper level cloud top temperature and/or cloud fraction. Furthermore, the changes are larger in the ascending orbits and are largest over subtropical land surfaces. The descending orbit differences are roughly half

the magnitude of the ascending orbit differences with most of the differences on the order of 5K or less. While it is difficult to assess whether the weighted cloud top temperature differences are dominated by the cloud top temperature or the cloud fraction is not known. But the bottom line is that there are strong changes between v6 and v6.55 in these products.

Figure 5.2.7 shows the changes in the fine cloud fraction and there are very large dependences in the vertical. In the three example pressure levels (346, 648, and 887 hPa), the amount of CloudFrc can drastically increase or decrease by 10s of relative percentage points such that other levels have compensating changes that lead to smaller differences in the total CloudFrc depicted in Figure 5.2.6. In other words, the amount of cloud that is shuffled around vertically is larger than one would infer from the total CloudFrc variable. The differences between v6 and v6.55 are larger as a function of height and less so as a function of ascending or descending orbit. Note the regionally significant differences between v6 and v6.55 at 648 hPa and 887 hPa where clouds could shift several degrees in longitude (e.g., the tropical eastern Pacific Ocean).It is difficult to assess whether this variable represents and improvement or not and will take a great deal of effort to understand and quantify, and ultimately would require independent data sets to assess whether the changes are beneficial, neutral, or adverse. The AIRS algorithm does not retrieve a vertical profile as this product implies, but does strongly suggest that these vertical redistributions are very significant between v6 and v6.55.

The changes in the cloud_phase_3x3 variable, both for ascending (Figure 5.2.8) and descending (Figure 5.2.9) orbits, are larger than what was initially expected for the liquid tests (-2 or -1) or the unknown category (0). In the case of the unknown category, the changes often exceed 5-10% of the total number of pixel counts especially in the subtropics and midlatitude storm tracks. There is an increase in unknown cases in v6.55 in the subtropics during the ascending orbit, while the opposite is true in descending orbits, except for a few regions. The changes for the ice cloud categories are much less substantial and in line with expectations, typically less than 1%. The exceptions are for some increases in a few to several % within high latitude NH land mass regions and over a few of the more prominent mountains ranges in the mid-latitudes, which is a known problematic area for excessive positive ice tests in v6. Both of these increases are unrealistic and is related to additional fictitious clouds being added over land surfaces with poorly characterized emissivity or in the presence of winter/diurnal thermal inversions that can cause some of the phase tests to fail. The problems with total CloudFrc detailed above and in the L2 evaluation appear to be cascading to cloud_phase_3x3 for the liquid and unknown cloud phase categories.

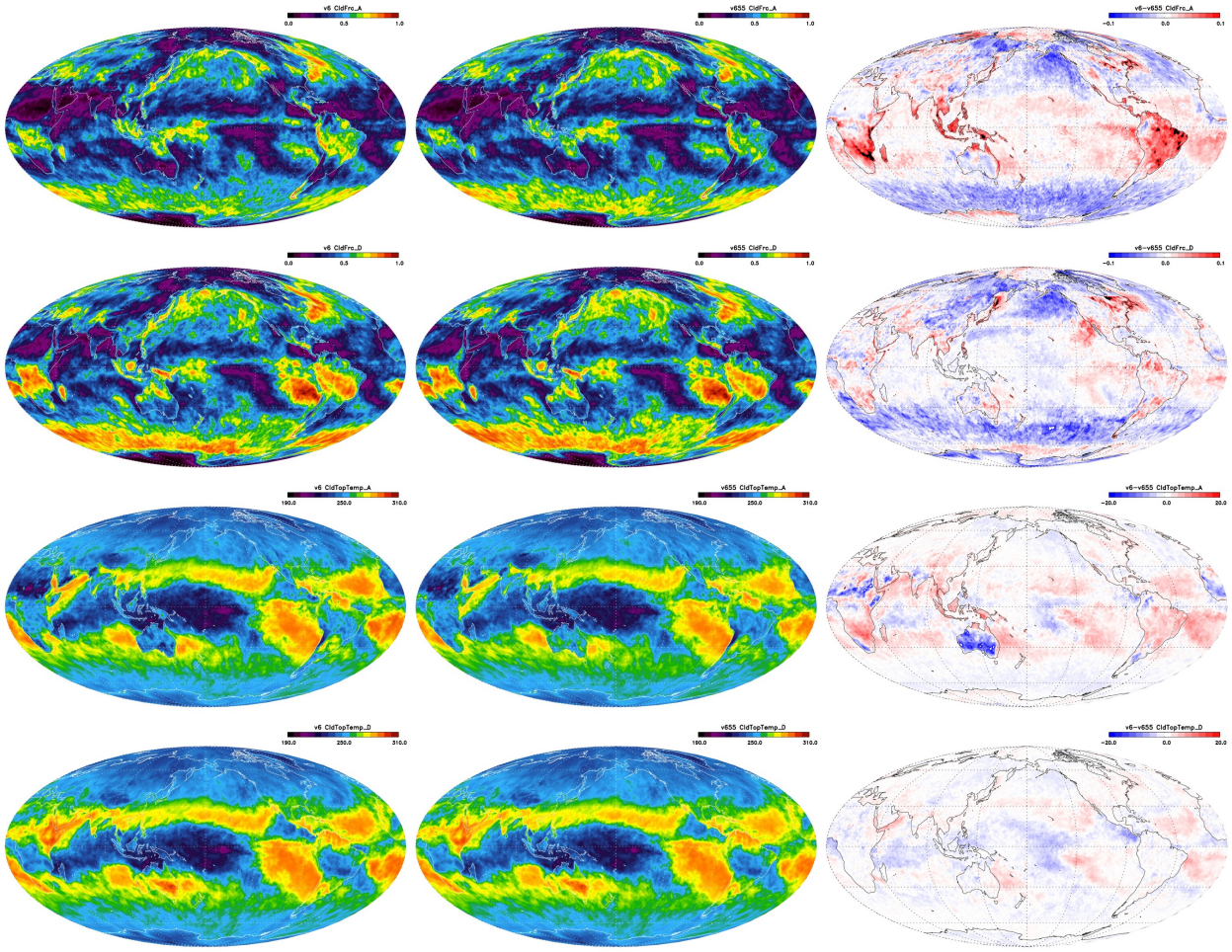


Figure 5.2.6. Left column v6, middle column v6.55 and right column v6.55 minus v6 for total CloudFrc_A (top row), CloudFrc_D (second from top row), layer weighted CldTopTemp_A (second from bottom row) and CldTopTemp_D (bottom row).

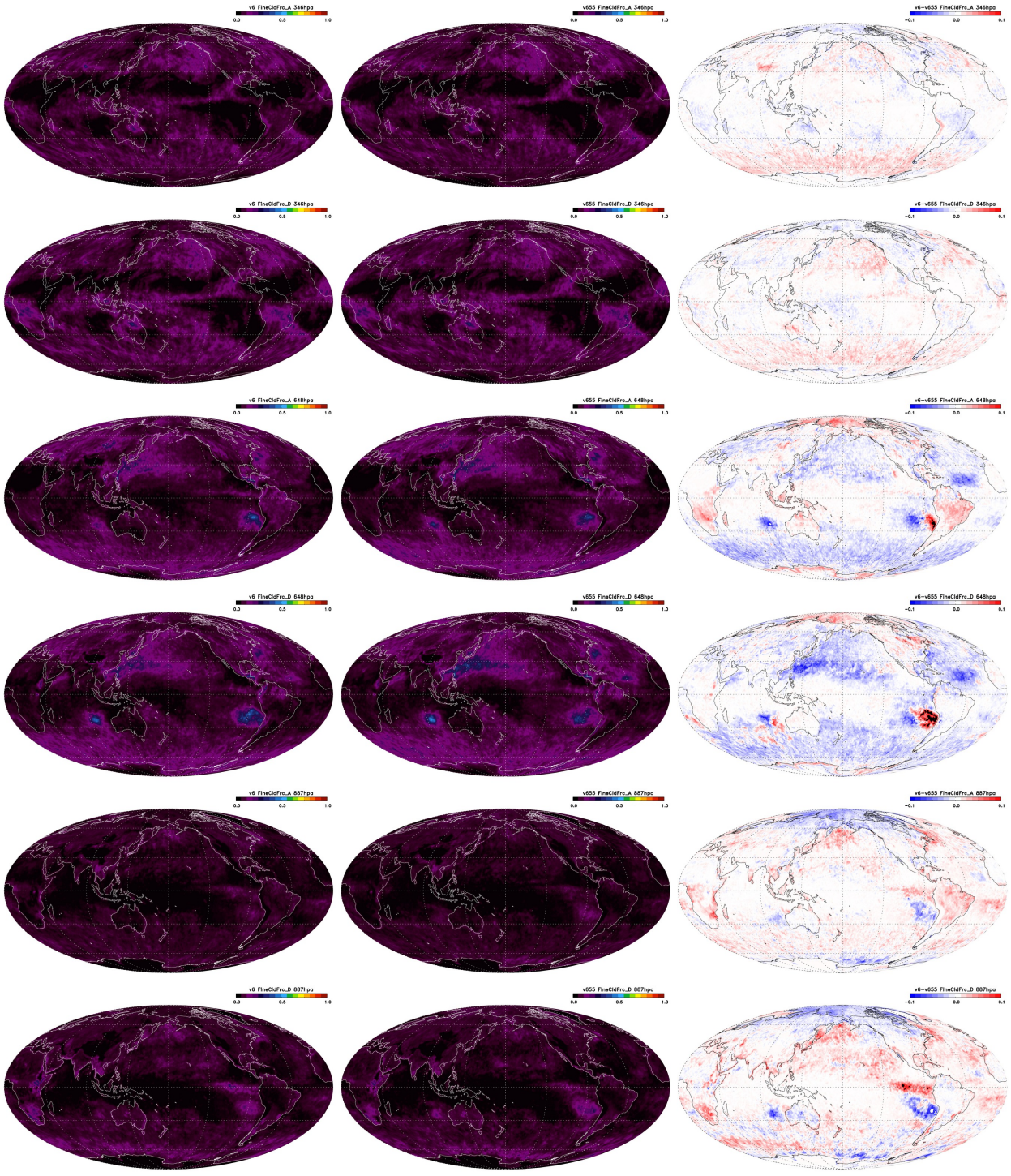


Figure 5.2.7. FineCloudFrc for v6 (left column), v6.55 (middle column), and v6 minus v6.55 (right column). The upper two rows are for 346 hPa, the middle two rows for 648 hPa, and the lower two rows for 887 hPa. Within each pair, the upper (lower) row represents the ascending (descending) orbit.

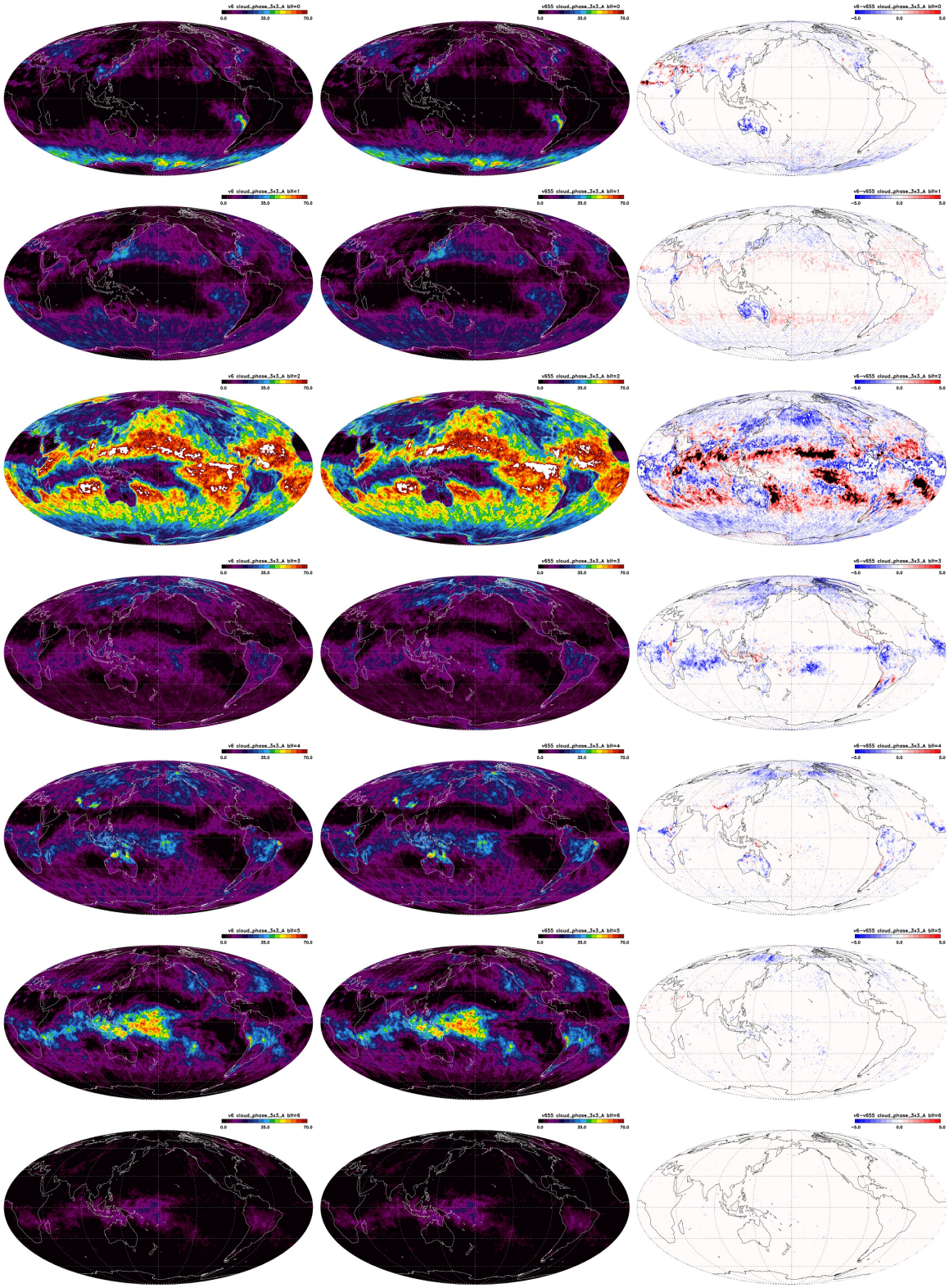


Figure 5.2.8. Cloud_phase_3x3_A for v6 (left column), v6.55 (middle column), and v6 minus v6.55 (right column). From top to bottom is the phase tests -2 and -1 (liquid), 0 (unknown), and +1, +2, +3, and +4 (ice).

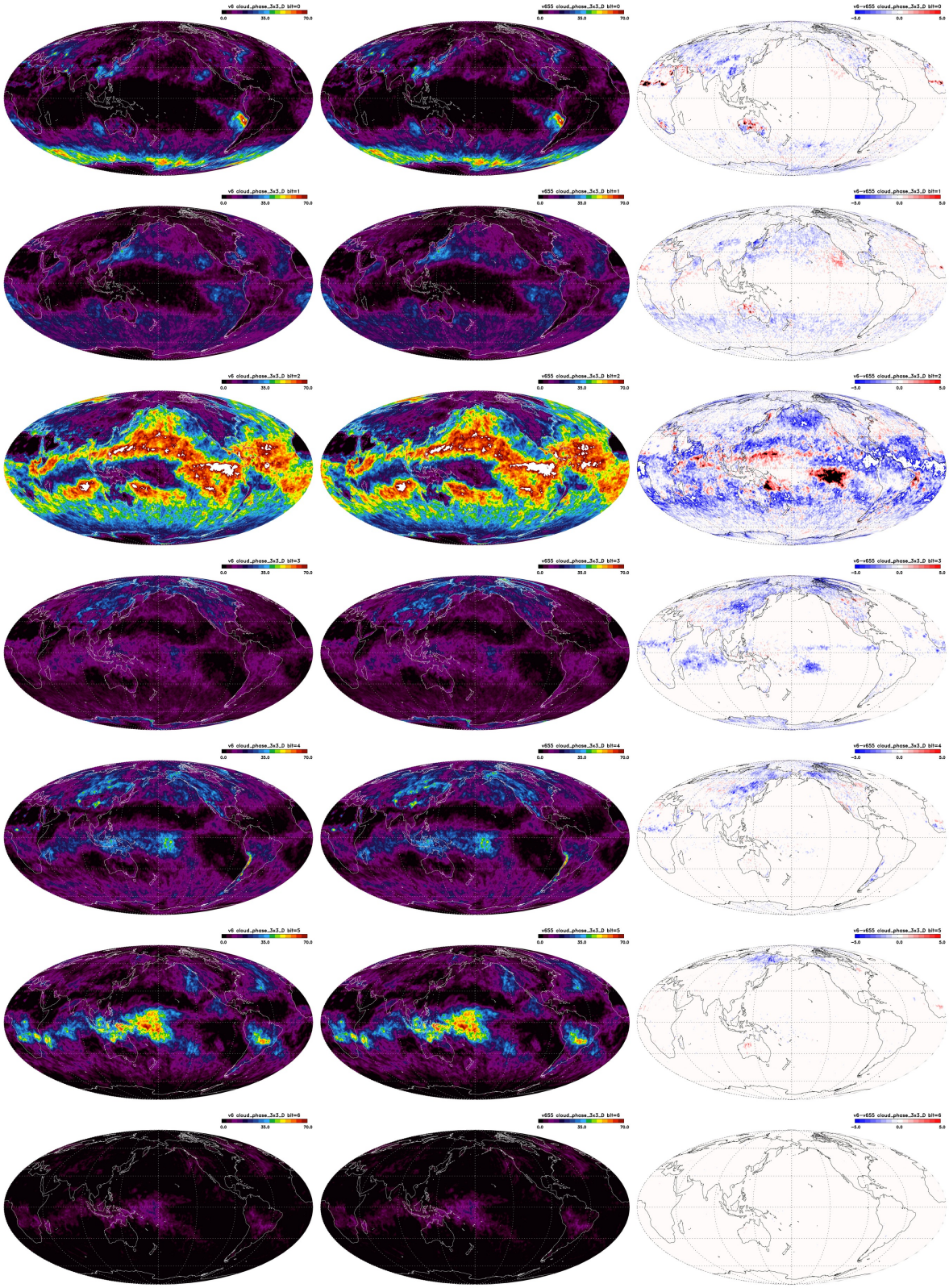


Figure 5.2.9 Cloud_phase_3x3_D for v6 (left column), v6.55 (middle column), and v6 minus v6.55 (right column). From top to bottom is the phase tests -2 and -1 (liquid), 0 (unknown), and +1, +2, +3, and +4 (ice)

5.3 Comparisons of L2 Near-surface Air Temperature and Humidity with In-situ Measurements

Contributor: Yixin Wen, Bjorn Lambrigsten

Data

L2 TSurfAir (henceforth **T**), L2 H2OMMRSurf (specific humidity in units of g/kg, henceforth **q**) and RelHumSurf (henceforth **RH**) from AIRS/AMSU (IR + MW) and AIRS (IR) are examined by comparing with the collocated in-situ measurements from MesoWest over land and International Comprehensive Ocean-Atmospheric Data Set (ICOADS) over ocean.

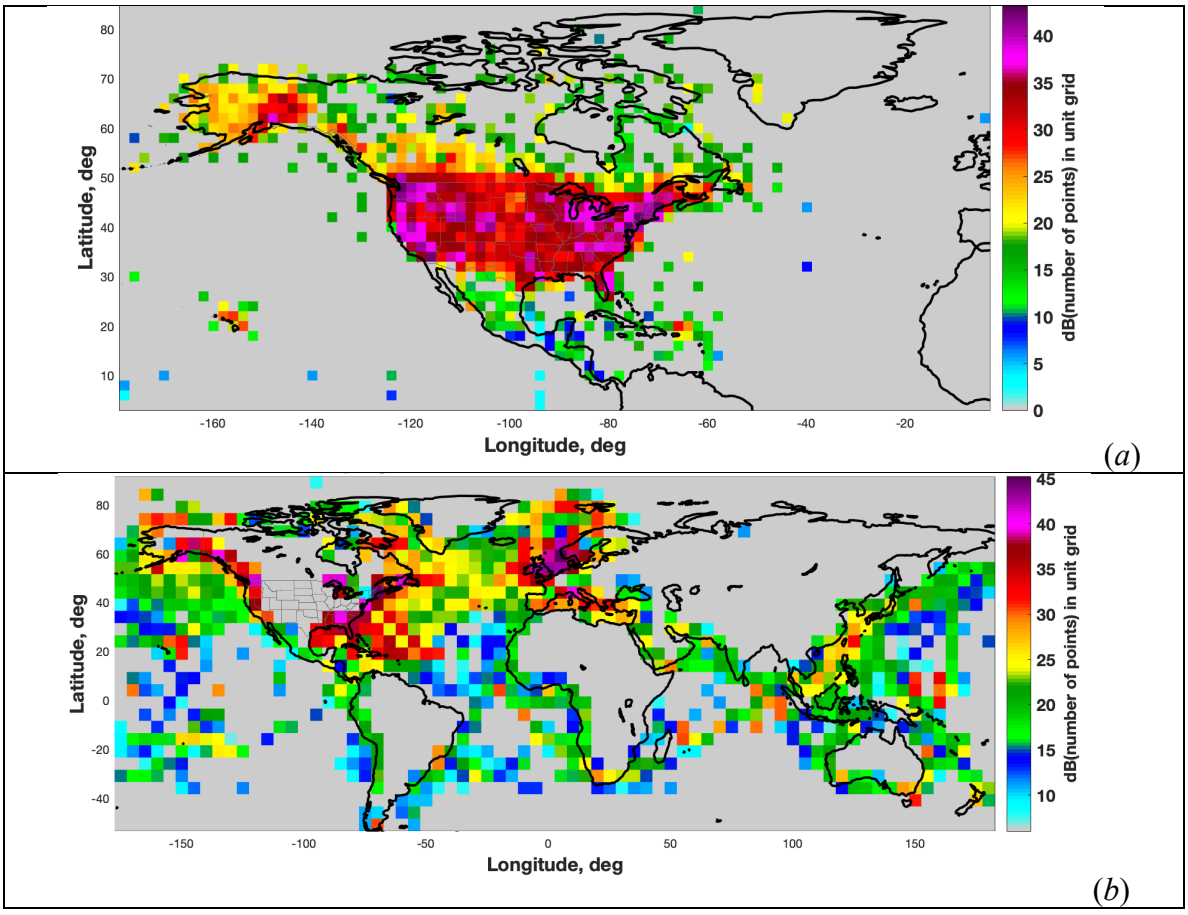


Fig. 5.3.1. Number of reports from (a) MesoWest and (b) ICOADS in July 2011. The number is in dB scale in each grid.

We analyze all AIRS measurements from January and July in 2003 and 2011 for each variable (T and q) and each AIRS data type (airx2sup and airs2sup) from the new version V655. Meteorological data from automated environmental monitoring stations archived in MesoWest platform provide "truth" for near surface temperature and relative humidity. MesoWest is a cooperative program to collect, archive, and distribute environmental observations across the country. Observations of more than 10k stations nationally from more than 70 networks are archived on MesoWest platform. Data is processed and quality controlled every 15 minutes. The number of reports from July 2015 is shown in Fig. 5.3.1a as an example of MesoWest measurement distribution.

The ICOADS, the world's most extensive surface marine meteorological data collection, contains marine surface data spanning the past three centuries. ICOADS data range from early non-instrumental ship observations to measurements initiated in the twentieth century from buoys and other automated platform types. In this research, the observations from ICOADS provide "truth" for near surface temperature and specific humidity over ocean. Fig. 5.3.1b shows the reports number from ICOADS in July 2015 as an example of ICOADS observation distribution.

Methodology

The AIRS L2 retrievals are collocated with MesoWest and ICOADS observations. Strict matching criteria are applied to MesoWest data because of its high spatial and temporal resolutions. The criteria are: temporally not more than 15 min before or after each AIRS pass and spatially not more than 13km from AIRS pixel. If both constraints are satisfied, the spatially nearest data pair is chosen. For ICOADS and AIRS matchup, the criteria are: temporally not more than 1 hour before or after each AIRS pass and spatially not more than 100 km from AIRS pixel. The spatially nearest data pair is selected if both constraints satisfied. The total numbers of matched pairs are list in Table 5.3.1. To analyze the data-pairs, the quality control is performed using the quality flag from the IR+MW products. Any matchups with data quality marked as 2 are discarded.

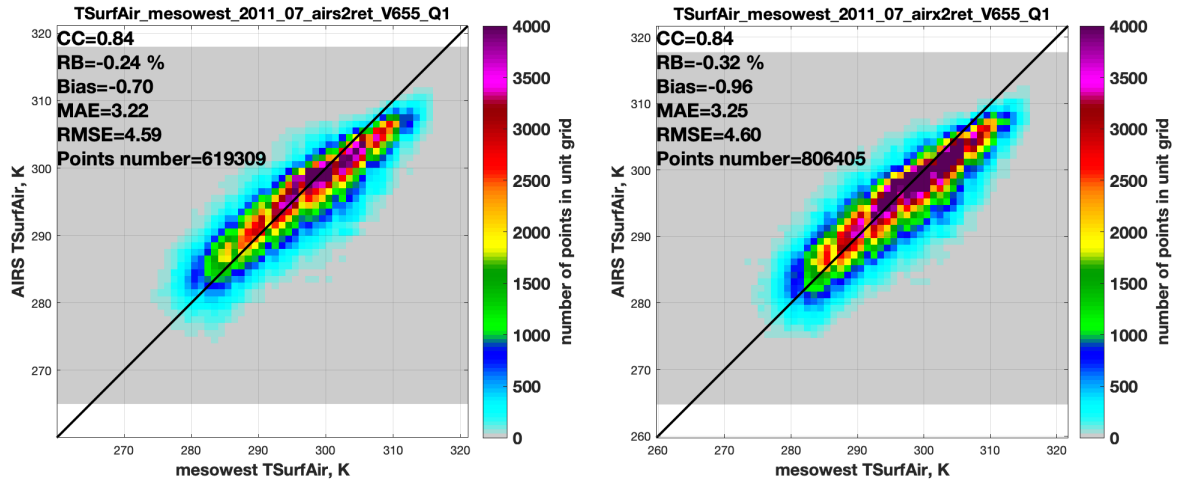
Table 5.3.1 Number of matched data-pairs between AIRS and MesoWest and AIRS and ICOADS.

Sampling	2003		2011	
size	January	July	January	July
MesoWest	161,034	177,165	892,066	929,364
ICOADS	205,357	432,995	238,125	573,910

Results and Conclusions

Comparison results shown in this section are based on July 2011 data. In Fig. 5.3.2, the colored data-density scatterplots of AIRS measurement T against in-situ measurements show that both T from IR only and IR+MW compare well with MesoWest measurements over land and with ICOADS measurements over ocean, although AIRS has a cold bias. Fig. 5.3.2a shows the AIRS T tends to be colder than the MesoWest observations from high air temperature (above ~304 K). Fig. 5.3.2b shows AIRS has high correlations with ICOADS T (CC=0.94), but AIRS has a consistent cold bias of -1.35K. The RMSE of AIRS T is better over ocean than that over land. IR only and IR+MW comparison results are slightly different. But IR+MW has higher yields than IR only temperature measurements.

(a) MesoWest



(b) ICOADS

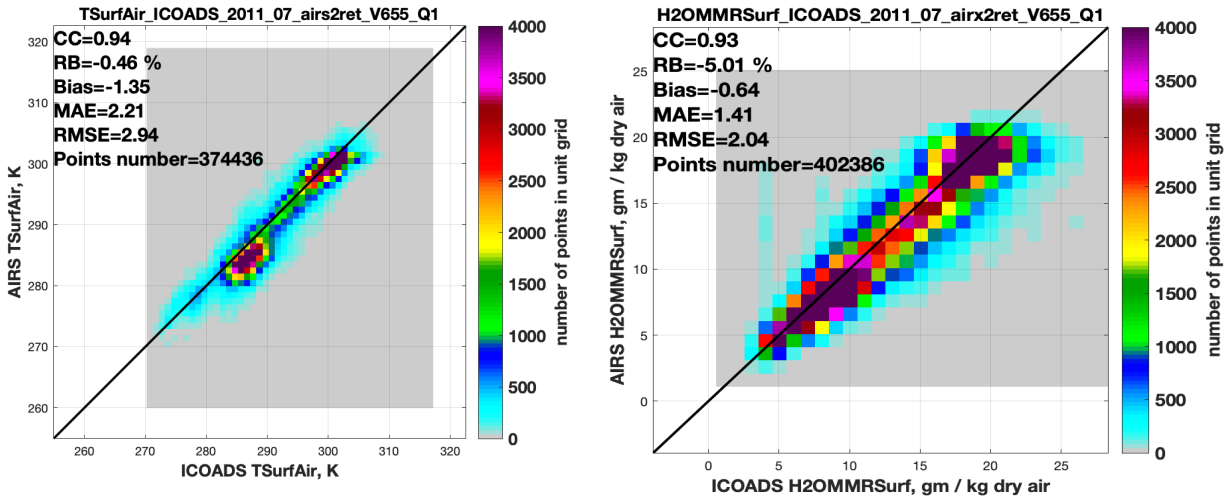
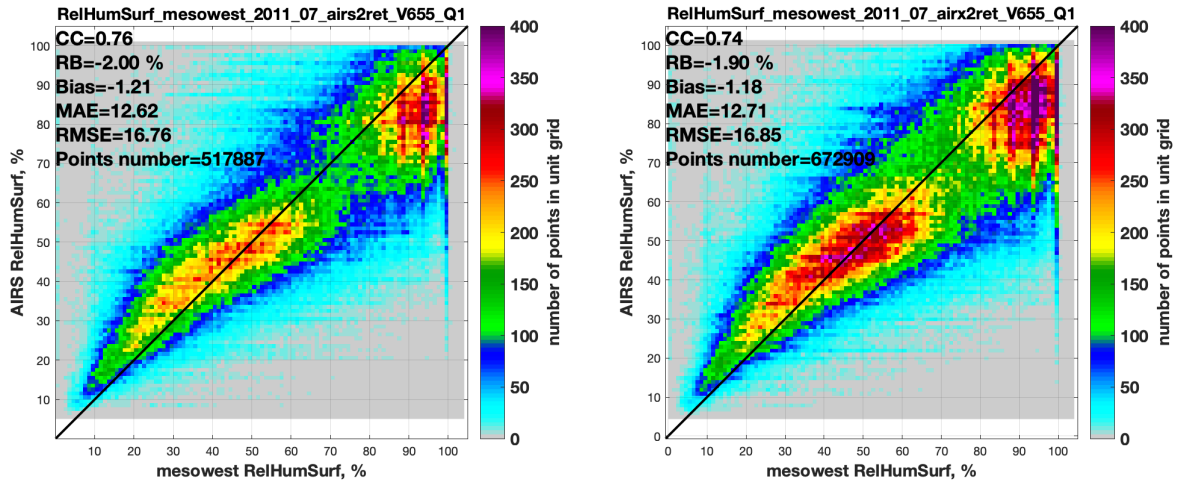


Fig. 5.3.2. Scatterplots with colored data density of near surface air temperature (left) IR-only and (right) IR+MW.

The scatterplot in Fig. 5.3.3a shows AIRS RH has a good agreement with MesoWest measurement. Although the relative bias (RB) of AIRS RH is neglecting (-2%), the scatterplot shows a wide spread range. The overestimation at RH range of 20%~40% cancels out the underestimation for RH >85%. Fig. 5.3.3b shows AIRS measurement q has a high CC of 0.93 and low RB of ~-5% with ICOADS observations over ocean. Similar to T, the comparison results of q from IR only and IR+MW are slightly different. But IR+MW has higher yields than IR only.

(a) MesoWest



(b) ICOADS

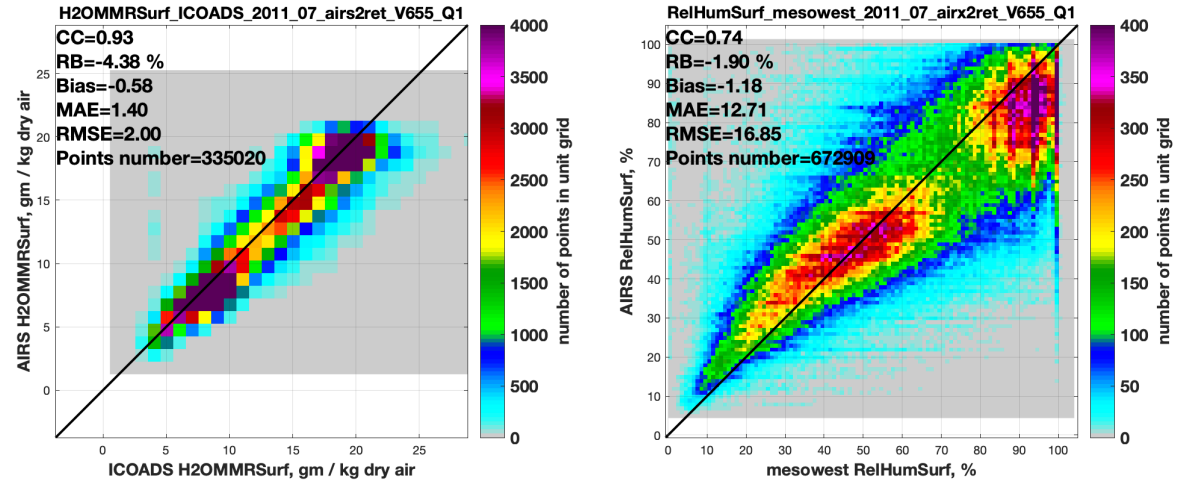
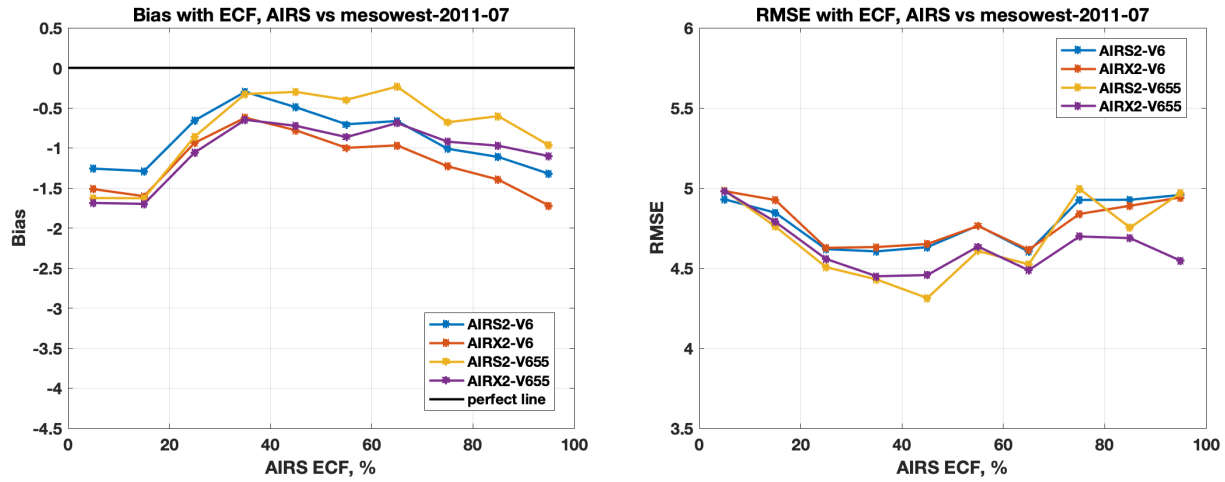


Fig. 5.3.3. Scatterplots with colored data density of near surface relative humidity from MesoWest and specific humidity from ICOADS (left) IR-only and (right) IR+MW.

Fig. 5.3.4a shows the mean biases, the RMSE, between the AIRS retrieved T (both V6 and V655) and collocated MesoWest measurements as a function of AIRS ECF values. The new version V655 shows the smallest bias and RMSE, indicating improvement compared with V6. The difference between IR only and IR+MW is small when ECF < ~30% and becoming larger when ECF > ~60%. Fig. 5.3.4b shows the comparisons between AIRS T and collocated ICOADS over ocean measurements as a function of ECF. The T from new version V655 MW+IR shows the largest negative bias and largest RMSE. Although T from V655 IR only also has the largest negative bias, it has the lowest RMSE among the four products when ECF < 50%, which indicating the new V655 IR-only T retrieval has improvements compared to V6 in cloudless conditions.
Fig. 5.3.5 shows the similar comparisons in Fig. 5.3.4 but for AIRS humidity retrievals. Both V655 data products present the smaller bias and RMSE than V6, indicating significant improvement compared to V6 over land and over ocean.

(a) MesoWest



(b) ICOADS

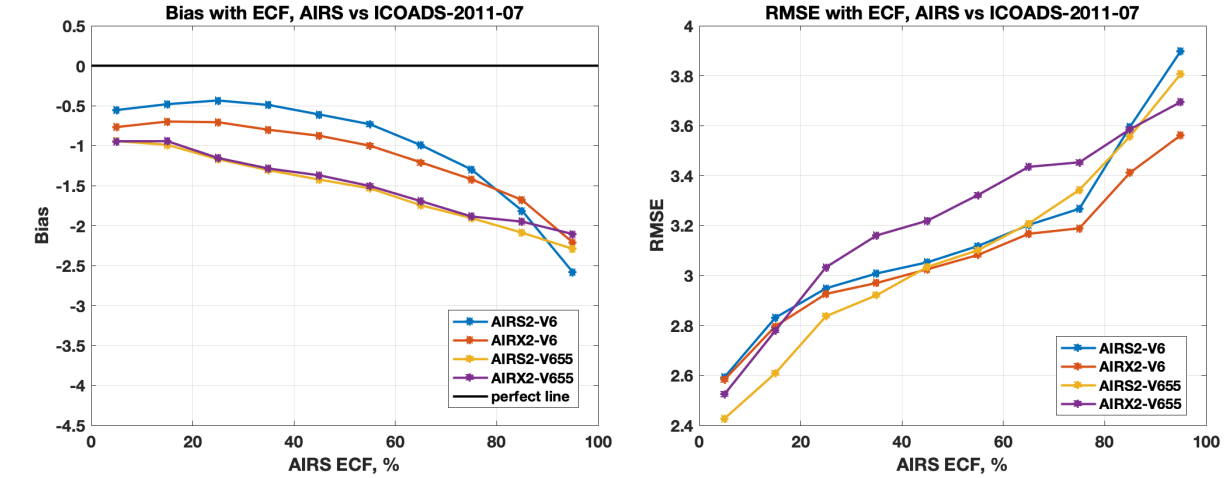
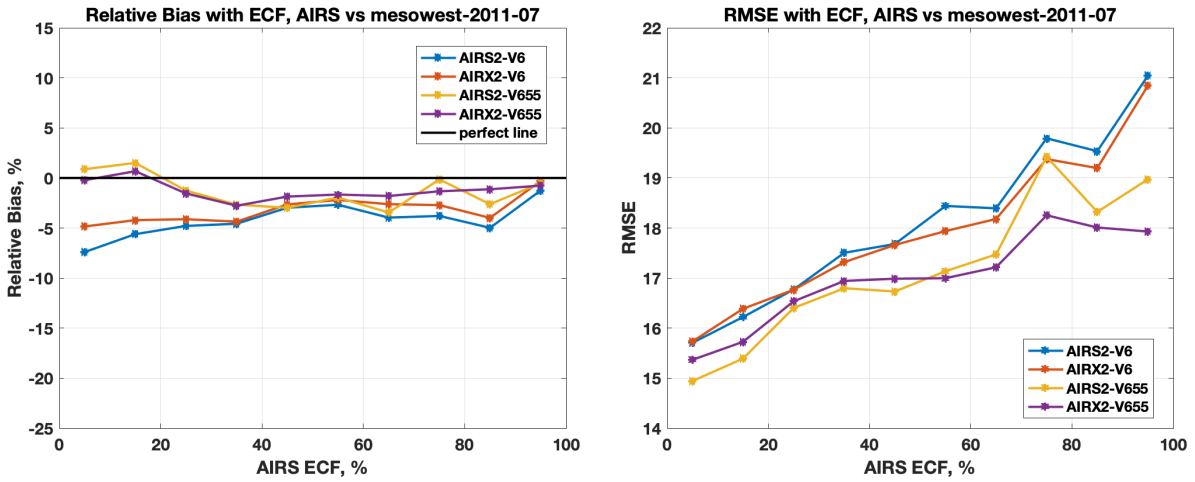


Fig. 5.3.4 The mean bias (left) and RMSE (right) with the collocated (a) MesoWest and (b) ICOADS near surface air temperature measurements (unit in K) by effective could fraction (ECF) from AIRS. Different colors indicate different AIRS retrievals: blue for V6 IR only, red for V6 IR+MW, yellow for V655 IR only, and purple for V655 IR+MW.

(a) MesoWest



(b) ICOADS

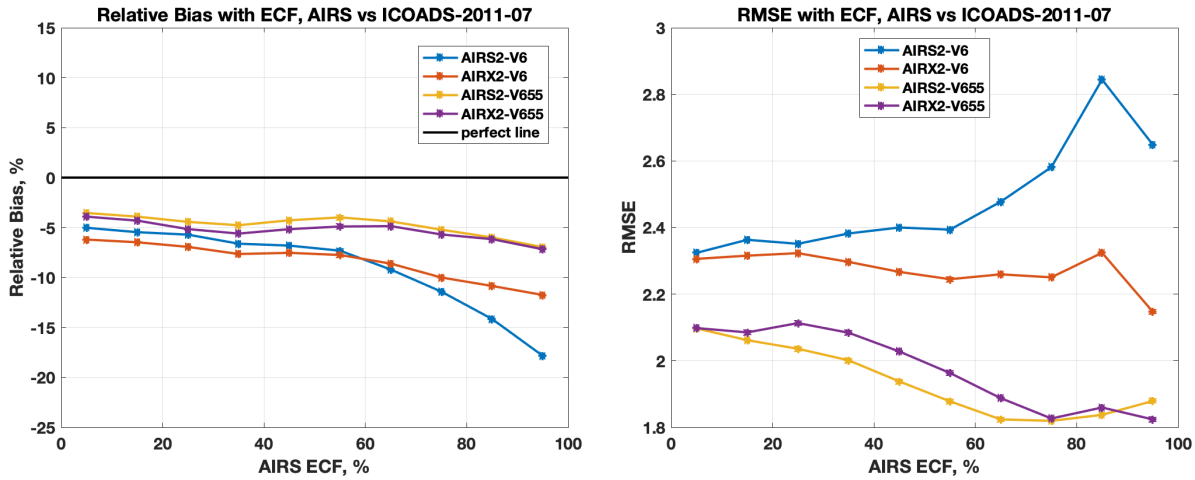


Fig. 5.3.5 The relative bias (left) and RMSE (right) with the collocated (a) MesoWest near surface relative humidity (%) and (b) ICOADS near surface specific humidity (g/kg) by effective could fraction (ECF) from AIRS. Different colors indicate different AIRS retrievals: blue for V6 IR only, red for V6 IR+MW, yellow for V655 IR only, and purple for V655 IR+MW.

5.4 Analyses on AIRS Ozone Vertical Profile and Total Ozone

5.4.1 Comparisons of L2 Ozone Vertical Profiles with Ozonesonde Observations

Contributor: Fredrick Irion

Data

Differences between AIRS 6.0.11 (hereafter V6) and 6.5.5 AIRS profile ozone retrievals were examined by comparisons against coincident ozonesondes. This analysis was meant to be a comparison of profiles and total columns between the AIRS versions and to check for pathologies in V6.5.5.

From the compilation of ozonesonde data collected by the World Ozone and Ultraviolet Radiation Data Centre (WOUDC.org), matchups were made to AIRS L2 ozone profile retrievals

made within 3 hours and 200 km of the sonde launch time and location. (Note that this can allow multiple AIRS observations matched to a single ozonesonde.) Figure 5.4.1 shows the location and number of ozonesondes launched matched with AIRS retrievals as used in this study. (We thank Vivienne Payne for providing the WOUDC database of ozonesonde observations.)

To compare with AIRS, ozonesonde mixing ratios were converted to "slab columns," that is, the integrated number of ozone molecules per square cm between two pressure levels on the AIRS forward-model grid. (This is the same data format as the AIRS support product for trace gases.) Only AIRS slab column observations with quality flags of 0 (best) or 1 (good) were used. Results were compared using relative bias:

$$relative\ bias = \frac{AIRS-sonde}{sonde} \text{ x } 100\ (\%) \qquad (5.4.1)$$

As seen in Figure 5.4.1, the number of sondes coincident to an AIRS observation can vary widely by launch site. The number of AIRS observations to a single sonde can vary widely as well. Rather than presenting averages that can be more weighted to those sonde observations with greater numbers of AIRS coincidences, we use a "median of the medians" approach which first calculates the median AIRS-ozonesonde bias from a single sonde. From these "single sonde" medians, we then calculate the median bias across all sondes launched from a particular site. For inclusion in site-by-site analyses, an ensemble of at least 5 AIRS ozone profiles had to be matched to a single sonde at all available levels, and each site had to have at least 10 such ensembles.

Figure 5.4.2 compares median AIRS relative bias against ozonesondes for V6 and V6.5.5, specific to the sonde location. V6 retrievals tend to be within ±20% in the stratosphere, but can be biased high in the lower stratosphere/upper troposphere. V6.5.5 retrievals are biased low in the mid stratosphere (~ -5 to ~ -20%) at pressures below 50mb, but tend to be biased high in the lower stratosphere (0 to ~40%). V6.5.5 retrievals are biased low near the tropopause at ~200mb, and high in the troposphere (~40%), although we note that tropospheric ozone is only some ~10% of the total column, and AIRS sensitivity to tropospheric ozone is low.

In addition to direct comparisons between AIRS and ozonesonde profiles, we have also applied AIRS ozone averaging kernels to the ozonesonde data ("kerned sondes"). The averaging kernel, **A** (a matrix), contains information on the vertical smoothing and sensitivity of a retrieval, and is calculated for each retrieved AIRS profile (both in Version 6 and 6.5.5). Details on the calculation of **A** for AIRS retrievals can be found in Maddy et al., (2008) and Section 6.2 in this report.

We apply the AIRS averaging kernel reported in the AIRS L2 support product to the sonde data on the 100-level AIRS support grid as follows:

$$\mathbf{x}_{est} = \mathbf{x}_0 + \mathcal{T}\mathbf{A}\mathcal{T}'(\mathbf{x}_T - \mathbf{x}_0) \qquad (5.4.2)$$

where: $\mathbf{x}_{est}$ is the temperature profile as AIRS "should have" seen it given its sensitivity to the *a priori* and limited vertical resolution,

$\mathbf{x}_0$ is the *a priori* temperature profile,

$\mathbf{x}_T$ is the radiosonde profile,

A is the averaging kernel,

$\mathcal{T}$ is a matrix describing the retrieval trapezoids (see Susskind et al., 2003), and

$\mathcal{T}' = \left[\mathcal{T}^T\mathcal{T}\right]^{-1}\mathcal{T}^T$ is the least-squares inverse of $\mathcal{T}$.

The "Degrees of Freedom" (DOF) is the trace of the averaging kernel matrix, and is a measure of the information content of the retrieval. Note that V6 and V6.5.5 have different *a priori*, and that the number of retrieval trapezoids on which the retrievals are performed has increased from 9 in V6 to 20 in V6.5.5. Note also that averaging kernels are specific to each AIRS retrieval. Thus, the "kerned sonde" profile will vary not just by the original sonde profile, but also by the AIRS observation.

Since ozonesondes do not extend to the lower pressures of the AIRS gridding, *a priori* AIRS ozone partial profiles are used to fill in missing elements of $\mathbf{x}_T$ (e.g., pressures below the sonde burst) in Equation 5.4.2. However, we do not extend $\mathbf{x}_{est}$ beyond the sonde pressures. Comparisons of ozone between AIRS and kerned sondes may not be as reliable in the mid-to-upper stratosphere since we are combining in-situ ozonesonde data and *a priori* as 'truth' for this region.

Figure 5.4.3 compares AIRS relative bias against sonde profiles that have had the AIRS averaging kernels applied. Recall that this is a comparison of what AIRS retrieves to what it "should have seen" given a limited sensitivity. For V6, a bias of ~ -5 to ~20% is seen through the profiles, suggesting that the retrieval is imposing a mostly positive bias. For V6.5.5, profiles in the mid-stratosphere (pressures less than ~30mb) agree reasonably well with the kerned sondes (within ±20%), suggesting that the negative bias seen in the same region with unkerned sondes may be because the *a priori* climatology may be biased low and the AIRS retrieval may not have the sensitivity to correct the difference. V6.5.5 biases tend to increase in the lower stratosphere through the troposphere, suggesting that changes in the retrieval methodology from V6 are imposing at least some of the positive biases seen in the more direct comparisons of Figure 5.4.2.

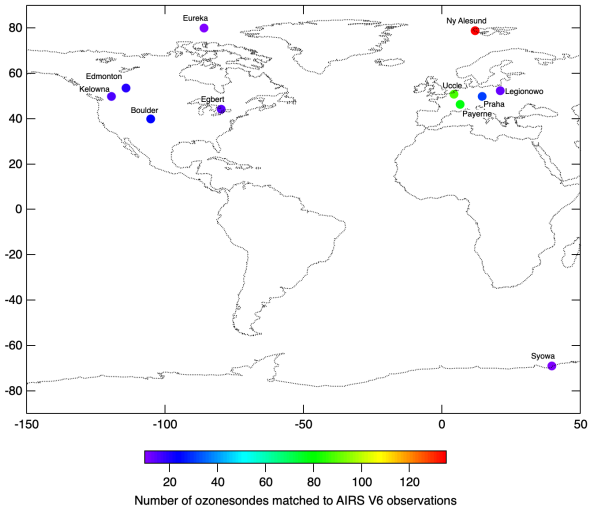


Figure 5.4.1: Location and number of ozonesonde launches matched to AIRS V6 observations as used in this study.

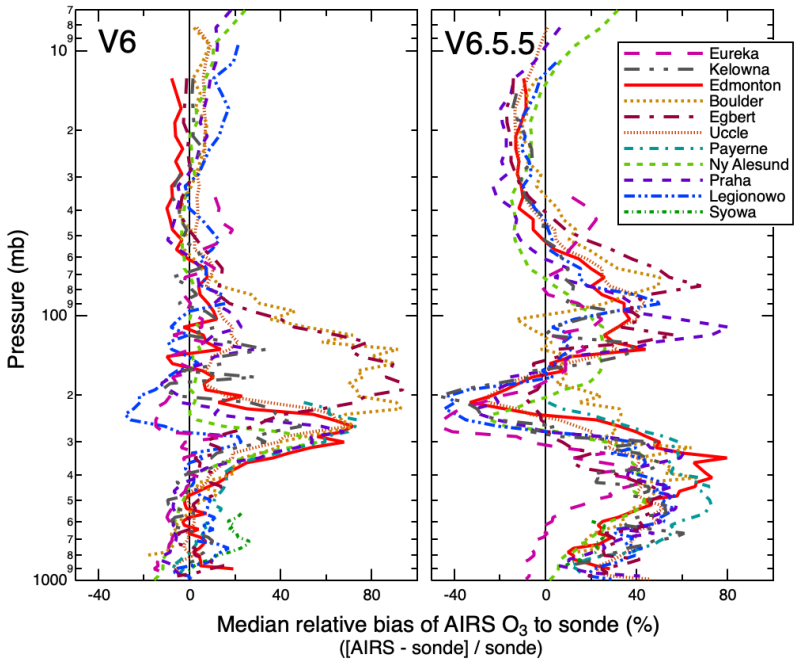


Figure 5.4.2, Left panel: Median profile bias of V6 AIRS ozone profile retrievals against coincident ozonesondes. Right panel: Same for V6.5.5 AIRS ozone. See text for explanation of data selection and reduction.

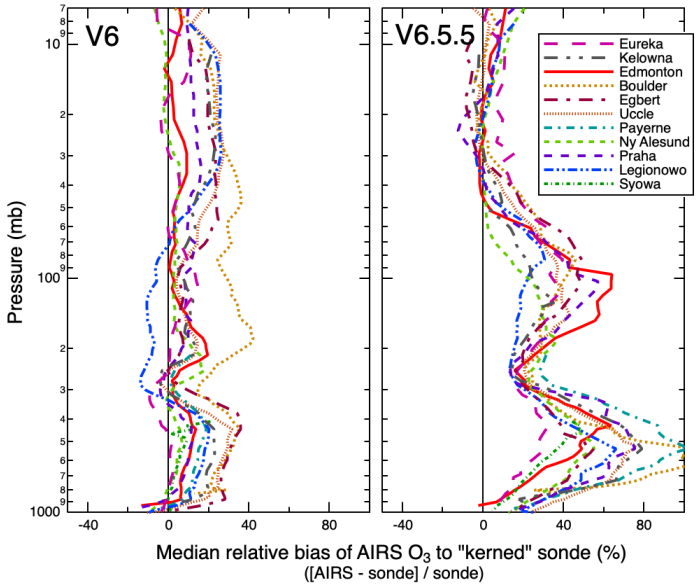


Figure 5.4.3, Left panel: Median profile bias of V6 AIRS ozone profile retrievals against ozonesondes with AIRS averaging kernel applied (see text for formula). Right panel: Same for V6.5.5 AIRS ozone profiles. See text for explanation of data selection and reduction.

5.4.2 Comparisons of AIRS L2 Total Ozone with OMI Retrievals

Contributor: Fredrick (Bill) Irion

Data and Methodology

For total column ozone, we validated AIRS data against two days of Ozone Monitoring Instrument (OMI) data, January 1, 2007, and July 1, 2007. OMI is on the Aura platform (about 16 minutes ahead of AIRS on Aqua), and we used Version 3 of the "Differential Optical Absorption Spectroscopy" (DOAS) Level 2 product (OMDOAO3). (See OMI Team, 2012, and Veefkind et al., 2006.) OMI has a 13x24 km footprint at nadir while the AIRS L2 nadir footprint is approximately a circle of 45 km diameter. Comparisons were one-to-one, with the AIRS observation of quality flag 0 (best) or 1 (good) geographically closest to an OMI footprint selected, but not more than 60 km away. As OMI relies on backscattered UV radiation, only sunlit measurements of AIRS could be used.

Figure 5.4.4 compares the relative bias ([AIRS – OMI] / OMI) in total column for January 1, 2007 for AIRS V6 and V6.5.5, while Figure 5.4.5 shows the same for July 1, 2007. Note that for both dates regional biases have been significantly reduced for V6.5.5. Table 5.4.1 summaries average relative biases of AIRS V6 and V6.5.5 ozone columns against OMI, number of matched observations, and average DOFs for the two days. Improvements from V6 to V6.5.5 are seen against OMI for the total column bias, and the standard deviation. Significant increases in the average DOFs are also seen, most likely from the increased number of basis functions in the retrieval (see section 3.1.3). However, the number of matched observations has decreased from V6 to V6.5.5, indicating that retrieval changes and/or modifications to quality control are reducing the AIRS retrieval yield.

Table 5.4.1: Summary of average relative biases and ancillary results for AIRS V6 and V6.5.5 against OMI total ozone columns

Date	AIRS Version	Average (AIRS-OMI)/OMI bias (% ± 1σ std. dev)	No of matched observations	Average DOF
Jan 1, 2007	V6	1.79 ± 7.19	277418	1.56 ± 0.28
	V6.5.5	1.45 ± 3.78	227394	2.48 ± 0.70
July 1, 2007	V6	3.79 ± 6.90	330949	1.64 ± 0.31
	V6.5.5	3.21 ± 4.31	276634	2.69± 0.67

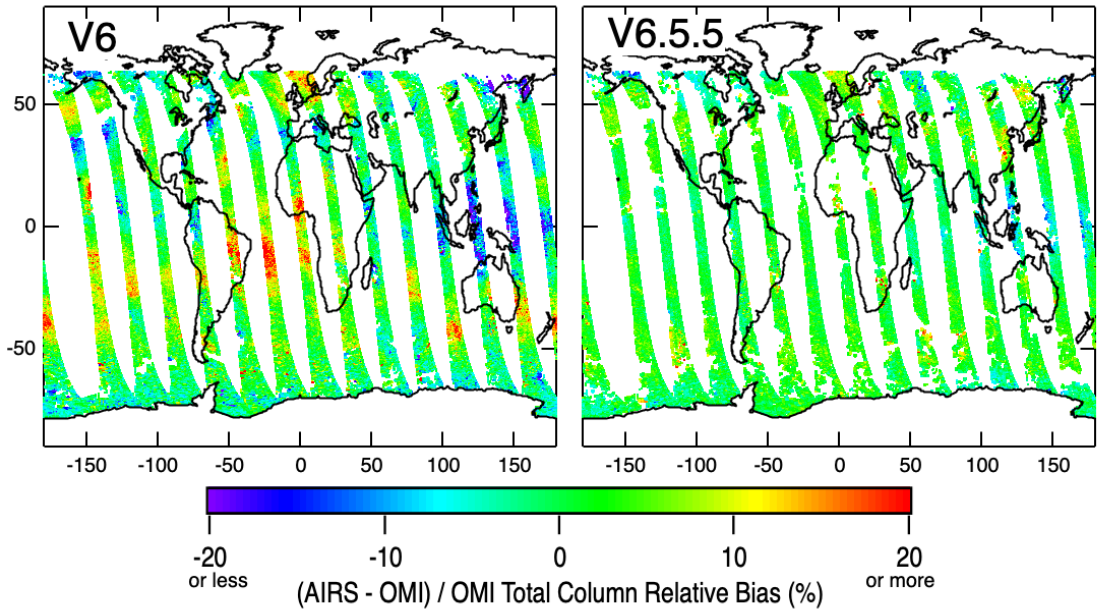


Figure 5.4.4. Left panel: AIRS V6-OMI relative ozone column bias for January 1, 2007. Right panel: AIRS V6.5.5-OMI relative ozone column bias. Note that the color scale has been restricted for clarity.

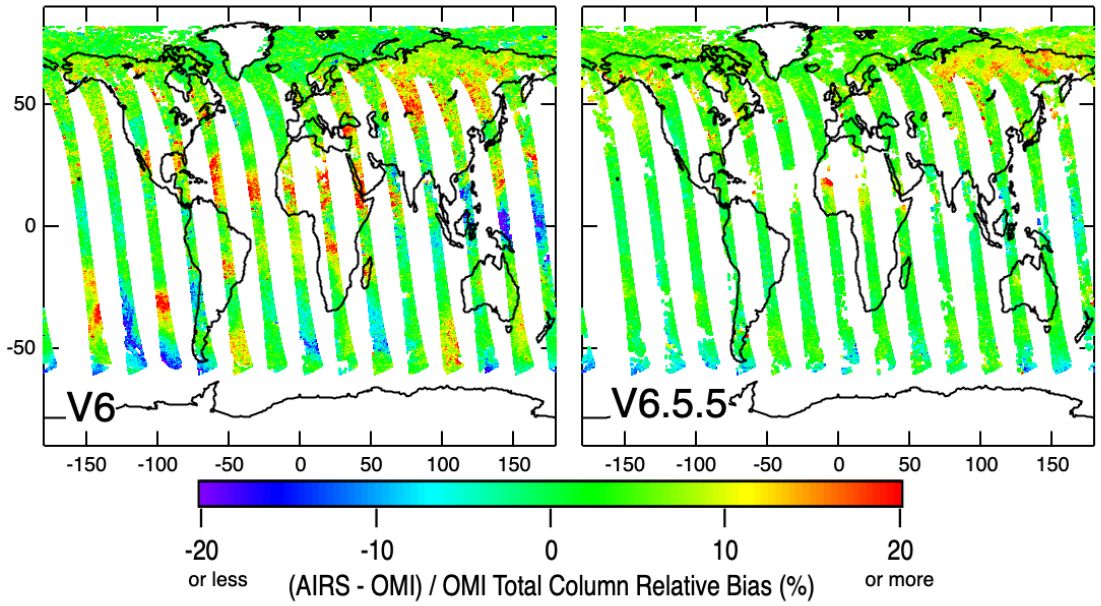


Figure 5.4.5, left panel: AIRS V6-OMI relative ozone column bias for July 1, 2007. Right panel: AIRS V6.5.5-OMI relative ozone column bias. Note that the color scale has been restricted for clarity.

5.4.3 Evaluation of V7 AIRS O_3 Retrieval and Information Content Over Antarctica

Contributor: Evan Fishbein

This section compares Antarctic stratospheric ozone from the version 6 and 7 AIRS retrieval systems. Unlike temperature and water vapor, ozone is retrieved only during the final retrieval step and uses a climatological first guess. This analysis will assess the ability of the version 7 ozone product to capture anomalous structures poorly represented in any annual-mean ozone climatology, specifically the southern hemisphere ozone hole using data from the Dumont d'Urville Station (DUS).

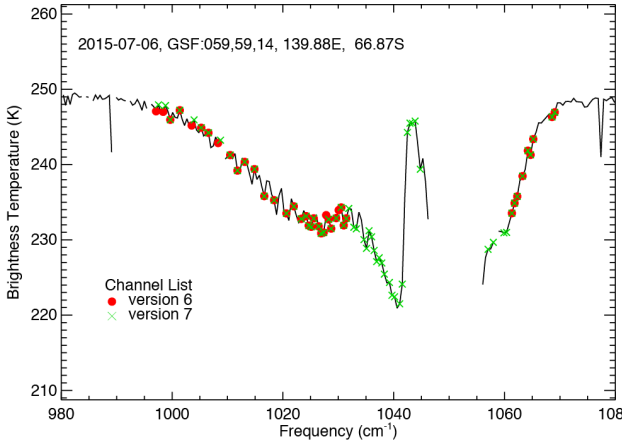


Figure 5.4.6: AIRS ozone spectra from granule 59, scan-line 59, footprint 14 on 06 July 2015. Green crosses are the channels used in the V7, red circles are used in V6, and green crosses over red circles are used in both.

5.4.3.1 Algorithm changes specific to O3

The version 7 algorithm has three changes specifically affecting the O_3 product:

1. The channels used to retrieve O_3,
2. The O_3 retrieval trapezoids, and
3. The O_3 climatology in the lower stratosphere.

Version 7 included more channels sensitive to O_3 in the upper troposphere and lower stratosphere in the 9.6μm O_3 band. Figure 5.4.6 showing an AIRS spectrum near DUS collected during the middle of the Antarctic winter, indicates the channels used in the V6 and V7 algorithms. The channels between 1030 and 1040 cm^{-1} sense O_3 in the cold lower stratosphere, upper troposphere, while the channels near 1043 cm^{-1} sense mid to upper stratospheric O_3.

Version 6 used 9 trapezoids, all but one below 10 hPa, while version 7 used 20 trapezoids. Figure 5.4.7 compares the two sets of trapezoids. The additional channels support retrieving with finer resolution in the upper troposphere and stratosphere. Note that while version 7 uses 20 trapezoids, the two trapezoids covering the mesosphere (above 1 hPa), are set during initialization and are not updated.

Prior to version 7, the O_3 first guess was interpolated from the middle atmosphere trace gas climatology for the COSPAR international reference atmosphere (Keating, 1990). Version 7 added a temperature-dependent first guess mimicking reduced lower stratospheric ozone of the ozone hole associated with conditions forming polar stratospheric cloud (PSC), leading to heterogeneous chemistry and ozone destruction. Unfortunately, while O_3 destruction is tied to cold temperatures, PSCs and sun light, O_3 amount depends on dynamics and the processing history of air parcels over the winter and spring. The climatology depends on month and latitude and has a nominal and O_3 -depleted PSC states. The temperature-dependent ozone first guess model is tied to 50 hPa temperature, through 3 conditions:

1) $O_{3\text{-normal}}$, if $T_{50hPa} > T_{threshold} + 3K$
2) $O_{3\text{-PSC}}$, if $T_{50hPa} < T_{threshold}$
3) $f\,O_{3\text{-normal}} + (1\text{-}f)\,O_{3\,PSC}$, if $T_{threshold} < T_{50hPa} < T_{threshold} + 3K$

$f = (T_{50hPa} - T_{threshold})/3K$

Figure 5.4.8 summarizes the differences in the new and old climatologies. The lower panel shows the latitude-dependence of the PSC and nominal O_3 mixing ratios at 50 hPa, the level of greatest dependence. The largest depletion is 1.5 ppmm and occurs at the pole. The middle panel

shows the annual cycle of the two climatologies at 75˚S, showing the greatest depletion occurs in mid-September. $T_{threshold}$ is also monthly-dependent and in July, the coldest month is 184.7K. The upper panel shows the vertical profiles of nominal (above 187.7K) and depleted (below 184.7K) O_3 at DUS. The difference is negative in the lower stratosphere, but positive in the upper and the maximum depletion is 0.5 ppmm.

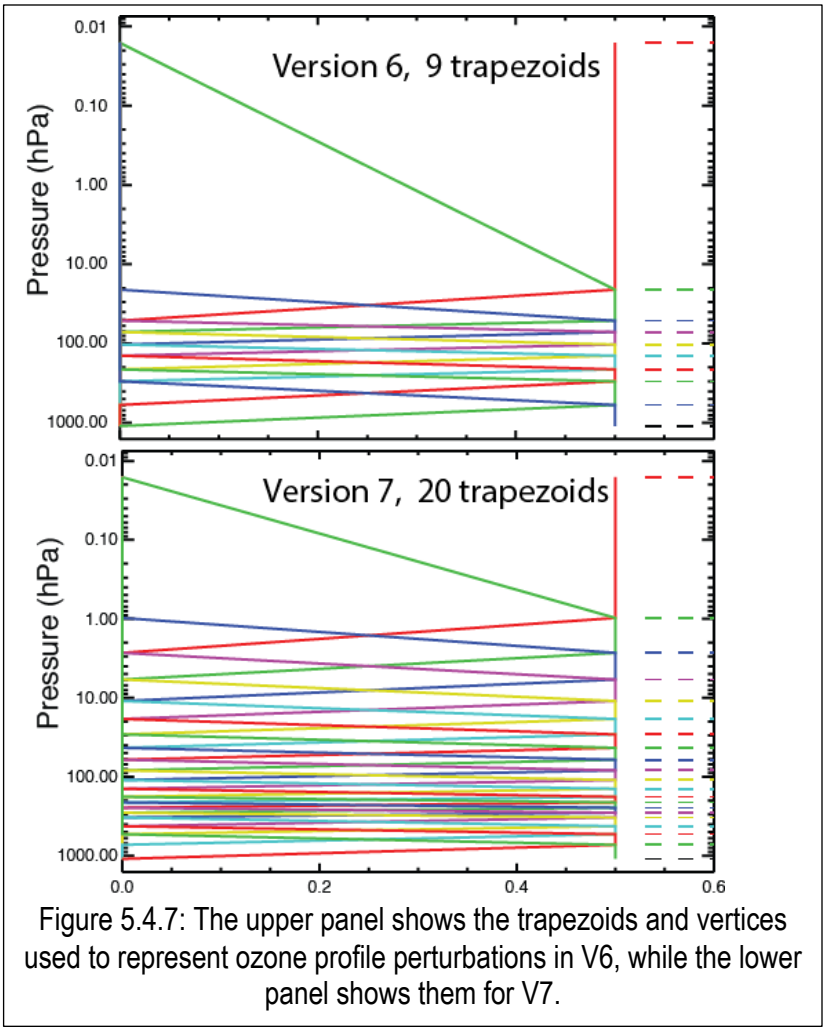


Figure 5.4.7: The upper panel shows the trapezoids and vertices used to represent ozone profile perturbations in V6, while the lower panel shows them for V7.

5.4.3.2 Dumont d'Urville Antarctic: Station data used in analysis

These analyses were to assess the capability of the version 7 algorithms to sense O_3 , as opposed to improving the initial guess and thereby the final solution. We identified anomalous Antarctic conditions, where a climatology would perform poorly. Figure 5.4.9 shows maps of total O_3 collected from the GOME-2 sensor on MetOp-A and the SCIAMACHY sensor on Envisat, Antarctic stations contributing to the WMO Global Atmosphere Watch (GAW) 2015 Antarctic Bulletins, and the time series of total O_3 collected at DUS during 2015 compared against the 20-year median O_3 time series. The 2015 O_3 hole recovered unusually slowly and the normally high O_3 around the edge of the polar vortex was absent by the end of August 2015. Normally total O_3 at DUS has its highest values during October, but O_3 did not recover over DUS December and was wildly fluctuating in November as the elongated O_3 hole rotated on and off DUS. In the remainder of this section we examine measurement diagnostics, compare profiles of O_3 and time series of total O_3 at DUS. DUS collects O_3 profiles one or two times monthly using ozone-monitoring balloon probes (ozonesondes), and collects total O_3 daily using an up-looking UV-visible Systeme d'Analyse par Observation Zenithale (SAOZ) spectrometer. DUS is sufficiently far north, that total O_3 is collected even during polar night at the south pole.

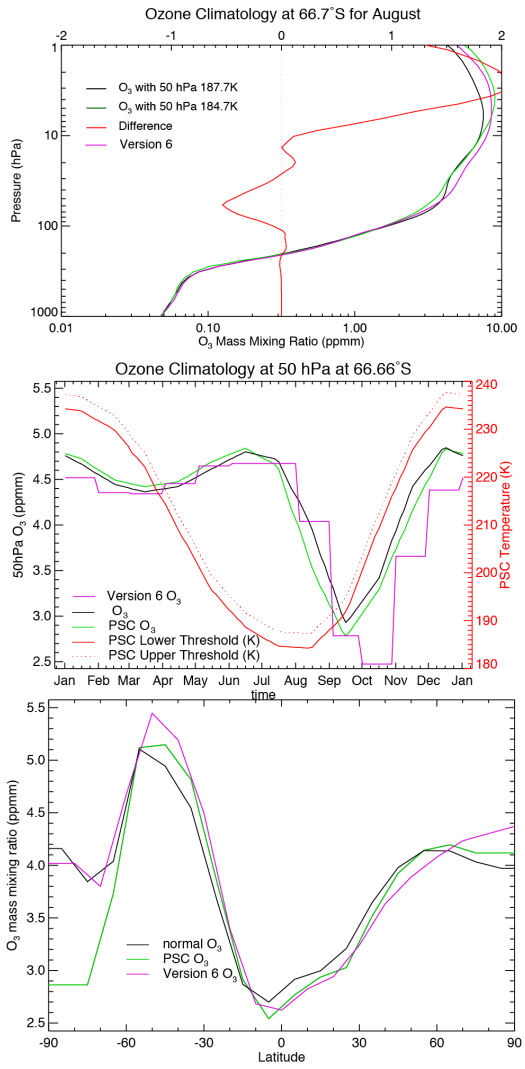


Figure 5.4.8: Comparison of v6 (magenta) and v7 nominal (black) and depleted (green) ozone climatologies. The bottom panel is the latitude-dependence of 50 hPa O_3 during August, the middle panel is the time-dependence at 67˚S (DUS) at 50 hPa, and the upper panel are the profiles at DUS. The red curve is the v7 depleted minus nominal. The red curves in the middle panel show the threshold temperatures for the nominal to depleted climatologies.

5.4.3.3 Averaging kernel comparisons

The averaging kernels map retrieval state vector elements j to vector elements i, $\boldsymbol{A}_{i\,j}$. The averaging kernel between layer mean ozone mixing ratios is:

$$\mathbf{A}' = \overline{\mathbf{X}}\,\mathbf{T}\,\mathbf{A}\,\mathbf{L}^{-1}\mathbf{T}^{T}\overline{\mathbf{X}}^{-1}, \qquad (5.4.3)$$

where $\mathbf{T}$ is the trapezoid matrix and $\mathbf{L} = \mathbf{T}\mathbf{T}^{T}$ is the square of the trapezoid matrix used in evaluating the pseudoinverse of the trapezoids. The trapezoid matrix is the evaluation of the M trapezoid functions, defined at M+3 vertices P_{i-1}, P_i, P_{i+1}, P_{i+2}, at the 100 AIRS levels (N). The index i and pressure P_i identify each averaging kernel (see Figure 5.4.10 for the connection between i and the trapezoid). The matrix $\overline{\mathbf{X}}$ is the diagonal matrix of layer mean ozone mixing ratios, and arises because the retrieval is formulated in terms of fractional change in layer-ozone-amount. $\mathbf{L}$ has the dimensions of the number of trapezoids (N × N), and its inverse is nonsingular, provided the number of trapezoids is smaller than the number of layers and each trapezoid extends over several layers. In the case of perfect averaging kernels, i.e. $\mathbf{A}$ is the identity matrix, the layer-mean averaging kernels, $\mathbf{A}'$, also reduces to the identity matrix. The log-ozone averaging kernels, $\mathbf{A}^{*} = \mathbf{T}\,\mathbf{A}\,\mathbf{L}^{-1}\mathbf{T}^{T}$, varies less with level in the atmosphere and are easier to compare in graphics. Figure 5.4.11 shows the strong state-dependence of averaging kernels from summer, winter and fall conditions for closest matched AIRS ozone profiles near DUS. The summer profile contrast with the winter profile, which has the greater total column ozone owing to descent of ozone-rich mid-stratospheric air into the lower stratosphere during the winter night. For all three conditions, the version 7 averaging kernels are usually sharper and have greater verticality. The greater ozone in the winter profiles, combined with greater density due to cooler temperatures, causes the winter averaging kernels to be sharper then summer kernels and to have much greater maximum values. The spring averaging kernels shows the 20 hPa and 200 hPa averaging kernels mostly unchanged compared to the winter

profiles because ozone is not destroyed at these levels, but from the 40 to 100 hPa levels, the destruction of ozone reduces the amount of ozone capable of absorbing infrared radiation and the shows the loss of sensitivity as indicated by having smaller amplitude averaging kernels.

5.4.3.4 Methodology

In the next section, ozonesondes and AIRS profiles are compared. Comparing AIRS and ozonesonde profiles are complicated by having different vertical resolutions, specifically the difference:

$$\delta x(P) = x_{\text{AIRS}}(P) - x_{\text{sonde}}(P), \tag{5.4.4}$$

has contributions from noise, biases and differences in vertical resolution. Differences associated with vertical resolution can be reduced by smoothing the ozonesonde profile, using vertical resolution information contained in the retrieval averaging kernel. Several problems and approaches are available. Smoothing a profile generally shifts ozone mass from one level of the

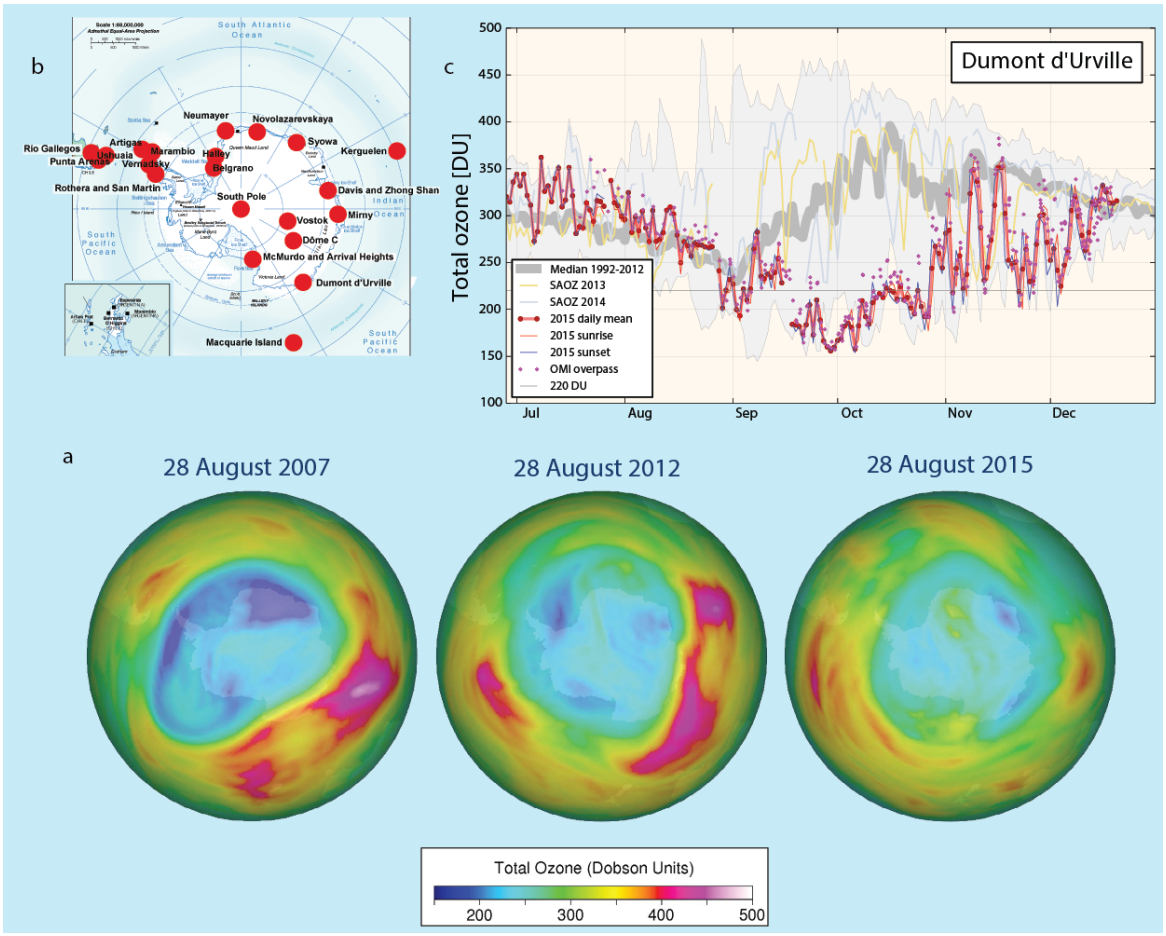


Figure 5.4.9: Total O_3 was unusual during winter 2015. Frame (a) compares total O_3 maps from 28 August 2007, 2012 and 2015. Frame (b) shows the location of total O_3 stations and frame (c) compares time series of total O_3 at Dumont d'Urville Station versus 20-year median. Figures are from Braathen, G., 2015: *WMO Antarctic Ozone Bulletin* no. 1 and 5, 2015.

atmosphere to another, changing the total ozone column. The degree is mass change depends on

the vertical representation (or discretization) of the profile, the units of the profile, e.g volume mixing ratio, mass mixing ratio, density or log mixing ratio and the smoothing kernel. Another problem is ambiguity with smoothing near the end of the profile. Also smoothing operation at levels where the row-sum (verticality) of the averaging kernel is small can be problematic; each will be expanded on in the following section.

We smooth the profiles using "kerning," which combines the averaging kernel applied to the profile with its complement applied to the retrieval background $x_{\text{background}}$:

$$x_{\text{sonde kerned}} = \mathbf{A}\, x_{\text{sonde}} \; + \; (\mathbf{I} - \mathbf{A})\, x_{\text{background}}, \tag{5.4.5}$$

The profile and kerned profile difference:

$$\delta x_{\text{sonde kerned}}(P) = x_{\text{AIRS}}(P) - \; x_{\text{sonde kerned}}(P)$$

is often used as a metric of retrieval performance. However, the kerned and retrieved profiles both relax to the background $x_{i\ \text{background}}$ and δx_i is when the verticality close to zero – agreement does not imply quality or useful information in the retrieved product, just that the averaging kernel is accurate.

Kerning a radiosonde near the ends of the profile is ambiguous once the smoothing kernel extends past the end of the profile. Five approaches are suggested:

1. Apply kerning only at levels where a large fraction (e.g. 90%) of the averaging kernel (verticality) is within the ozonesonde. The problem for broad weighting function is most of the ozonesonde profile isn't compared.
2. Apply kerning at levels where a smaller fraction (e.g. 40%) of the averaging kernel (verticality) is within the ozonesonde, repeating the ozonesonde end value over the entire range of averaging kernel. This produces a biased estimate, especially if the ozonesonde has noise or are large lapse rate at the ends.
3. Merge the ozonesonde profile with a climatological profile, possibly adjusted to match total column ozone from another validation source, and kerning the smoothed profile. The problem here is agreement may be improved near the ends, because the merged profile is more like the retrieval background profile.
4. Truncate averaging kernels outside the extend of the averaging kernel. Numerically similar to 3, depending on how ozonesonde and background profiles are merged and results in more background entering the kerned profile near its top.
5. Apply a truncated averaging kernel normalized by the verticality of the truncated and untruncated averaging kernel. Greater vertical resolution results for the truncated averaging kernels.

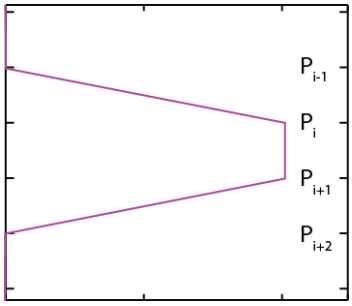


Figure 5.4.10: Averaging kernel vertices

In this analysis, we use approach 4.

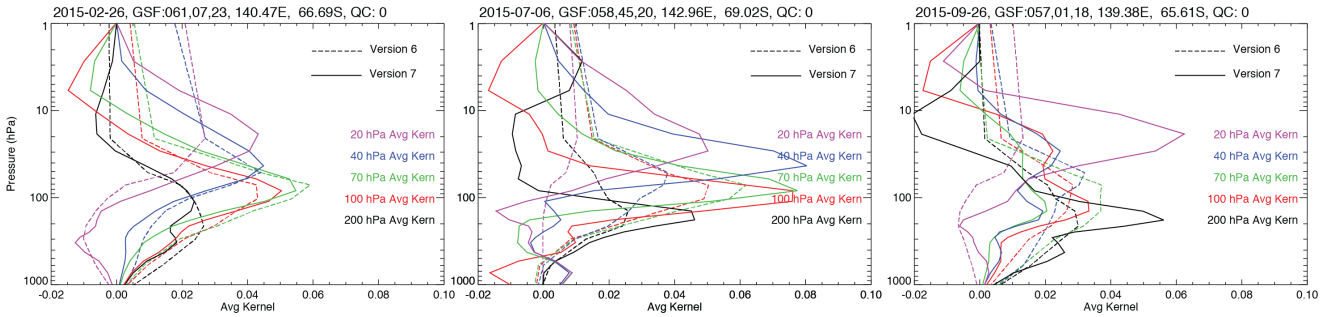


Figure 0: Comparison of version 6 and 7 ozone averaging kernels in the upper troposphere and lower stratosphere for version 7 (solid lines) and version 6 (dotted lines) for summer (2015-02-26), winter (2015-07-06) and spring (2015-09-26). Averaging kernels are show for the trapezoids with bottom vertices at 20, 40 70, 100 and 200 hPa.

5.4.3.5 Kerned profile comparison methodology

Figure 5.4.12 compares ozonesonde profiles with retrieved profiles from the version 6 and 7 products for summer, winter and spring conditions. The upper panel shows the profiles, the lower set shows differences. The ozonesonde profile is shown at its original resolution, layer-mean mixing ratios and kerned using the v6 and v7 averaging kernels and initial guesses. The layer means are calculated for the 100 AIRS layers and located at the log-mean pressure:

$$\bar{P}_i = \frac{P_{i+1} - P_i}{\mathrm{Log}\left(\frac{P_{i+1}}{P_i}\right)}, \qquad (5.4.6)$$

and interpolated to the AIRS 100 levels, but only those levels below the highest level of the ozonesonde. At the resolution of this figure, the layer means provide an accurate summary of the ozonesonde profile.

Comparing the summer and winter ozonesonde profiles, the descend of O_3-rich middle stratospheric O_3 in the winter, without appreciable O_3 destruction leads to a sharp O_3 gradient in the upper troposphere and lower stratosphere. The lower panel shows the differences between the retrieved profiles and the kerned and layer-mean profiles. We include differences with the layer mean profiles to assess the ability of the retrieved profiles to characterize the state of the atmosphere and differences between the retrieved and kerned profiles compare what is estimated and what is expected. The summer panels illustrate an expected and layer-mean comparison. The difference between the kerned and retrieved profiles (lower panel) approach zero as the sensitivity decreases and the differences between the retrieved and kerned profiles has less oscillation, illustrating that small-scale structure in the ozonesonde profile, which is not resolved in the AIRS profile is reduced by the kerning. The version 7 profiles have smaller oscillations in the upper-troposphere through mid-stratosphere, and this suggests that the additional trapezoids in the version 7 state vector have not make the retrieval less stable. The layer-mean profiles agree more than the kerned in the lower troposphere and near the top of the ozonesonde. This is also a region of reduced sensitivity (the averaging kernels shown in Figure 5.4.11 show this), and this suggests possible misrepresentation of the information content by the averaging kernels or error associated with ambiguity with kerning at the ends of the ozonesonde profile.

The winter and spring comparisons have larger differences in the upper troposphere, and the version 7 is generally better in lower stratosphere. However, the winter averaging kernels have less sensitivity in the lower stratosphere than during the spring, the time of the ozone hole. The layer-mean comparison has smaller differences than the kerned sometimes, and this suggests misrepresentation of information by the averaging kernel. There are also larger oscillations in the spring mid and upper troposphere in the version 7 profiles, suggesting that with less ozone in the profile above this level, the number of channels sampling this part of the atmosphere may be fewer and the retrieval is more unstable. This is surprising because the spring averaging kernels

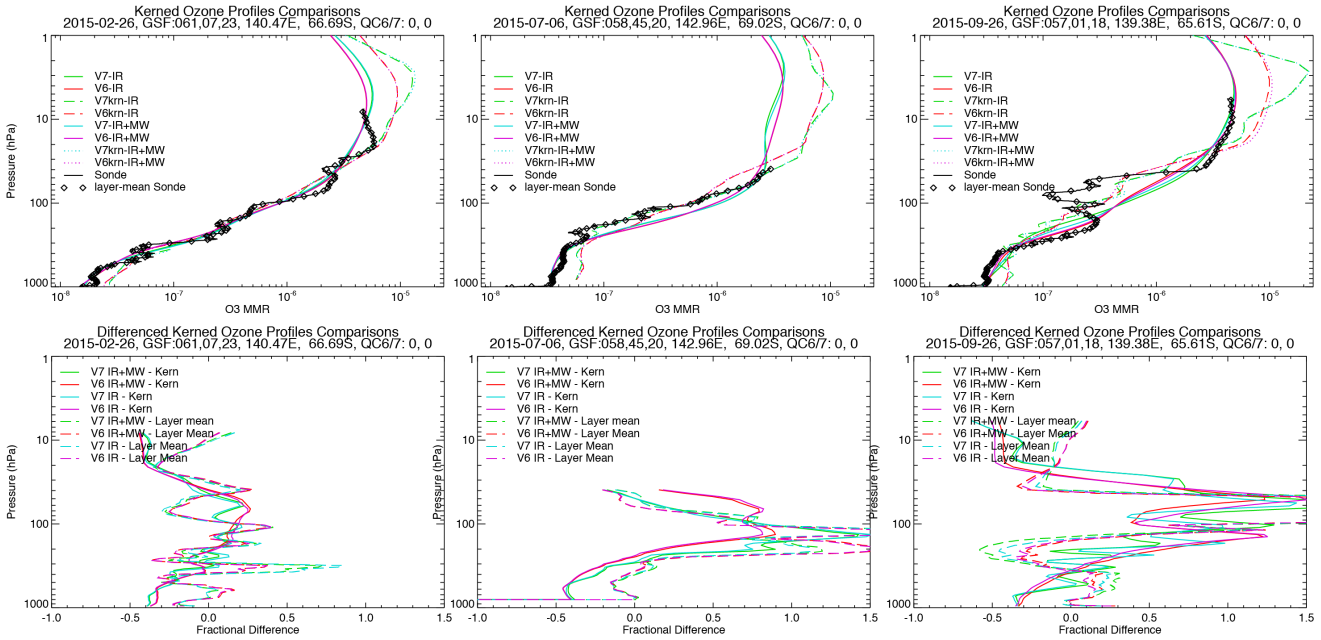


Figure 5.4.12: Comparison of version 6 and 7 ozone profiles for the same footprints in Figure 5.4.11 (left to right, Summer, Winter and Fall). The upper row shows ozonesonde profiles. Black and black diamonds are the ozonesonde and layer-mean ozonesonde profiles. Version 6 IR+MW are red curves, version 6 IR-only are magenta curves, version 7 IR+MW are green curves and version 7 IR-only are cyan curves. Kerned profiles use the same colors, but are shown as dashed curves. The lower row shows differences, solid curves are differenced with kerned ozonesondes; dashed curves are differenced with layer-mean profiles. The line colors are the same as in the upper row.

indicate the least vertical resolution during the spring and we expect the kerned profile to relax closer to the retrieved and first guess.

5.4.3.6 Total O3 comparison methodology

Comparison of time series of total ozone, illustrate the ability of AIRS to detect the O_3 hole, but the validation of averaging kernels using collocated ozonesondes and total O_3 is challenging. The ozonesonde samples part of the profile while the SOAZ spectrometers measures the entire column. Total O_3 for the ozonesonde is:

$$\text{total } O_{3\text{ sonde}} = \text{total } O'_{3\text{ sonde}} + \text{total } O_{3\text{ correction}}, \tag{5.4.7}$$

where:

$$\text{total } O'_{3\text{ sonde}} = \int_{P_{\text{top}}}^{P_{\text{surface}}} X_{O_3\text{ sonde}}(P)\,\frac{dP}{g(z(P))}$$

is the total O_3 contained in the ozonesonde profile, and

$$\text{total } O_{3\ \text{correction}} = \text{total } O_{3\ \text{SAOZ}} - \text{total } O'_{3\ \text{sonde}}$$

is the additional O_3 above the top of the ozonesonde profile. We expect retrieved total O_3 to match the O_3 contained in the kerned profile; two approaches are used here. In the first, the total is calculated over the entire extent of the kerned profile

$$\text{total } O_{3\ \text{sonde kerned}} = \int_0^{P_{\text{surface}}} X_{O_3\ \text{sonde kerned}}(P)\ \frac{dP}{g(z(P))}\ , \tag{5.4.8}$$

and includes contributions from the first guess above the top of the ozonesonde. In the second, the total for the kerned ozonesonde is evaluated to the top of the ozonesonde profile, P_{top}, and the $\text{total } O_{3\ \text{correction}}$ is added:

$$\text{total } O_{3\ \text{sonde kerned corrected}} = \int_{P_{\text{top}}}^{P_{\text{surface}}} X_{O_3\ \text{sonde kerned}}(P)\ \frac{dP}{g(z(P))} + \text{total } O_{3\ \text{correction}} \tag{5.4.9}$$

We expect the $\text{total } O_{3\ \text{sonde kerned}}$ to have closest agree.

5.4.3.7 Total O_3 comparisons

Figure 5.4.13 summarized the comparisons of total O_3. The upper left two panels shows histograms of total O_3 for the DUS SAOZ UV spectrometer, V6, V7, first-guess, kerned and corrected kern; the lower panels shows histograms of differences from SAOZ. The first guesses are unbiased estimates of the ensemble, but the 49 DU of variability in the ensemble is not captured in the difference, which is only reduced to 40 DU; the V6 and V7 first guesses are not statistically different. The V6 and V7 ensembles are biased low by 4 and 12 DU, the V7 has a larger bias, but the variability of the difference is reduced from 33 DU to 32 DU.

The total O_3 from the kerned profiles agree poorly with the retrieved profiles; kerning reduces the agreement by more than 71 DU and 65 DU in V6 and V7; the total O_3 from retrieved profiles are in better agreement than from the kerned. The corrected kerned profiles agree better to the retrieved products, but not as well as the unkerned profiles. Kerning tends to decrease the total O_3, the correction increases it, but because to ozonesonde profiles do not extend fully into the region of AIRS O_3 sensitivity, the truncated averaging kernels provide a poor estimate of the AIRS sensitivity. The conclusion is that the error in kerning profiles which do not extend high enough, (i.e. above 10 hPa), is greater than the errors in total O_3 of AIRS retrieved products.

The right panels in Figure 5.4.13 show time series of total O_3, the upper panel shows the SAOZ time series, the V6 and V7 MW+IR and IR-only series, and the totals from the ozonesondes corrected by the averaging kernels with and without the SAOZ correction. Versions 6 and 7 both track the annual cycle, V7 residuals are slightly worse than V6. The kerned totals have a weaker annual cycle and do not follow the retrieved time series, but the corrected kerned time series do. This indicates that the errors in the retrieved profiles can be explained by state-dependent errors in the retrieved profile. The largest errors occur during the late winter, early fall and are likely arising from a lack of sensitivity to O_3 variability, and that the initial guess is further from the true state. Improved agreement between the corrected kerned ozonesondes and retrievals, indicates some validation of the information content assessment of the averaging kernels, but is limited because the ozonesondes to not reach high enough into the region of AIRS sensitivity.

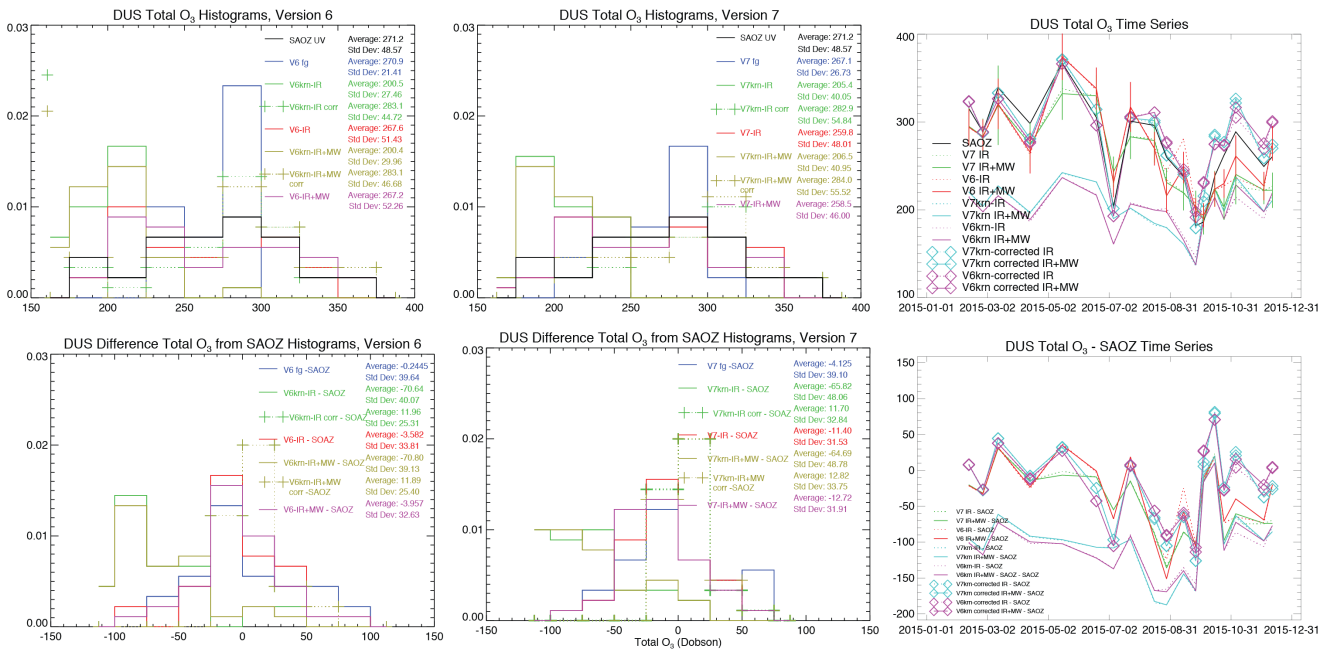


Figure 0: Comparison of version 6 and 7 total O_3. The upper row shows total O_3 from left to right as histograms of version 6 O_3, version 7 and time series. The lower row are histograms and time series differenced from the SAOZ up-looking UV spectrometer. For the histograms, black curves are total O_3 from the SAOZ up-looking UV spectrometer, the blue curves are the V6 and v7 first guess, red curves are the IR-only retrieved total, magenta curves are the IR+MW total O_3, green curves are the kerned ozonesonde and yellow curves are the "corrected" kerned ozonesonde (see text for details). For the kerned profiles, the IR-only curves are marked with "+", but the differences between IR and IR+MW products are small and the curves over-lay. The mean and standard deviation for each histogram is shown on the upper left of each panel. The right column shows the times series of V6 and V7 total O_3 (upper panel) and differenced with SAOZ total O_3 (lower panel). Version 7 are shown as the green curves, V6 by the red curves, and kerned by the cyan (V7) and magenta (V6). Corrected kerned are marked with diamonds. IR+MW are shown as solid lines, while IR-only are shown as dotted lines.

5.4.3.8 Summary of Dumont d'Urville Ozonesonde and Total O_3

Summarizing the key results of this section:

1. Averaging kernels vary with atmospheric conditions. Spring ozone hole conditions have the poorest sensitivity in the lower stratosphere
2. Kerning of ozonesondes tends to reduce differences in the mid and lower stratosphere as compared with layer mean differences.
3. Version 7 shows significant improvements in resolution and agreement with ozonesonde in the lower stratosphere.
4. Total O_3 is derived from the retrieved profiles and although the profiles have improved, total O_3 has not.
5. The change in O_3 climatology has no significant effect on the final result
6. Kerning ozonesonde profiles does not improve comparisons of total O_3 because a significant fraction of the column is above the top of the profile.

5.4.4 AIRS L3 Total Ozone Compared to OMPS

Contributor: Joel Susskind, John Blaisdell, Lena Iredell

AIRS Version 6.55 total ozone is significantly improved compared to AIRS Version 6 total ozone, and is consequently a much more valuable product. AIRS Version 6.55 total ozone now agrees extremely well with total ozone from the operational UV instrument OMPS. Because of this agreement, AIRS total ozone measurements can be used to complement those of OMPS in very important ways. AIRS is an IR instrument and does not require sunlight to observe ozone, as do UV instruments. AIRS produces twice daily fields of total ozone, which better represent the movement of total ozone in a given day. Much more significant is the fact that AIRS produces total ozone products during polar winter, which is a considerable time period not observed by UV instruments. In particular, the ozone hole over Antarctica develops by early August, although UV instruments cannot observe the onset. AIRS Version 6.55 for the first time can accurately depict the onset of the Antarctic ozone hole from IR observations. This is extremely important scientifically.

Data and Methodology

In this section we compare total ozone from AIRS Version 6.55 and AIRS Version 6 AIRS-Only (the current operational system) to the OMPS Version 2.1 Daily Gridded Product. Three selected days during the 2018 ozone hole season are shown in Figures 5.4.14 through 5.4.16. Each of these plots shows three spatially gridded products on the top row; the new AIRS version 6.55, the reference OMPS product, and the currently operational AIRS version 6 AIRS-Only. The bottom row shows the pairwise differences between these three products for grid boxes in common.

Results:

The improvement of Version 6.55 over Version 6 AIRS-Only with respect to OMPS is obvious. Because Version 6.55 AIRS agrees well with OMPS where they have data in common, we have high confidence that the AIRS data in nighttime areas is of equal quality, as the IR algorithm does not depend on daylight. The 2018 season was chosen in part because data from this year can in no way have entered into training for the neural network, compilation of the ozone climatology, or determining any emissivity constraints, as the Version 6.55 system was frozen before the time period containing this data.

The improvements to the ozone products in Version 6.55 fall into two distinct but complementary categories or retrieval and Quality Control (QC) changes. Retrieval changes to the algorithm include: a new climatology which distinguishes ozone hole conditions from normal conditions; adjusting the ozone climatology to the troposphere for this profile; use of additional vertical functions to better represent the profile; additional use of the highest-peaking ozone channels most of the time; and reduced damping. The ozone retrieval also benefits greatly from improved polar stratospheric temperatures in the v6.55 neural network. Quality Control (QC) changes include the development of an entirely separate QC procedure for ozone, recognizing that determination of the important stratospheric part of the ozone profile is not highly dependent on accurate tropospheric temperature which the main QC is designed to assess. The ozone QC in Version 6.55 relies on evaluating radiative closure in the ozone channels, detection of dust (which can contaminate the ozone channels), flagging huge changes in the first ozone iteration (which likely indicate a problem with the temperature retrieval rather than the ozone retrieval),

and spurious adjustments in the emissivity retrieval in the ozone spectral region, which also contains silicate features. The application of the new QC procedures allows for the reduced damping in the retrieval, as we no longer have the runaway retrievals occasionally found in Version 6 which the stronger damping had attempted to control. Some of these improvements are explicitly visible in large scale regions of the Level 3 products selected for these figures. Removal of the occasional outlier values found in Version 6 is not as apparent in the Level 3 images but is also an important improvement to the product.

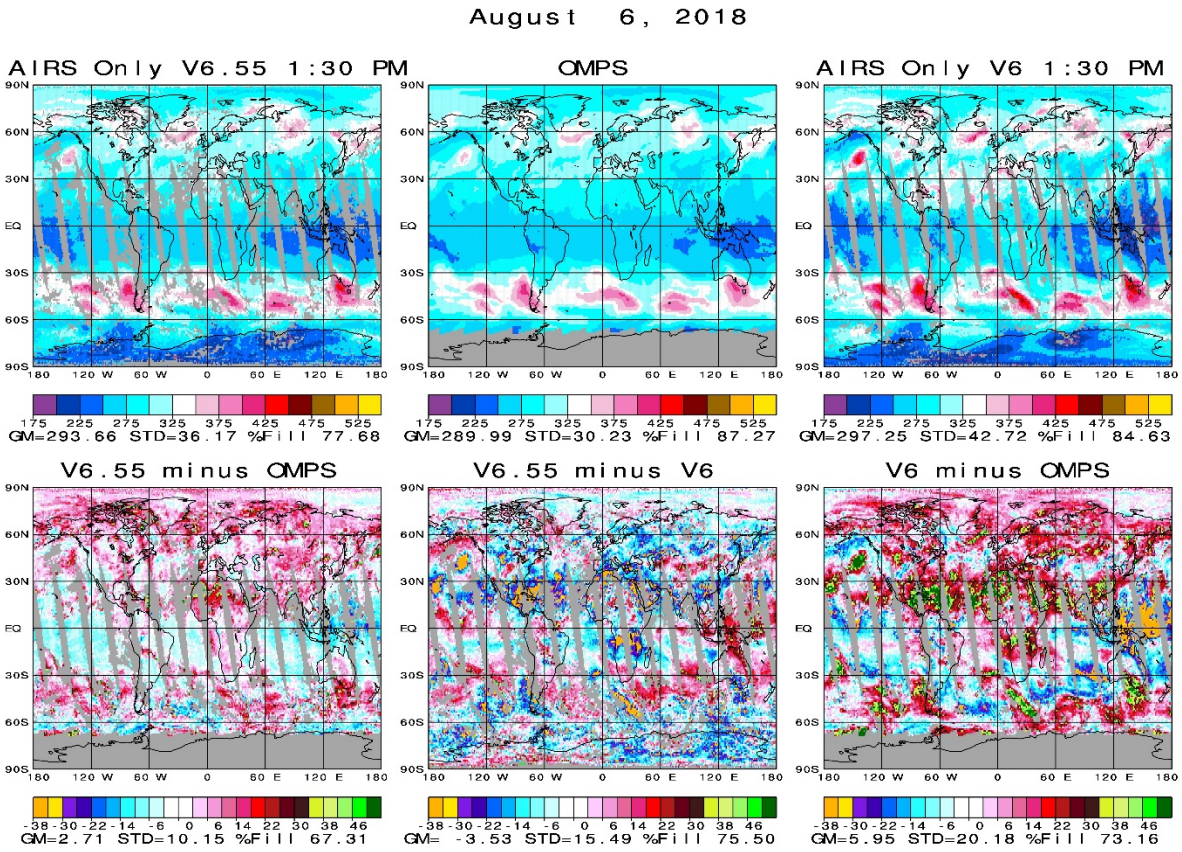


Figure 5.4.14 Daily gridded ozone comparisons for August 6, 2018, early in the formation of the 2018 ozone hole. Note the enormous improvements in the latest version in the low ozone region over the tropical western Pacific Ocean, where the overdamped Version 6 could not reach sufficiently low total ozone values and retrieval improvements have ameliorated this issue. Note also the improvement over northern Africa and the adjacent Atlantic Ocean, where the QC rejection of dusty cases has removed the incorrect ozone values from the Saharan dust outbreak area. These improvements in the tropical areas were necessary for the AIRS total ozone product to be competitive with other data sources. Note also the hint of a forming ozone hole in the OMPS data, with three small areas along the terminator with values below 250 DU. These areas are precisely consistent with the northernmost extent of the low ozone region depicted in the Version 6.55 total ozone product, which for this day is not very different from the Version 6 depiction.

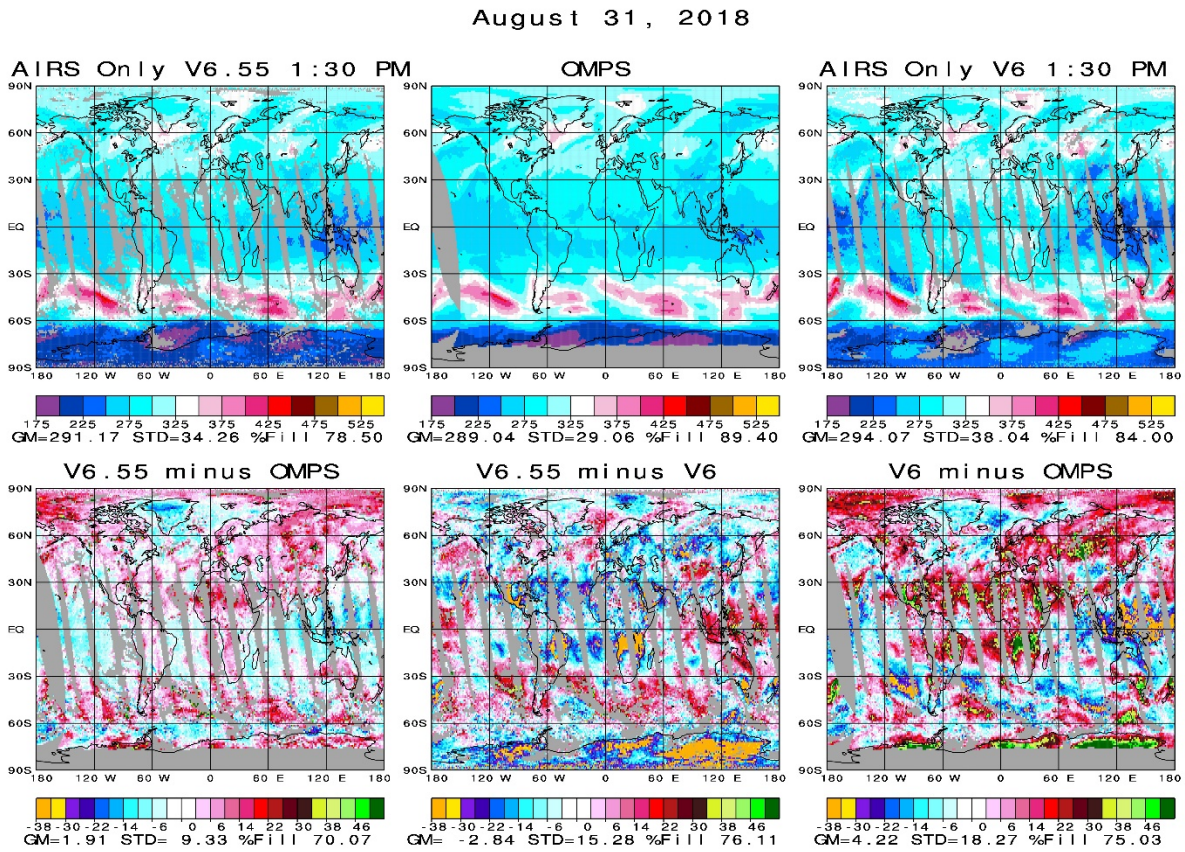


Figure 5.4.15. Daily gridded ozone comparisons for August 31, 2018, somewhat later in the formation of the 2018 ozone hole. On this day we see similar improvement as in Figure 5.4.14 for the tropical areas, but here we have a significant change in the ozone hole depiction. The areas measured by OMPS as below 200 DU are matched in the AIRS data, and we can see that these low values do not extend as far as the South Pole. On this day there are significant differences over Antarctica between Version 6 AIRS-Only and Version 6.55 AIRS. Anomalously high retrievals of total ozone over Antarctica in Version 6 have been corrected by a combination of using the new climatology and excluding the strongest ozone channels from the retrieval in regions where the surface temperature is nearly equal to the stratospheric temperature.

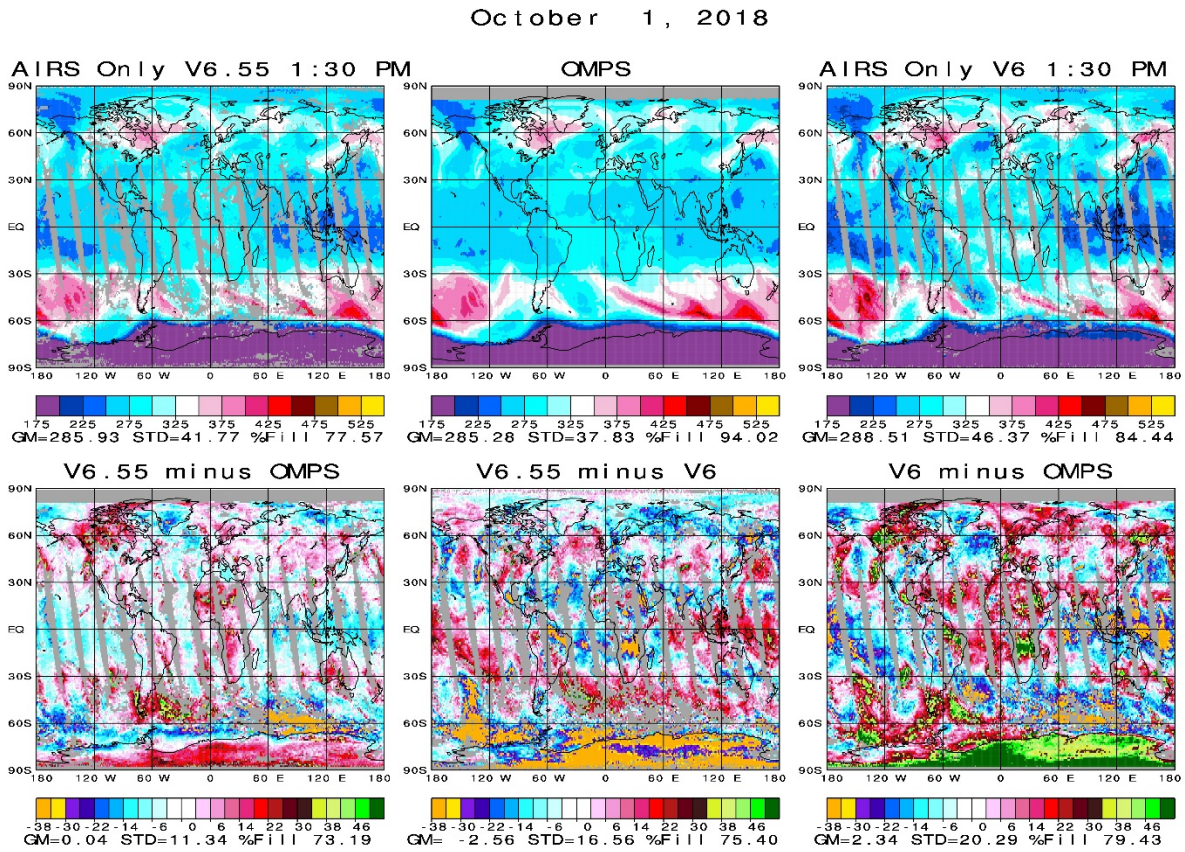


Figure 5.4.16. Daily gridded ozone comparisons for October 1, 2018, near the lowest ozone values and greatest spatial extent of the 2018 ozone hole. The vast majority of the points over Antarctica are now within 20 DU of the OMPS values, even in this extreme case, and examination of the V6.55 minus V6 difference plot shows that the retrieval algorithm updates in Version 6.55 made significant improvements in nearly all parts of the globe. Some smaller discrepancies remain in cloudy regions of the Southern Ocean which are particularly challenging for IR retrievals.

5.5 Analyses on AIRS L2 Cloud-Clear Radiance (CCRad)

Contributor: R. Chris Wilson, and Mathias Schreier

Data and Methodology

This analysis compares cloud-cleared radiances (CCRs) computed with AIRS/AMSU (IR+MW) to radiances computed with only AIRS (IR) for AIRS L2 versions 6 and 655. The CCRs are spectrally convolved to MODIS channels 22, 24, 28, 32, 33, 34, and 35, while clear MODIS radiances are spatially convolved to the AIRS Field of Regard (FOR). These channels are chosen because the convolution procedure between AIRS and MODIS is most accurate here and AIRS doesn't have spectral gaps in this region. All AIRS CCRs for January 14th and 15th, for 2003 and 2015 are compared to clear MODIS pixels. Figure 5.5.1 shows the spectral response functions that convolve AIRS CCRs to MODIS channels.

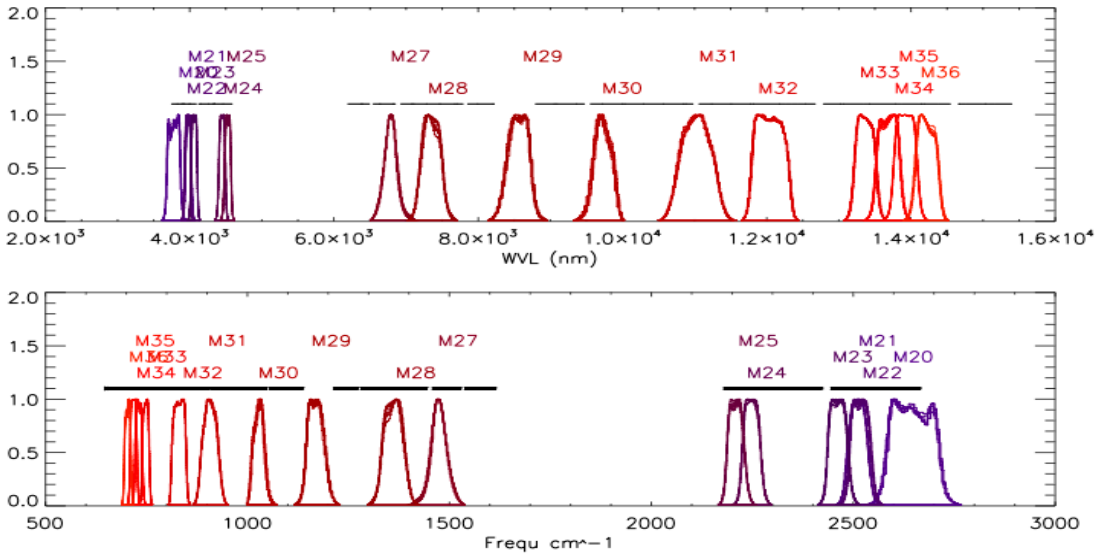


Figure 5.5.1. The spectral response function for MODIS channels.

The method for convolving and comparing clear MODIS radiances with AIRS CCRs is validated by convolving all the MODIS pixels in the AIRS field of regard with spectrally convolved Level 1b AIRS radiances. Without spectral gaps in AIRS, and perfect knowledge of the of the AIRS spatial response function the matchups should have zero difference. Table 5.5.1 shows the bias, standard deviation, and percent of matchups between -0.5 and 0.5K. These MODIS channels were selected because there were not spectral gaps in AIRS and the collocation method generally had near zero bias and nearly 90 % of matchups with less than 0.5K differences. The other dates showed similar results to the results in Table 5.6.

Table 5.5.1. Bias, standard deviation, and percent of matchups between -0.5 and 0.5K for cases in January 14, 2003.

Modis Channel	Wavenumber (cm^{-1})	Main Influence	Bias (K)	Std (K)	% diff between -.5K to .5K (%)
21	2525	Sfc/cld temp	-0.18	1.29	83
22	2465	Sfc/cld temp	-0.09	1.06	92
24	2236	CO_2/Atm temp	-0.46	0.93	99
28	1363	H_2O	-0.33	0.56	98
32	831	Sfc/cld temp	0.07	0.75	89
33	749	CO_2	-0.02	0.55	95
34	731	CO_2	-0.05	0.47	97
35	719	CO_2	-0.60	0.47	99

Results:

We compare all observations where at least 5 percent of the FOR is clear as determined by MODIS. Table 5.5.2 shows the differences for each MODIS channel when AIRS clear scenes are spectrally convolved to that respective MODIS channel, and all clear MODIS are spatially weighted by the AIRS spatial response function. This is for the version 655 L2 code processed for January 14, 2003. The IR+MW generally has a lower standard deviation then the IR only

cases. Differences between the IR+MW and IR only are similar to the accuracy of the collocation method. We found that the other 3 dates generally had similar values and showed little difference between IR+MW and IR only.

Table 5.5.2. Differences between the AIRS V655 L2 CCRs for each MODIS channel for cases in January 14, 2003. Both IR+MW and IR only results are shown.

Channel	Bias (K) IR+MW	Bias IR only	STD IR+MW	STD IR only
22	0.84	0.81	2.92	2.96
24	-0.04	-0.04	0.67	0.67
28	0.10	0.11	1.26	1.30
32	0.88	0.85	2.62	2.69
33	0.74	0.74	1.63	1.67
34	0.53	0.53	1.17	1.19
35	-0.34	-0.34	0.91	0.93

Table 5.5.3. Same with Table 5.5.2 but for AIRS V6 L2 CCRs.

Channel	Bias (K) IR+MW	Bias IR only	STD IR+MW	STD IR only
22	0.88 (5)	0.85 (4)	3.03 (4)	3.06 (3)
24	-0.03 (-52)	-0.03 (-34)	0.68 (2)	0.68 (1)
28	0.14 (29)	0.13 (13)	1.31 (4)	1.33 (2)
32	0.98 (10)	0.95 (11)	2.72 (3)	2.76 (3)
33	0.80 (8)	0.78 (6)	1.69 (4)	1.71 (2)
34	0.56 (7)	0.53 (4)	1.21 (3)	1.21 (2)
35	-0.32 (-8)	-0.34 (-4)	0.94 (3)	0.94 (2)

We examine the Version 6 processing algorithm in Table 5.5.3 with the same observation criteria as Table 5.5.2. For this version the IR only bias is generally smaller than the IR+MW, but has a larger standard deviation. The values in parenthesis is the percent increase (positive) / decrease (negative) compared to the similar calculation for Version 655. All Channels except 24 and 25 showed an increase in bias and standard deviation for Version 6 compared to Version 655. Other dates showed a similar relationship in the bias, but the standard deviation had no real difference between the IR and IR+MW.

The AIRS CCR product contains Quality Control (QC) flags for all channels. QC=0 means the expected error is less then 1K, QC=1 means that the expected error is between 1 and 2.5K, and QC=2 means the expected error is greater than 2.5K. The expected error is based on a regression between the internal retrieval indicators, and agreement with ECMWF profiles. This study used a modified QC to assess the spectrally collocated AIRS CCRs. The expected error for the collocated AIRS is the average of the expected error for all AIRS channels located in the MODIS spectral response function. Each AIRS and MODIS channel matchup is then assigned a QC value equal to 0, 1, or 2 based on the above QC description. In Figure 5.5.2 and Figure 5.5.3 we compare the brightness temperature difference between convolved AIRS CCRs and clear MODIS spectra as a function of QC using data on two different days: January 15, 2003 and 2015. Comparing V655 to V6 (left to right) we see that V655 has a slightly lower error for QC equal to two, and QC equal to one for channels 34 and 35. Both processing versions (V6 and

V655), and both instrumental setups (IR+MW and IR only) demonstrate that the QC correctly labels QC equal to 2 and 0, but some cases in QC equal to one should have been labeled zero.

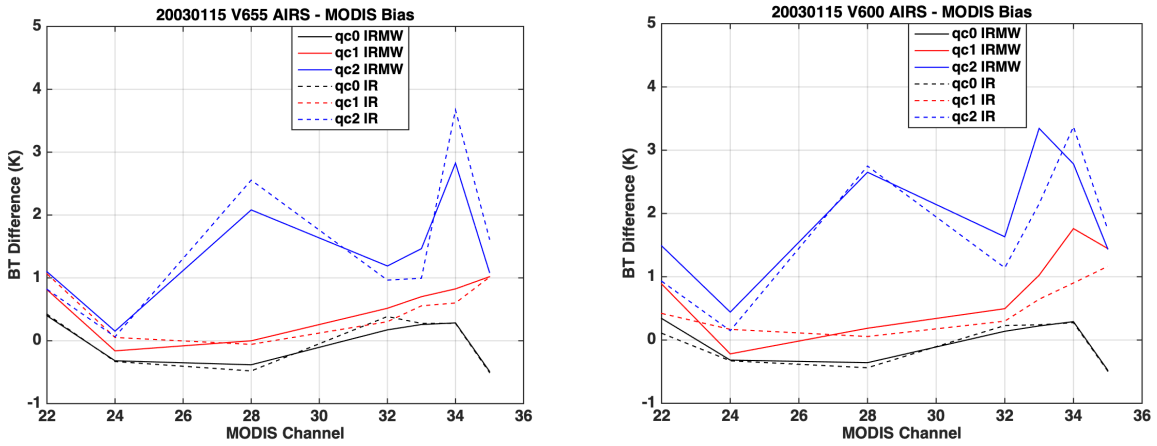


Figure 5.5.2. Difference between AIRS CCRs and MODIS channel bright ness temeperatures for cases on January 15, 2003 for AIRS V655 (left) and AIRS V6 (right) with different QC flags.

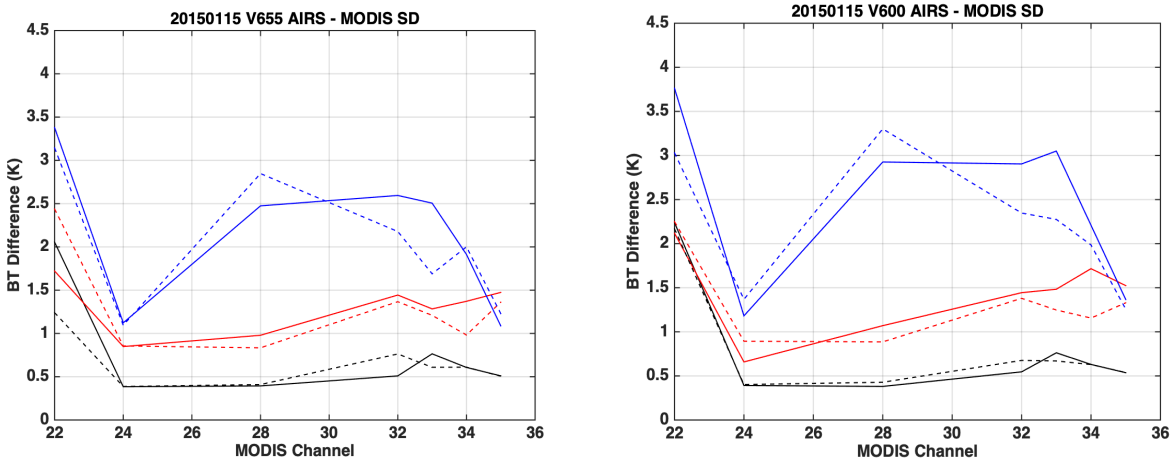


Figure 5.5.3. Same with Figure 5.5.2 but for cases on January 15, 2015.

Figure 5.5.4 displays the percent of cases that are QC level 0, 1, and 2 for the AIRS CCRs spectrally convolved to MODIS channel 22. This channel is sensitive to the surface or cloud top temperature. We chose to display January 3, 2015 for V655 (Figure 4a) and V6 (Figure 4b), but the rest of the dates had similar QC relationships. V655 IR+MW has less cases with QC equal to 0 and more cases with QC equal to 2 compared with the V6 IR+MW. The V6 IR loses almost all of its QC equal to zero cases, while the V655 IR only preserves a much higher percentage of QC equal to zero cases.

Figure 5.5.5 is similar to Figure 5.5.4, but for AIRS CCRs spectrally convolved to MODIS channel 28. MODIS Channel 28 is sensitive to water vapor. Both Versions (V655 and V6) and instrument setups (IR+MW and IR) have almost no QC equal to two. Like Figure 5.5.4, V655 IR+MW has less QC equal to zero compared with V6. In this case V6 IR doesn't lose as much QC equal to zero compared to AIRS CCRs convolved to MODIS channel 22. In fact, V6 IR has more QC equal to zero than V655.

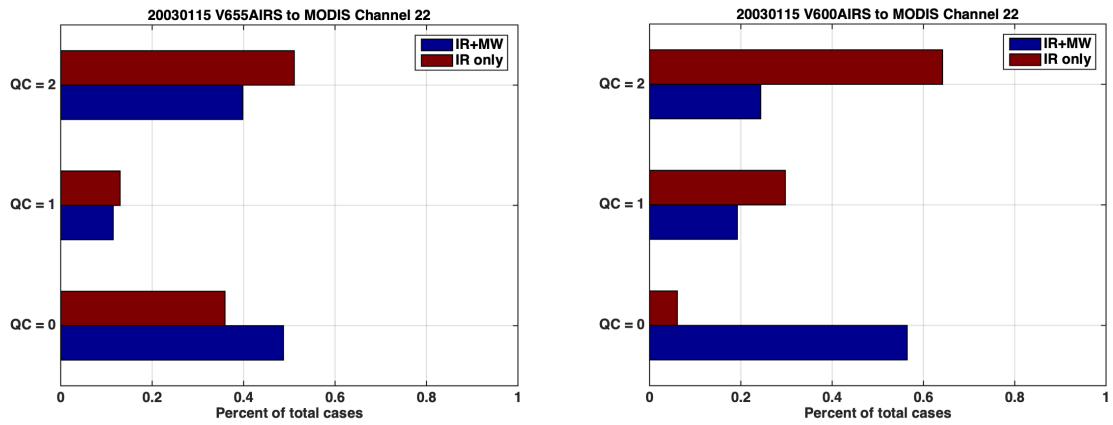


Figure 5.5.4. The percent of cases that are QC level 0, 1, and 2 for the AIRS CCRs spectrally convolved to MODIS channel 22. The left panel shows results for AIRS V6.55 and the right panel for AIRS V6.

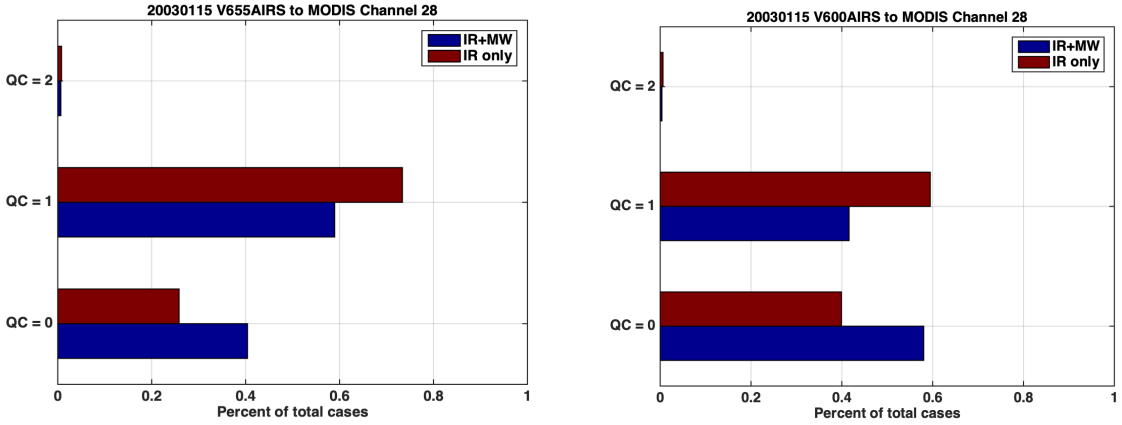


Figure 5.5.5. Same with Figure 5.5.4 but showing results with respect to MODIS Channel 28.

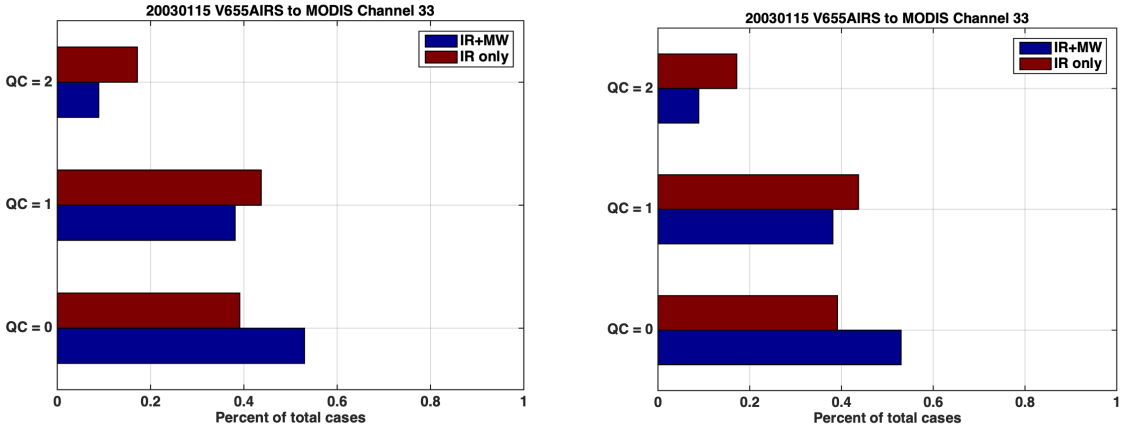


Figure 5.5.6. Same with Figure 5.5.4 but showing results with respect to MODIS Channel 33.

Figure 5.5.6 is similar to Figures 5.5.5 and 5.5.4, but for AIRS CCRs spectrally convolved to MODIS channel 33. This region of the spectra is sensitive to CO_2. Like Figure 5.5.5, V655 has less QC equal to 0 for IR+MW and IR compared to V6. Unlike in Figure 5.5.5 though, V655 has

less QC equal to 1 for IR and nearly the same QC equal to 1 compared to V6. There is more QC equal to 2 for V655 compared to V6 in this case.

We have shown two separate methods for estimating the error in AIRS CCRs. The first methods compare CCRs from AIRS to clear MODIS radiances in the same FOR. The second method uses the QC flag to get an estimated error. Table 5.5.4 examines how well the expected error from the QC is incorrect as determined by comparing the CCRs to clear MODIS radiances for Version655. This table shows that comparing the AIRS CCRs against MODIS can identify cases when QC equal to zero has an error that is too high. This is most evident in Channels 33:35. Also this comparison identifies some cases when the QC identifies cases with too much error. Generally the IR+MW is slightly better at estimating the correct QC for V655. Table 5.5.5 is similar to Table 5.5.4, but is for V6 instead of V655. Generally, the values for V6 are similar to V655, with some values higher and some values smaller.

Table 5.5.4. AIRS V6.55 CCRs vs. MODIS Clear V655 for January 15, 2015. Frac 0 is the percentage of cases with QC equal to zero that had an error greater than 1K. Frac 1 is the is the percentage of cases with QC equal to one that had an error greater then 2.5K. Frac 2_1 is the percentage of cases with QC equal to two that had an error less then 2.5K. The first value is IR+MW and the value in parenthesis is IR.

Channel	Frac 0 (%)	Frac 1(%)	Frac 2_1 (%)
22	13(13)	10(12)	17(14)
24	4(4)	1(1)	32(38)
28	2(2)	2(2)	14(11)
32	5(13)	6(6)	15(11)
33	57(72)	12(10)	26(14)
34	11(13)	10(4)	18(22)
35	12(12)	10(7)	48(65)

Table 5.5.5 Same with Table 5.5.4 but for AIRS V6.

Channel	Frac 0 (%)	Frac 1(%)	Frac 2_1 (%)
22	14(13)	12(7)	18(13)
24	4(5)	<1(2)	28(29)
28	2(3)	2(2)	16(12)
32	6(8)	6(6)	16(13)
33	52(67)	16(11)	28(26)
34	12(13)	22(7)	19(19)
35	13(14)	11(6)	40(44)

Conclusion and Summary:

To summarize, the differences of the CCRs from the IR+MW and the IR were less than the accuracy of the collocation scheme for QC equal to 0 and QC equal to one. We generally assume then that the IR+MW and IR are essentially similar to the accuracy of this validation scheme for both V655 and V6, although V655 was a little more accurate. The QC methods are accurate for QC equal to 0,1, and 2 for channels 22:28. QC equal to 1 had slightly more error for channels 33-35 for V6 compared to V655. V655 generally identifies less QC equal to 0 cases

and more QC equal to 2 cases compared to V6. The current QC processing system misidentifies a significant portion of QC equal to 2 cases for both V655 and V6. This was especially the case in the region of MODIS channels 33-35. V655 does a better job preserving the QC equal to 0 cases when using on the IR, compared to V6 which loses more QC equal to 0 cases for some channels.

5.6 Analyses on AIRS L3 Outgoing Longwave Radiation (OLR)

Contributor: Joel Susskind, John Blaisdell, and Lena Iredell

Data and Methdology:

This section compares 2018 calendar year Version 6 and Version 6.55 AIRS and CERES monthly mean products for OLR and Clear Sky OLR. The prototype Version 6.55 product, which is an AIRS-Only system with a few AMSU A1 channels in the neural network startup, is compared against the operational Version 6 AIRS-Only product to evaluate the effects of code and parameter updates between Version 6 and Version 6.55. More importantly, these AIRS products are compared against the reference Energy Balanced and Filled (EBAF_Ed4.0) product from the Clouds and the Earth's Radiant Energy System (CERES). Version 6.55 OLR products are very similar to Version 6 OLR products, which already showed excellent agreement with CERES.

AIRS and CERES produce their OLR products by completely different methodologies. CERES measures two broadband radiances and applies corrections and assumptions to estimate a time and angle averaged OLR. AIRS retrieves a physical state from each of its twice daily observations, and then calculates Outgoing Longwave Radiation based on that retrieved state, with use of an OLR Radiative Transfer Algorithm (RTA). AIRS generates single field of view values of Level 2 OLR under all conditions in which the cloud parameter retrieval converged, which is about 99% of the time, and includes these Level 2 OLR products in the gridded Level 3 product.

The Version 6 Performance and Test Report (Dang, et al., 2017) showed that over many years, the AIRS and CERES EBAF_Ed2.6 products showed very similar trends and anomaly features. This report instead focuses on results on the one complete year of data for which we have run Version 6.55, with a goal of looking for any seasonal patterns that might be found in the data.

Results:

Figure 5.6.1 shows the difference of the monthly area weighted global mean Level 3 Version 6.55 AIRS-Only product from the reference CERES dataset, as well as the difference of the Version 6 AIRS-Only product from the same reference dataset. The biases between the datasets are slightly reduced in Version 6.55 from that of Version 6. Version 6.55 also has a more consistent month-to-month bias, which supports the idea of a slight improvement in the Version 6.55 product over Version 6 AIRS-Only. The bias between the products is just outside the estimated uncertainty of 2.5 W/m^2 of the reference CERES product (Loeb et al, 2017).

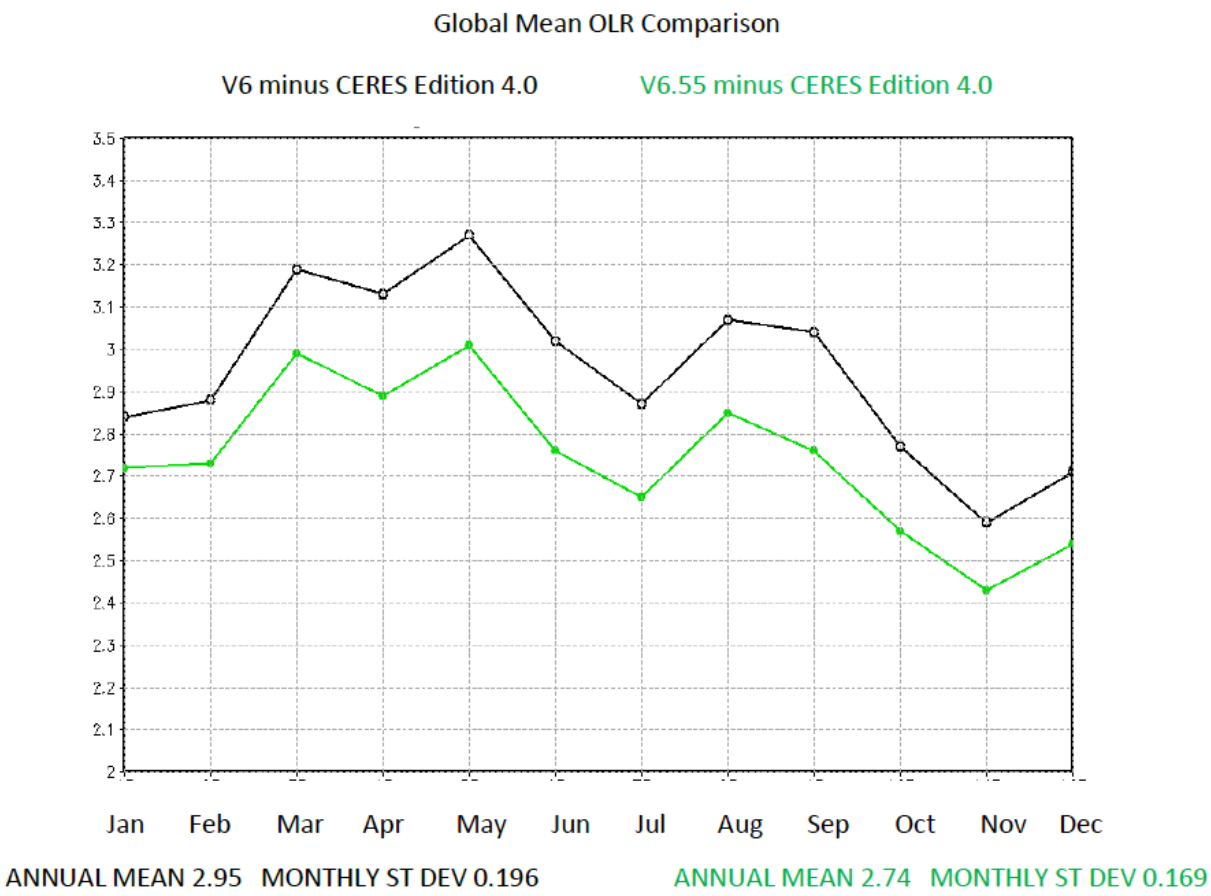


Figure 5.6.1. Differences between AIRS OLR and CERES OLR for every month of 2018, in W/m^2.

Figure 5.6.2 shows the values of CERES OLR for four representative months, January, April, July, and October 2018. Figures 5.6.3 through 5.6.6 show spatial comparisons of monthly AIRS Version 6 AIRS-Only and Version 6.55 Level 3 OLR products with CERES for the four representative months. In each of Figures 5.6.3 through 5.6.6, the top panel shows the comparison of the monthly product from the Version 6.55 system with the EBAF Edition 4 monthly CERES product. The second panel shows the comparison of the previously released Version 6 AIRS-Only product with the Edition 4 CERES product, and the third panel shows the difference between the first two. We note that the differences among all systems are small for all displayed months, with the Version 6.55 system global mean values very slightly closer to the CERES global mean values. Significant differences exist in areas near the edges of sea ice, where Version 6.55 has improved surface class determination using auxiliary ice cover information from the GFS forecast to mitigate the loss of AMSU-A2. The agreement over ocean is seen to be excellent in all cases. The largest biases are over hot land, where diurnal differences of surface skin temperature, as well as of OLR, are largest. AIRS 1:30 PM and 1:30 AM averaged values of OLR are not entirely representative of the variability over the entire day.

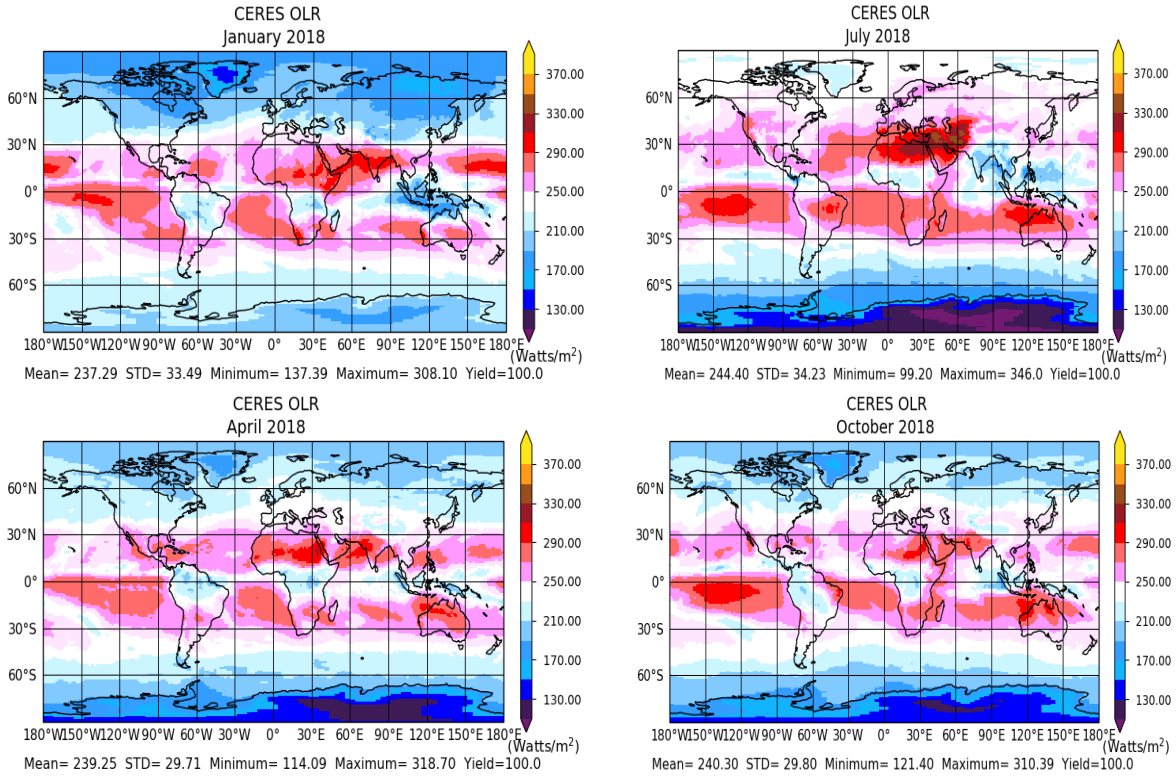


Figure 5.6.2. Seasonal variation of OLR in the CERES reference data set.

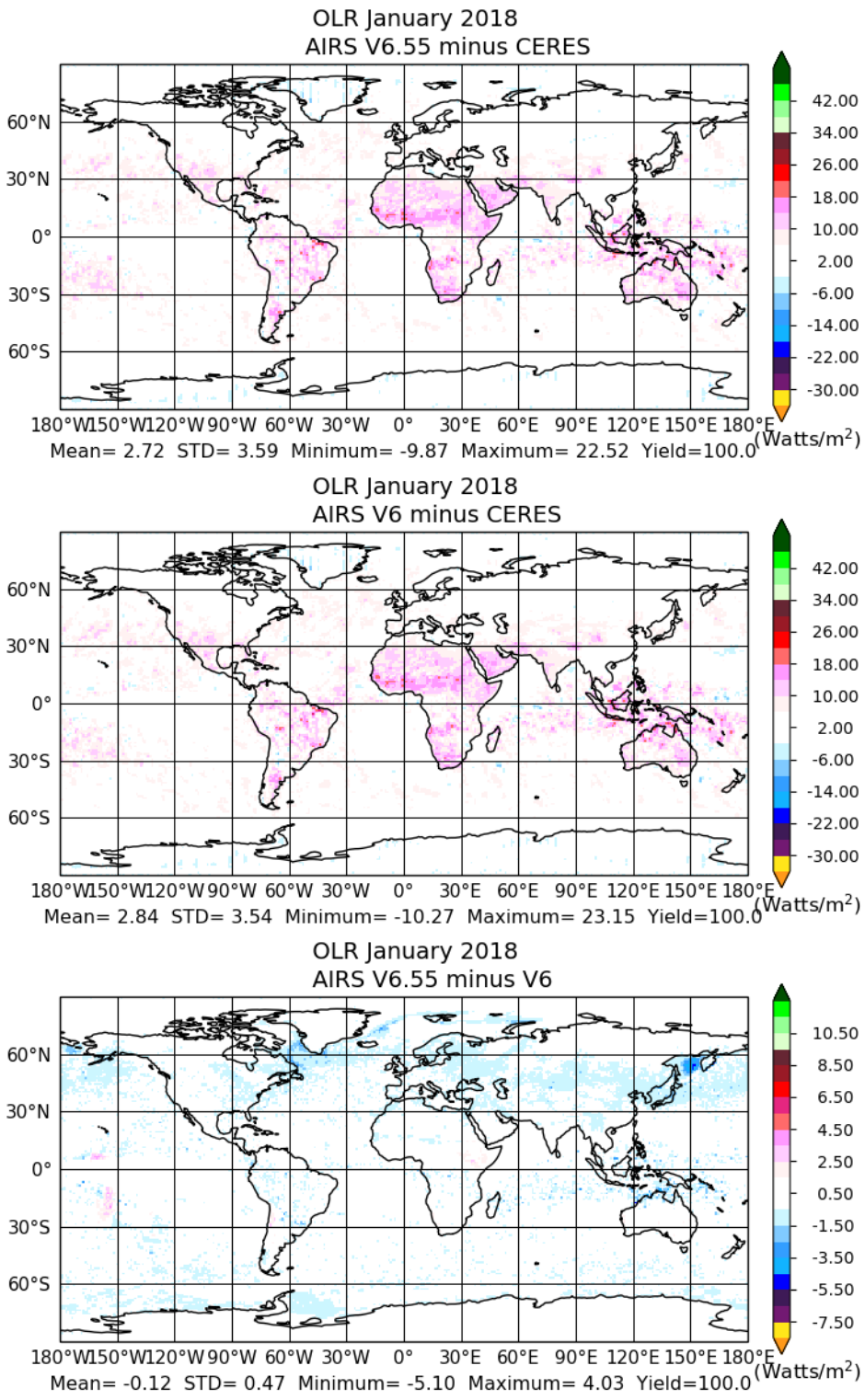


Figure 5.6.3. The OLR differences from CERES EBAF_Ed4.0 for the month of January 2018 are small, and very slightly smaller in Version 6.55 than in Version 6.

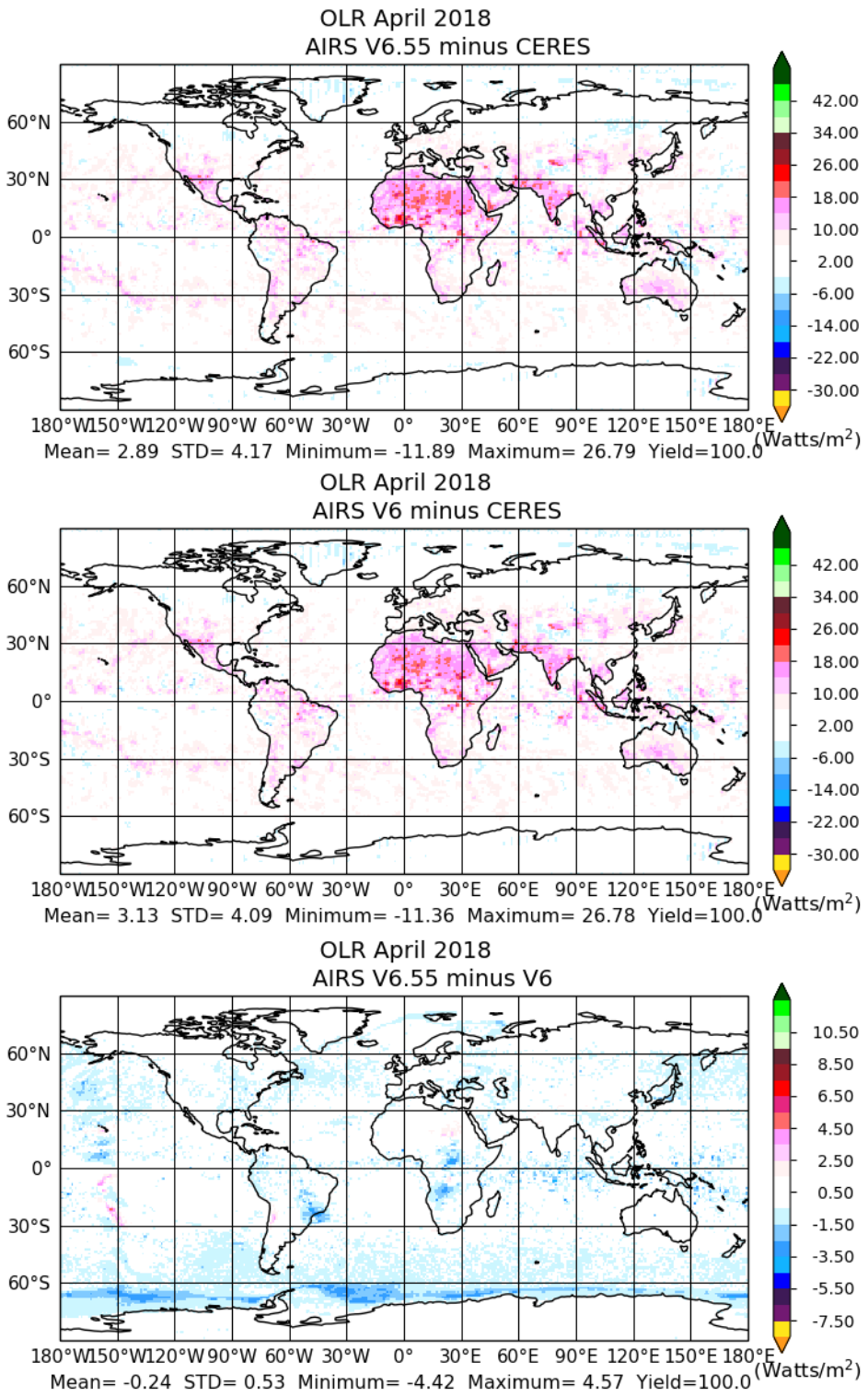


Figure 5.6.4. The OLR differences from CERES EBAF_Ed4.0 for the month of April 2018 are very slightly smaller in Version 6.55 than in Version 6, with the largest differences along the Antarctic ice edge where the surface class determination has been improved.

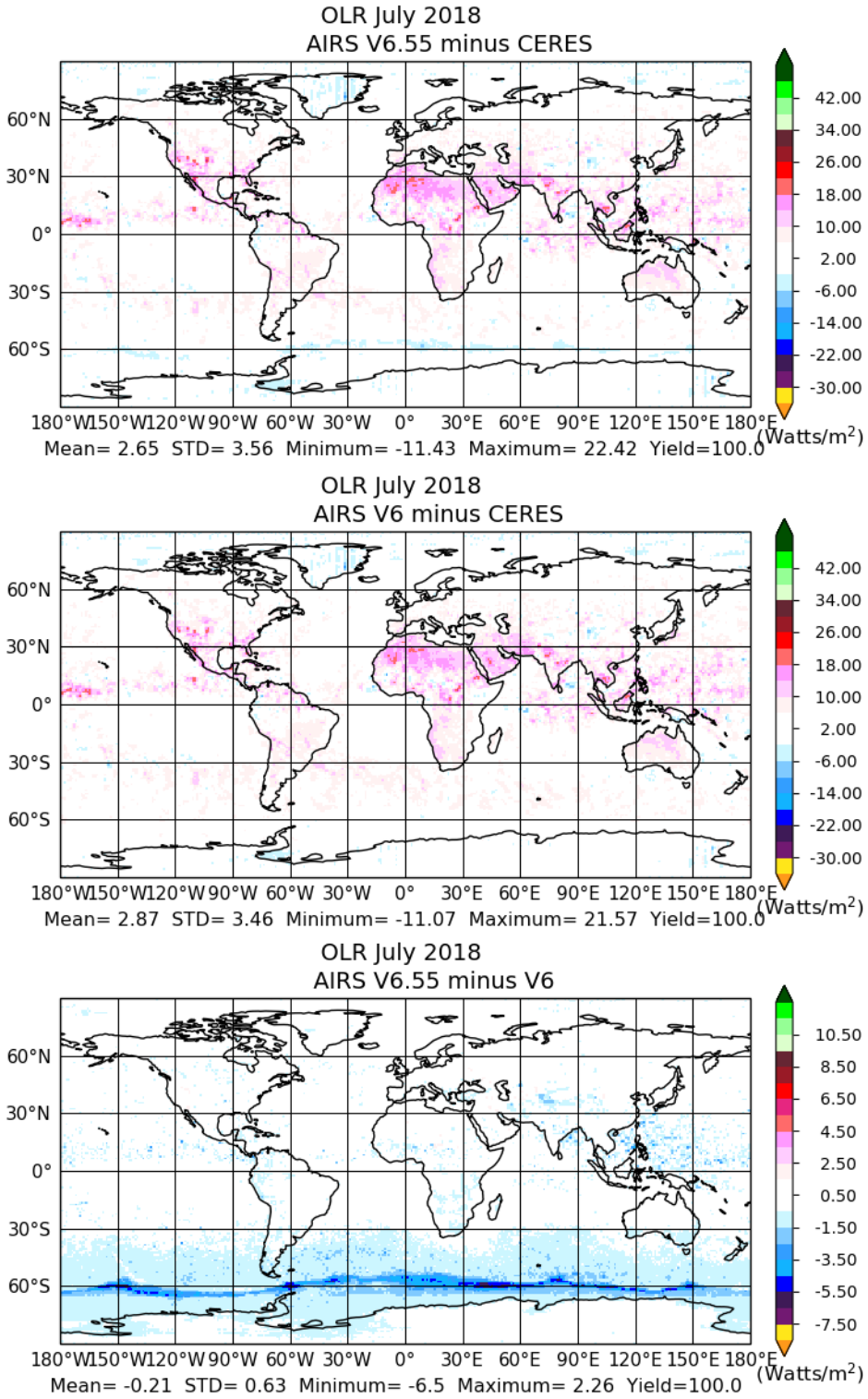


Figure 5.6.5. The OLR differences from CERES EBAF_Ed4.0 for the month of July 2018 are very slightly smaller in Version 6.55 than in Version 6, with the only significant changes occurring along the Antarctic ice edge.

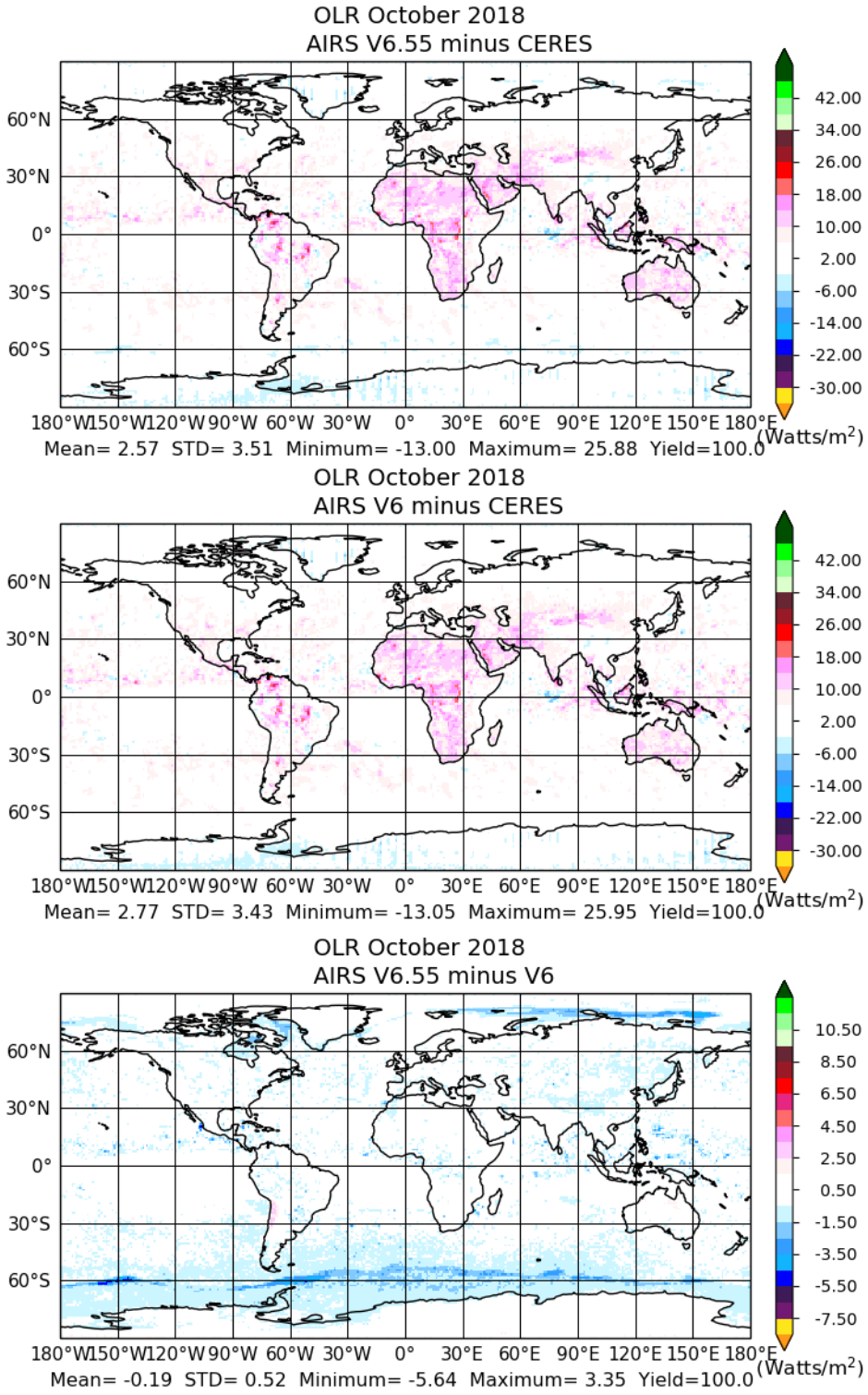


Figure 5.6.6. The OLR differences from CERES EBAF_Ed4.0 for the month of October 2018 are very slightly smaller in Version 6.55 than in Version 6, with the largest differences occurring along ice edges.

Figure 5.6.7 shows the monthly area weighted global mean differences between the AIRS and CERES Clear Sky OLR products. While interesting, comparison between AIRS and CERES monthly mean Clear Sky OLR products are less meaningful than comparisons between the OLR products because AIRS and CERES use different methodologies to decide which cases are to be included in the gridded Level 3 Clear Sky OLR product. The Version 6.55 system shows a more negative bias here than the Version 6 system, but the bias remains fairly constant through the year.

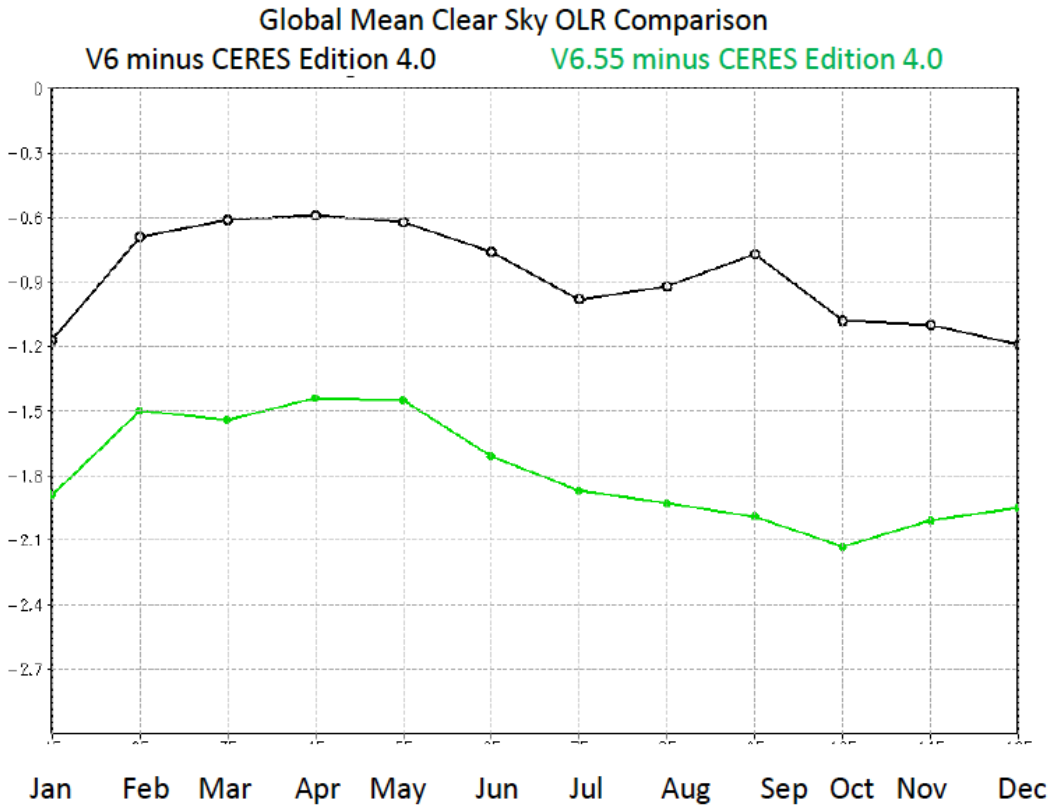


Figure 5.6.7. Difference of AIRS Clear Sky OLR from CERES Clear Sky OLR for every month of 2018, in W/m2.

Figure 5.6.8 shows the values of Clear Sky OLR for the four representative months in the CERES data set corresponding to the time frame. Figures 5.6.9 – 5.6.12 show the differences in Clear Sky OLR products between Version 6.55 and Version 6 from the CERES monthly data and from each other.

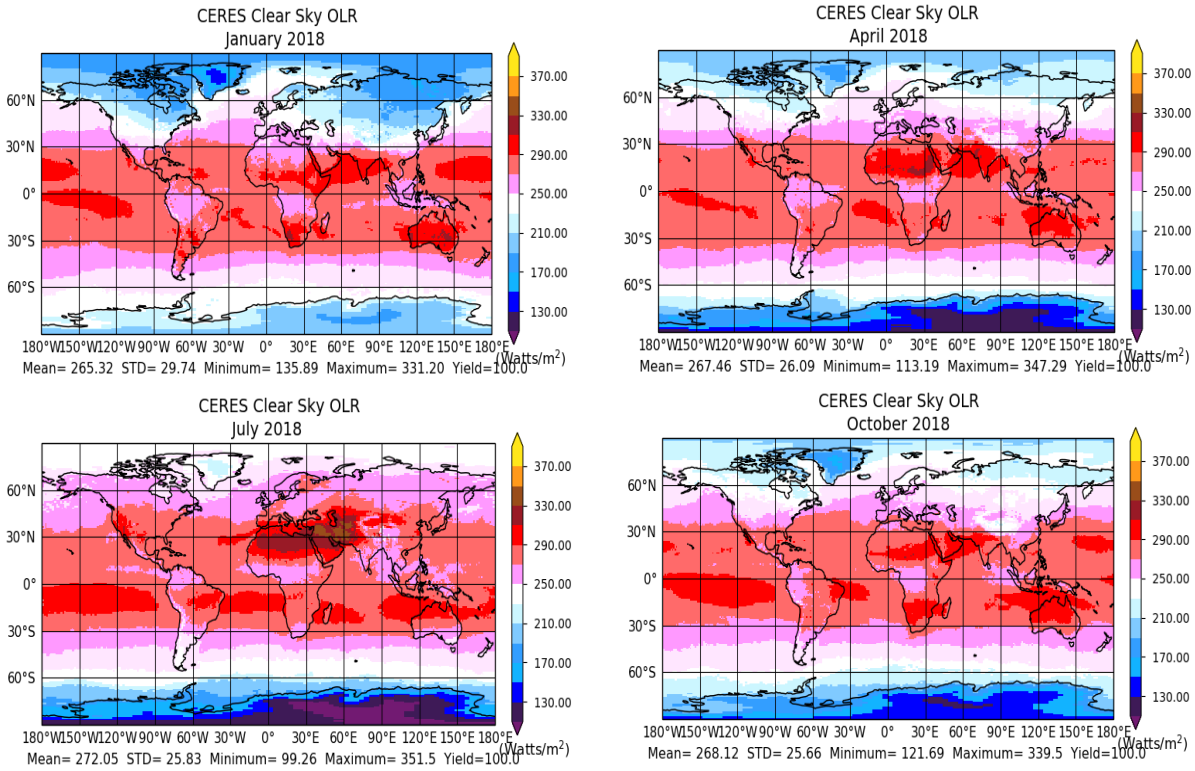


Figure 5.6.8. Seasonal variation of Clear Sky OLR in the CERES reference data set.

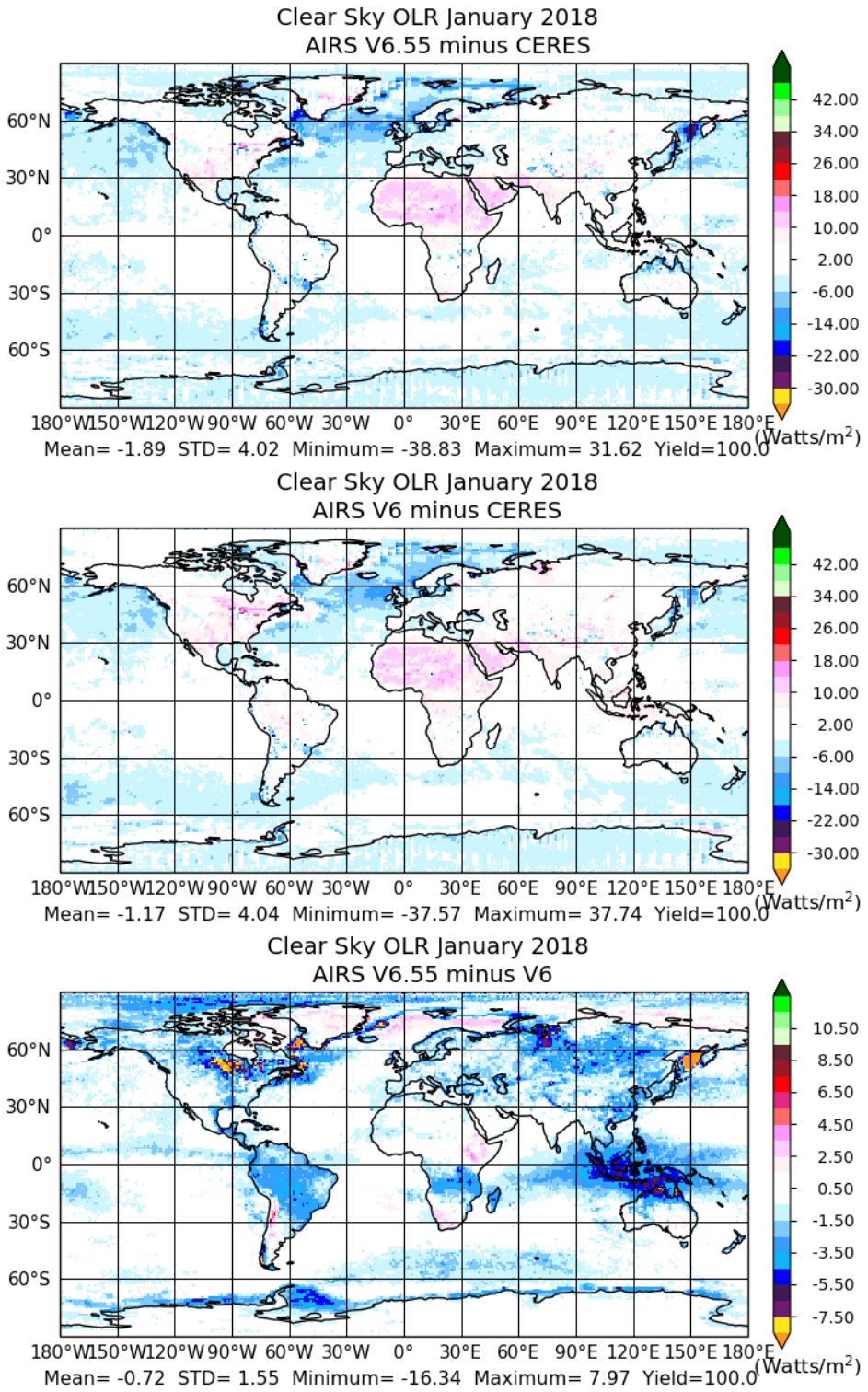


Figure 5.6.9. Clear Sky OLR for January 2018. Changes in retrieval methodology and QC for AIRS have resulted in reduced global mean Clear Sky OLR values compared to Version 6, primarily in convective areas and along ice edges.

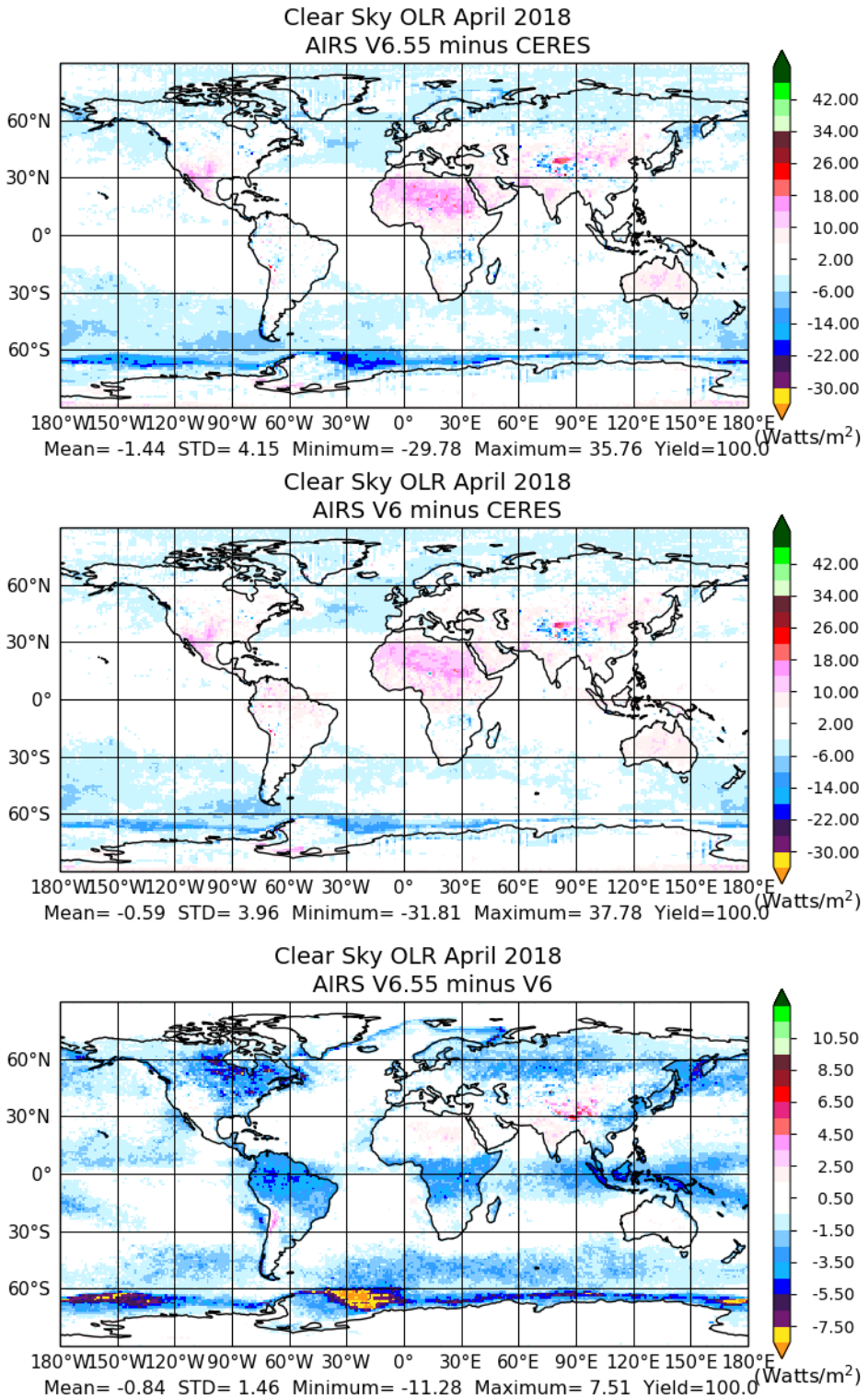


Figure 5.6.10. Clear Sky OLR for April 2018. Changes in retrieval methodology and QC for AIRS have reduced the global mean Clear Sky OLR values compared to Version 6, with the largest changes in convective regions and off the coast of Antarctica.

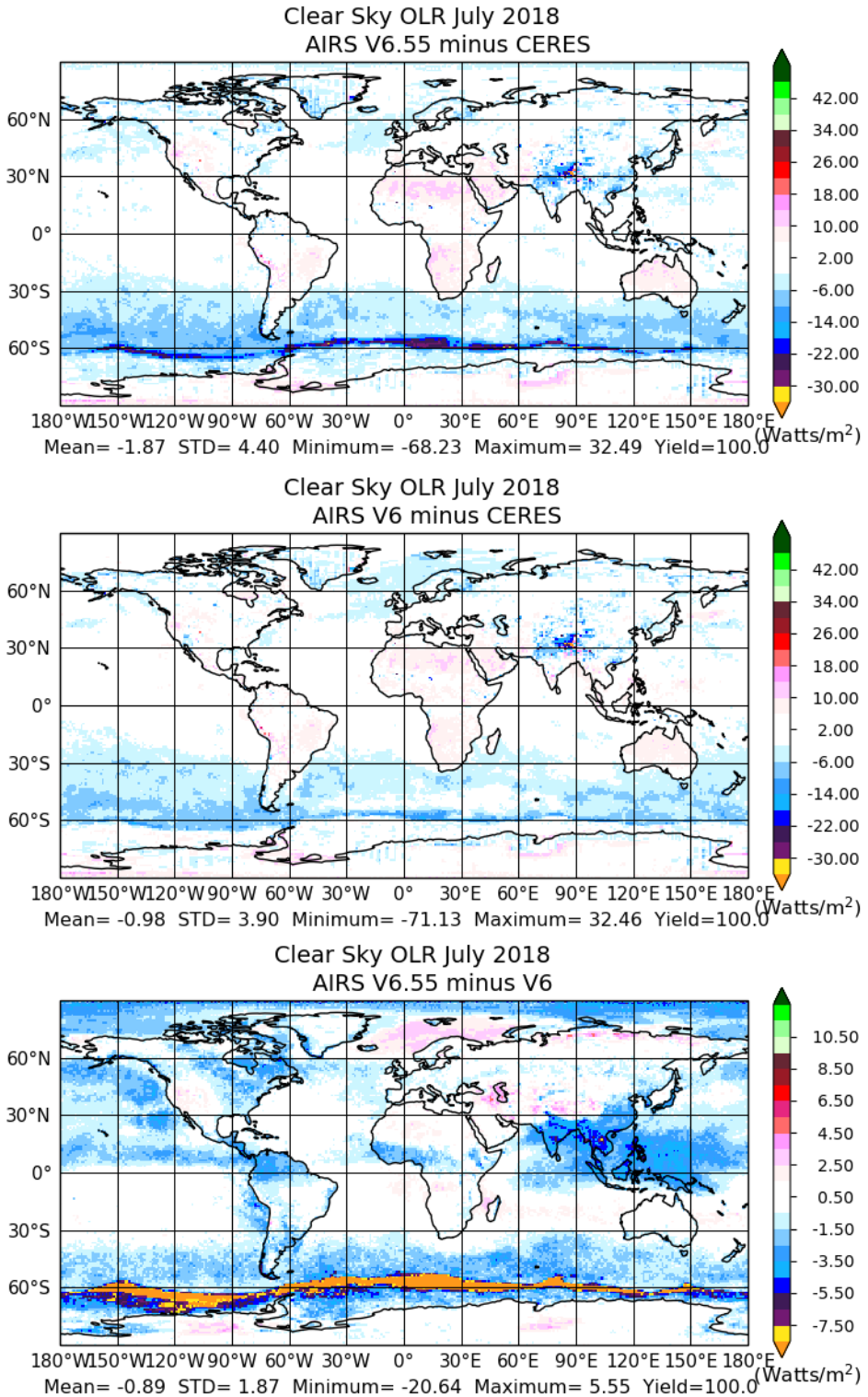


Figure 5.6.11. Clear Sky OLR for July 2018. Changes in retrieval methodology and QC for AIRS have slightly reduced the global mean Clear Sky OLR values compared to Version 6. The largest differences from Version 6 are in convective regions and off the coast of Antarctica, but there are significant differences in the northern mid-latitude and polar regions as well.

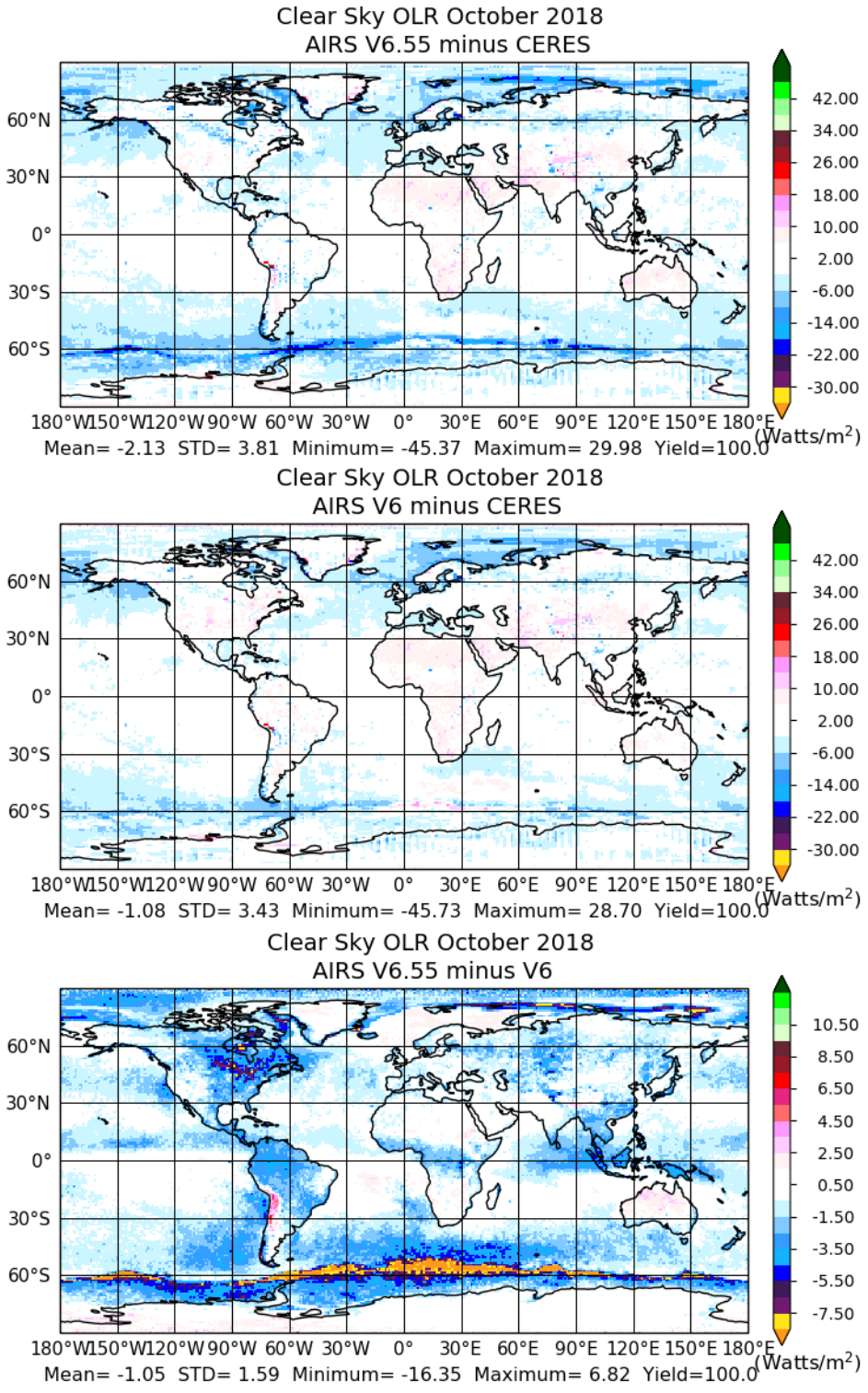


Figure 5.6.12. Clear Sky OLR for October 2018. Spatial patterns of differences between Version 6.55 and Version 6 are similar to those in July.

Longwave Cloud Radiative Forcing (LWCRF) is defined as the difference between OLR and Clear Sky OLR, and represents the effects of clouds on OLR. Figure 5.6.13 shows the monthly mean differences between LWCRF computed from the AIRS and CERES OLR and Clear Sky OLR products. As with Clear Sky OLR, LWCRF is affected by the fact that AIRS and CERES use different methodologies to decide which cases are included in their gridded Level 3 Clear Sky OLR product.

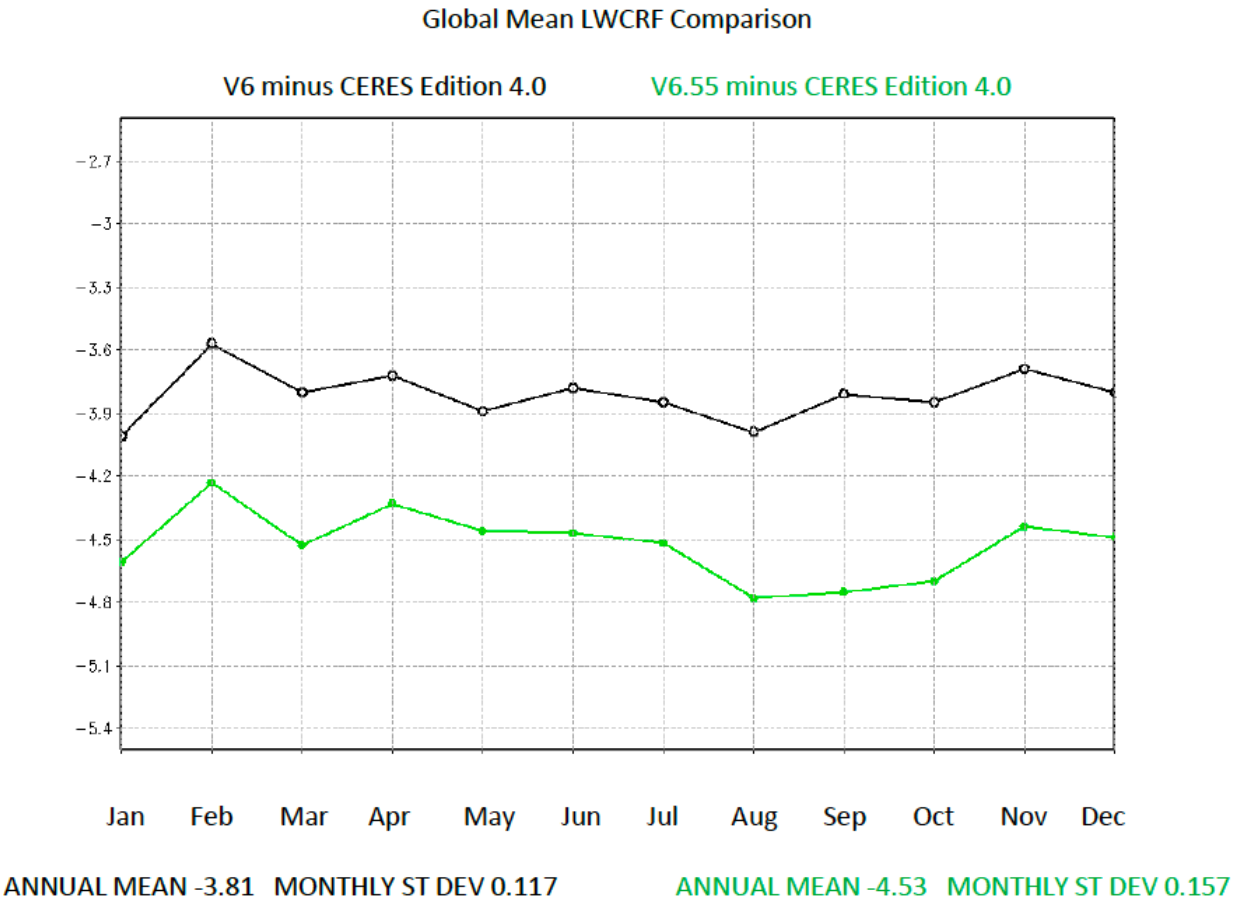


Figure 5.6.13. Long-wave Cloud Radiative Forcing obtained by differencing OLR and Clear Sky OLR.

Table 5.6.1 shows the frequency domains of each of the 16 spectral bands used in computation of OLR for the month of January 2018. This table includes the global mean annual mean spectral band fluxes (W/m^2) and the percentage contribution of each spectral band to OLR, and also includes the spectral contribution of each band to Clear Sky OLR and LWCRF. Almost half of OLR comes from spectral bands 1-3, extending from 100 cm^{-1} to 630 cm^{-1}, an extensive spectral domain in which AIRS does not make any observations. The largest approximation made in the current methodology used to compute AIRS OLR is that the cloud spectral emissivity is frequency independent. This assumption most likely does not hold at the longer wavelengths in spectral bands 1-3, especially in the presence of cirrus clouds. The sixth column in Table 5.6.1 shows that the presence of clouds in a field of view in the longwave spectral domain not observed by AIRS decreases total flux by about 5.0 W/m^2 on the average. Consequently, computation of OLR does not depend significantly on the assumed values of cloud emissivity in bands 1-3.

Table 5.6.1. Flux (W/m²) and Percentage Contributions of Spectral Bands to Clear Sky OLR, OLR, and LWCRF.

Band Number	Frequency Range	Clear Sky OLR		OLR		LWCRF	
		Flux	% of total	Flux	% of total	Flux	% of total
1	100-350	34.78	13.17	34.35	14.31	0.43	1.77
2	350-500	42.84	16.22	40.82	17.00	2.02	8.42
3	500-630	38.74	14.67	36.19	15.07	2.55	10.62
4	630-700	10.33	3.91	10.32	4.30	0.01	0.05
5	700-820	31.85	12.06	28.52	11.88	3.32	13.84
6	820-980	45.76	17.32	38.90	16.20	6.86	28.55
7	980-1080	16.62	6.29	14.24	5.93	2.38	9.90
8	1080-1180	17.24	6.53	14.29	5.95	2.95	12.29
9	1180-1390	16.77	6.35	14.34	5.97	2.43	10.11
10	1390-1480	2.42	0.92	2.24	0.93	0.18	0.75
11	1480-1800	2.93	1.11	2.76	1.15	0.18	0.73
12	1800-2080	2.33	0.88	1.96	0.82	0.37	1.52
13	2080-2250	0.76	0.29	0.60	0.25	0.16	0.68
14	2250-2380	0.05	0.02	0.05	0.02	0.00	0.00
15	2380-2600	0.42	0.16	0.31	0.13	0.10	0.42
16	2600-3260	0.32	0.12	0.24	0.10	0.08	0.33

5.7 Analyses on Drought and Flu Season Prediction Applications of AIRS Data

Since its launch in 2002, AIRS data have been widely applied in various research and application fields. It is still an ongoing effort to document the changes on these research and applications using AIRS observations from V7 retrievals. Two examples are given in this section: the drought detection and flu season prediction using AIRS data.

5.7.1 Drought Application of AIRS Data

Contributor: Alireza Farahmand, Yixin Wen

Data

The AIRS Level 2 daily and monthly 0.5 degree aggregated near surface air temperature (Ta), and near surface RH are used to derive Vapor Pressure Deficit (VPD).

Methodology

Daily values of V655 and V6 relative humidity, temperature, and VPD for IR-only retrievals were calculated separately and compared to MesoWest ground observations in July 2011. MesoWest is a cooperative program to collect, archive, and distribute environmental observations across the country. The program has grown steadily since 1997. Observations from over 70 networks, and more than 10000 stations nationally. MesoWest data are processed & quality

controlled every 15 minutes (Horel et al., 2002). Figure 5.7.1 is an example of MesoWest number of observations recorded for July 2011.

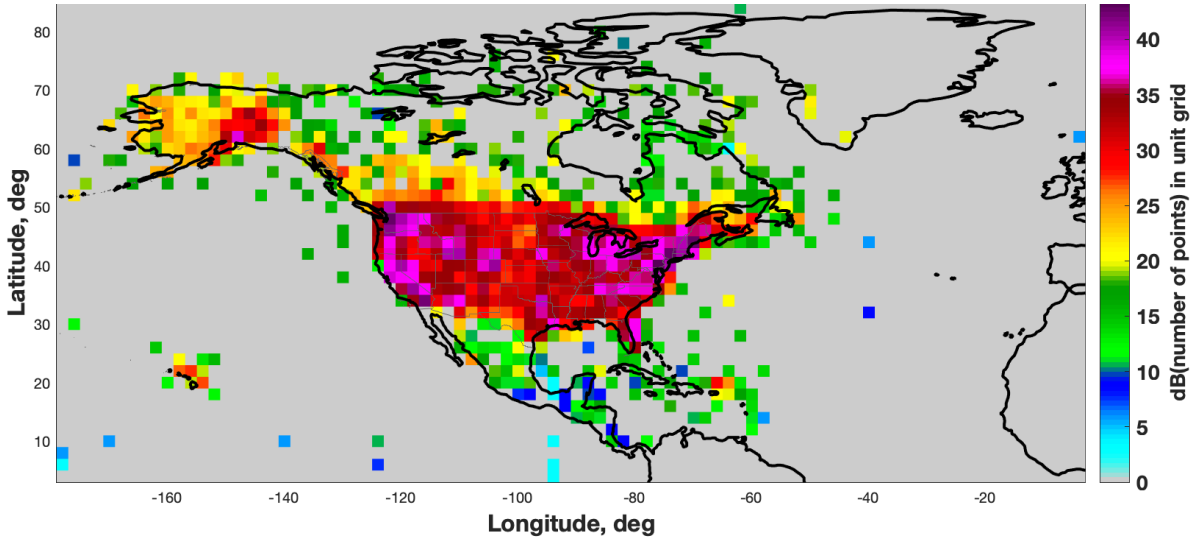
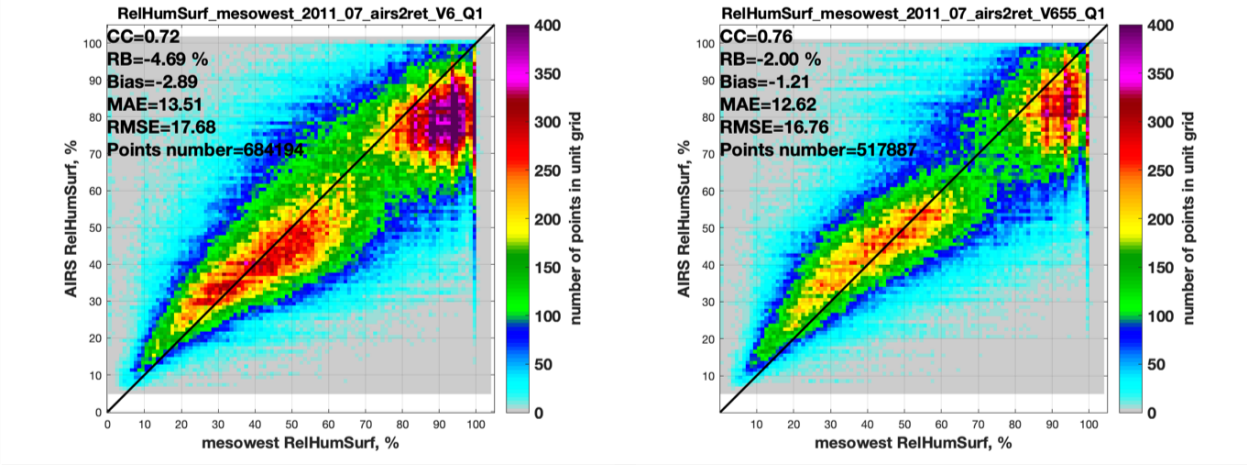


Figure 5.7.1: Mesowest number of records for July 2011

Afterwards, monthly anomaly values of 0.5-degree aggregated V655 and V6 were calculated separately for IR-only retrievals and compared for 2011 major drought in Texas. Since entire V655 climatology is not available yet, V6 climatology was used to derive V655 and V6 anomalies. More discussions about the case studies can be found in Farahmand et al. (2015) and Behrangi et al. (2015).

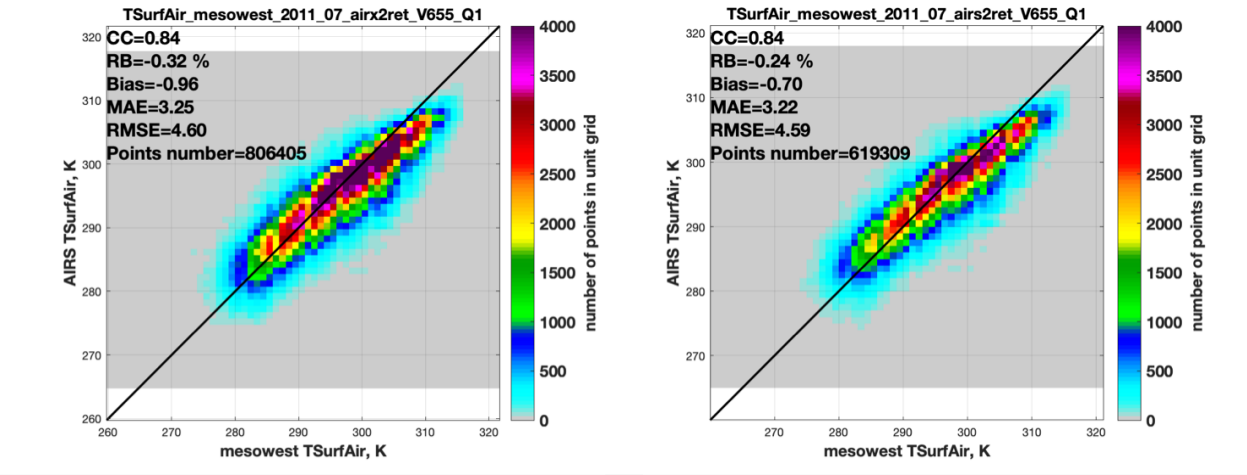
Conclusions

1) V6 and V655 relative humidity, temperature and VPD compare very well with MesoWest observations. Furthermore, no significant difference is shown between V6 and V655 statistics when compared to ground-based observations.



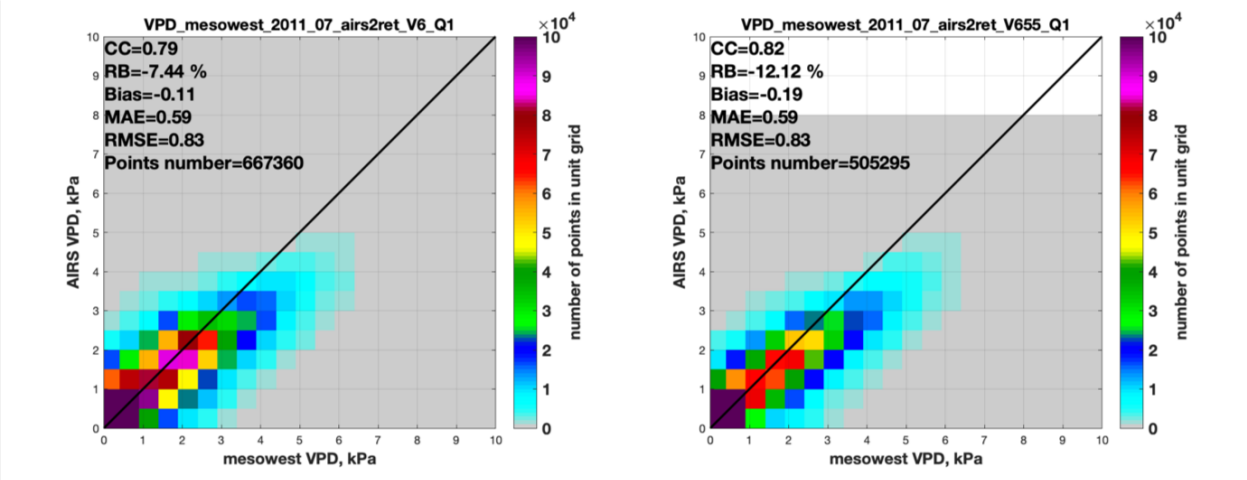
V6 V655

Figure 5.7.2: L2 relative humidity comparisons with Mesowest for July 2011



V6 V655

Figure 5.7.3: L2 temperature comparisons with Mesowest for July 2011



V6 V655

Figure 5.7.4: L2 VPD comparisons with Mesowest for July 2011

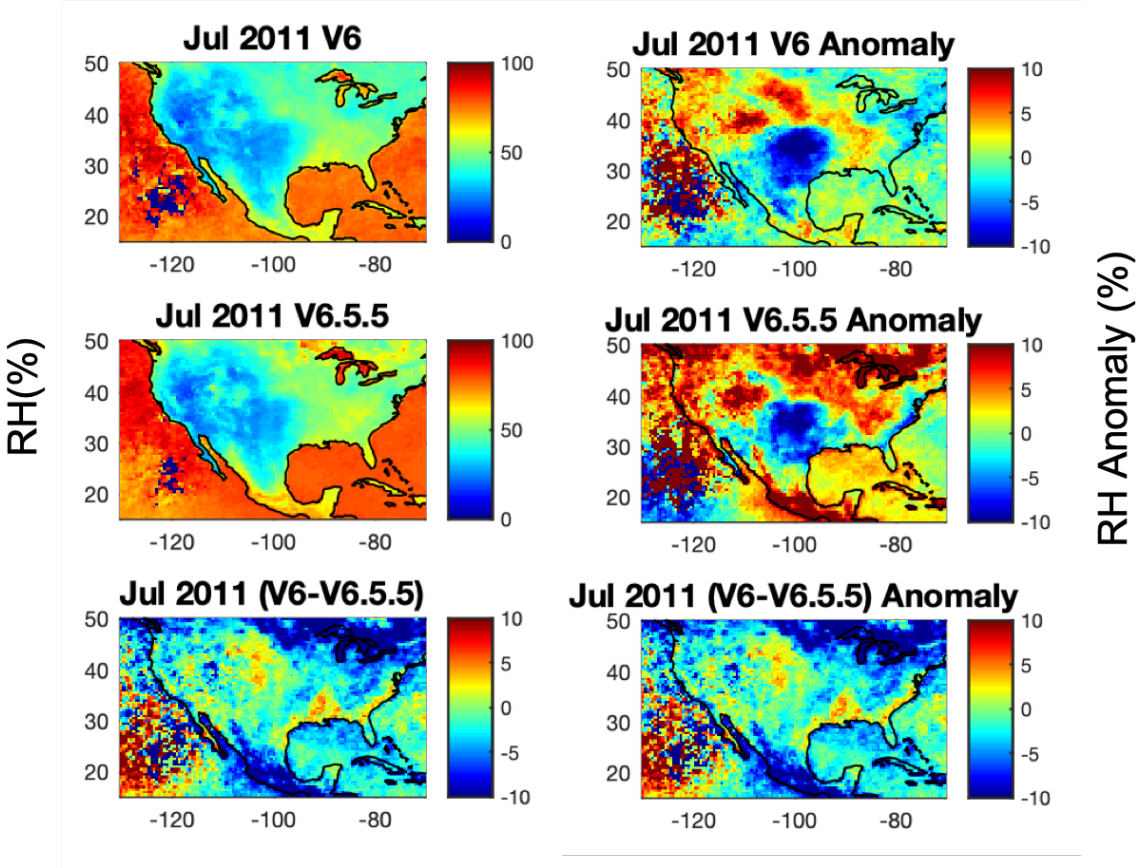


Figure 5.7.5: Relative humidity actual (left) vs anomalies (right) for July 2011. Top: V6, Middle: V6.5.5, Bottom: V6 minus V6.5.5

2) V6 and V655 are compared very well in capturing the anomaly locations and time series. No major impact on drought analysis was identified going from V6 to V655.

- Near surface relative humidity: V655 shows slightly larger values than V6
- Near surface air temperature: V655 shows slightly larger values than V6
- Near surface VPD: V655 shows slightly larger values than V6

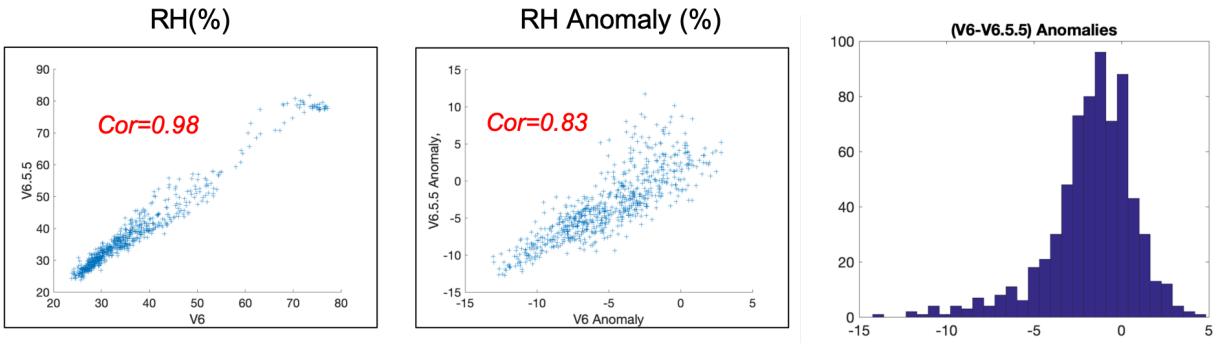


Figure 5.7.6: Comparisons on the monthly V6 vs V6.5.5 relative humidity actual (left), anomaly (middle), and the histogram of relative humidity anomaly differences between V6 and V6.5.5 (right) for July 2011 over Texas.

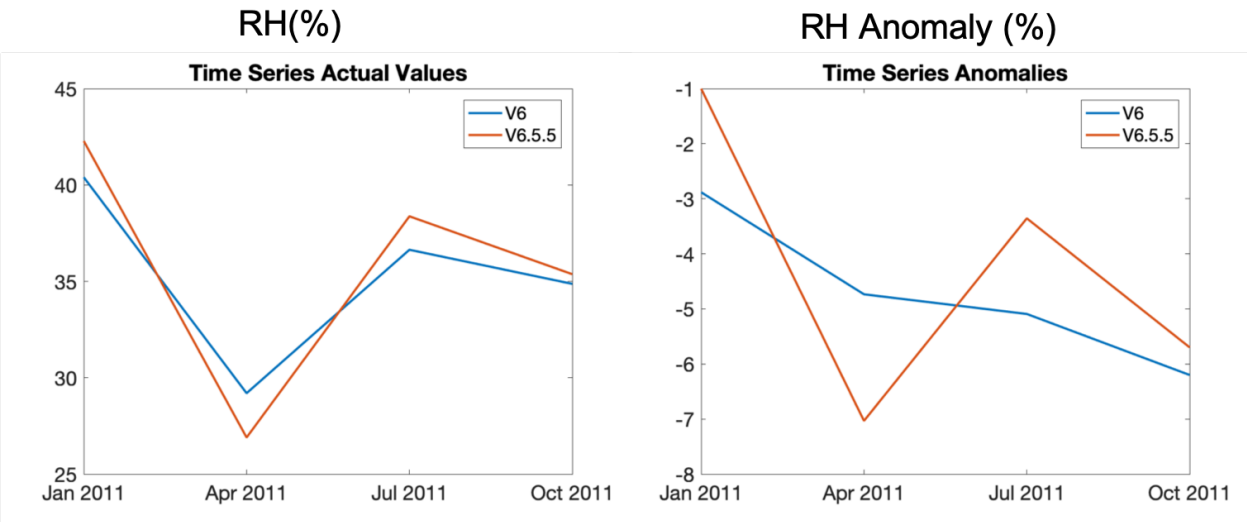


Figure 5.7.7: The time series of RH actual and anomalies during the Texas drought event in 2011. Red: V.6.5.5. Blue: V6.

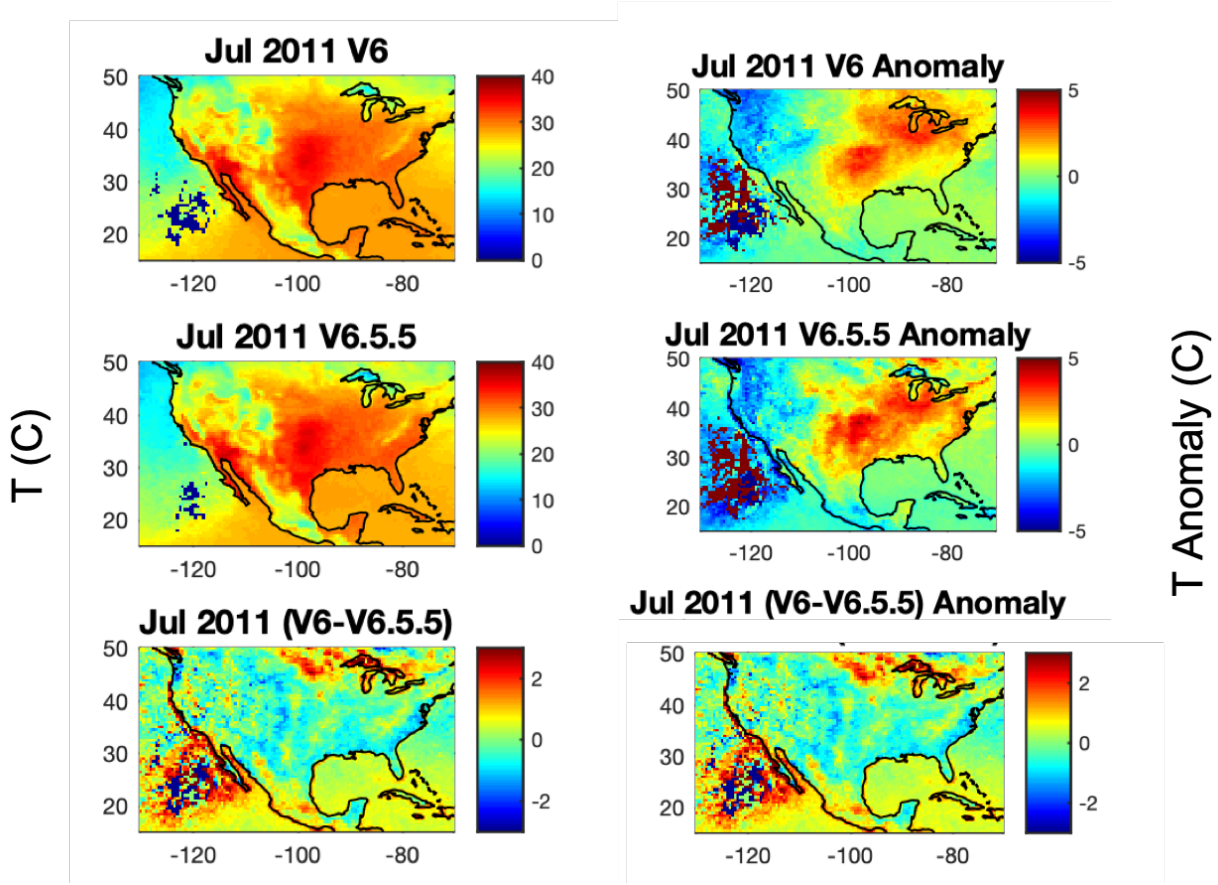


Figure 5.7.8: Temperature actual (left) vs anomalies (right) for July 2011. Top: V6, Middle: V6.5.5, Bottom: V6 minus V6.5.5

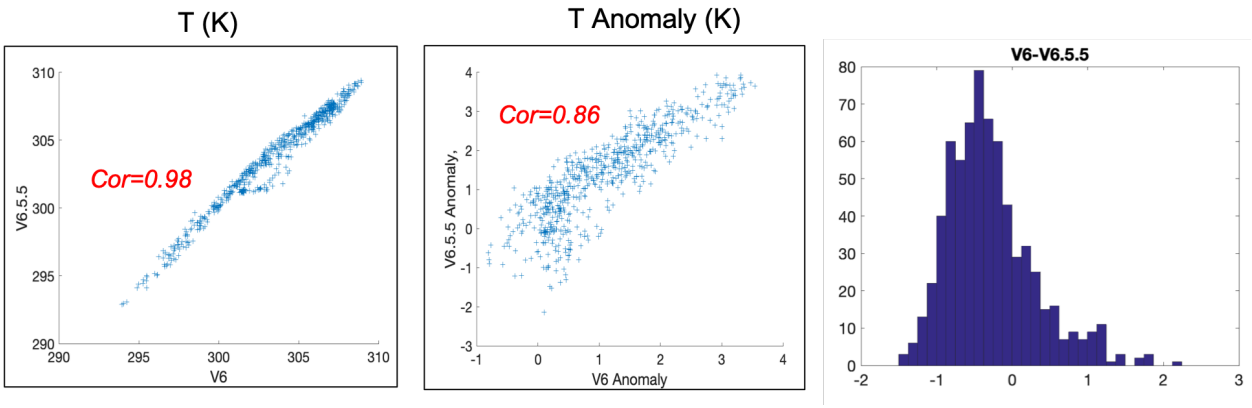


Figure 5.7.9: Comparisons on the monthly V6 vs V6.5.5 temperature actual (left), anomaly (middle), and the histogram of temperature anomaly differences between V6 and V6.5.5 (right) for July 2011 over Texas.

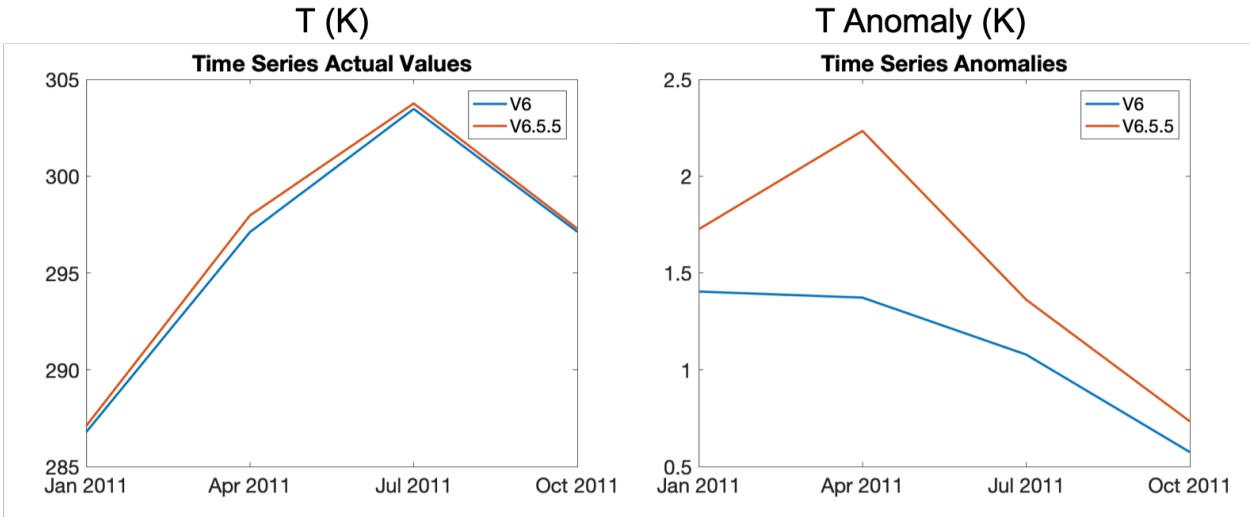


Figure 5.7.10: The time series of T actual and anomalies during the Texas drought event in 2011. Red: V.6.5.5. Blue: V6.

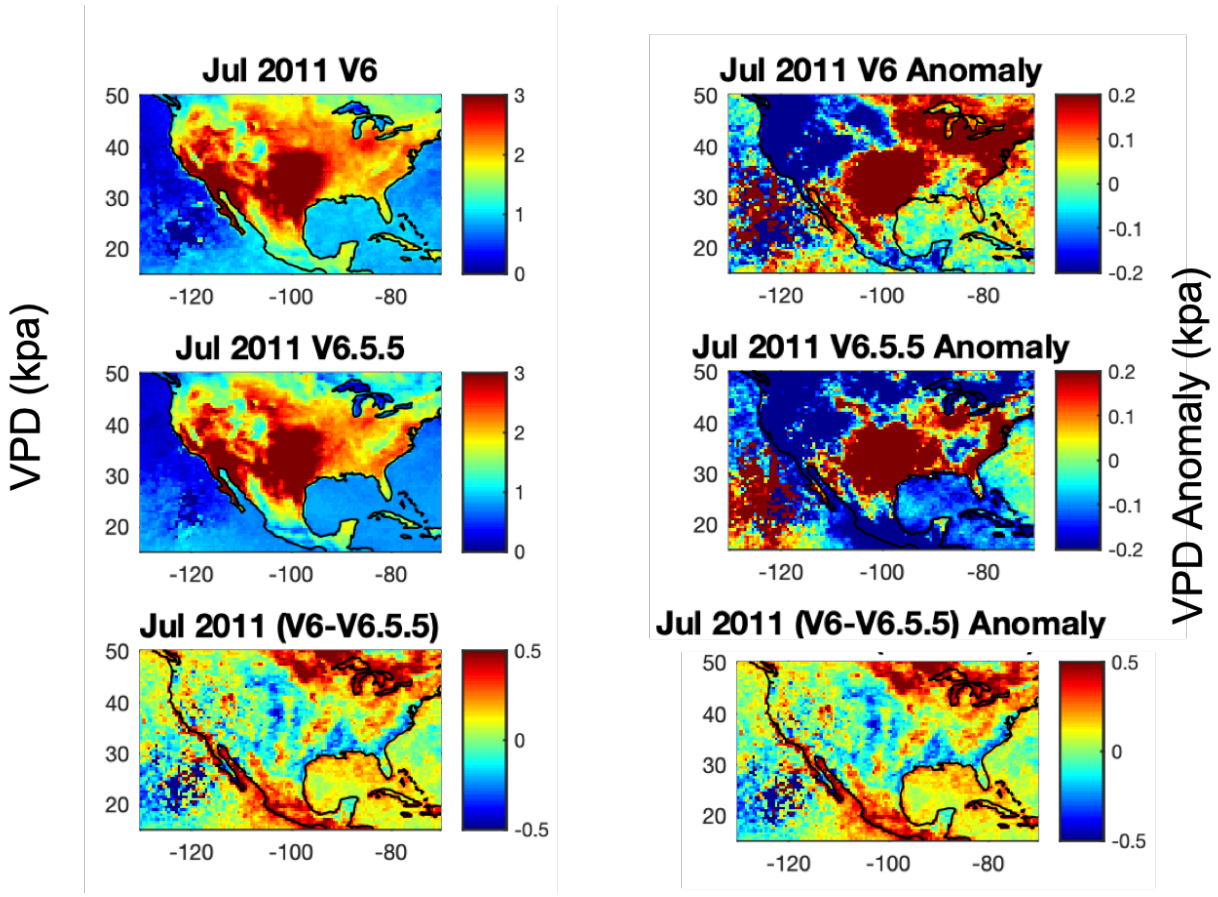


Figure 5.7.11: VPD actual (left) vs anomalies (right) for July 2011. Top: V6, Middle: V6.5.5, Bottom: V6 minus V6.5.5

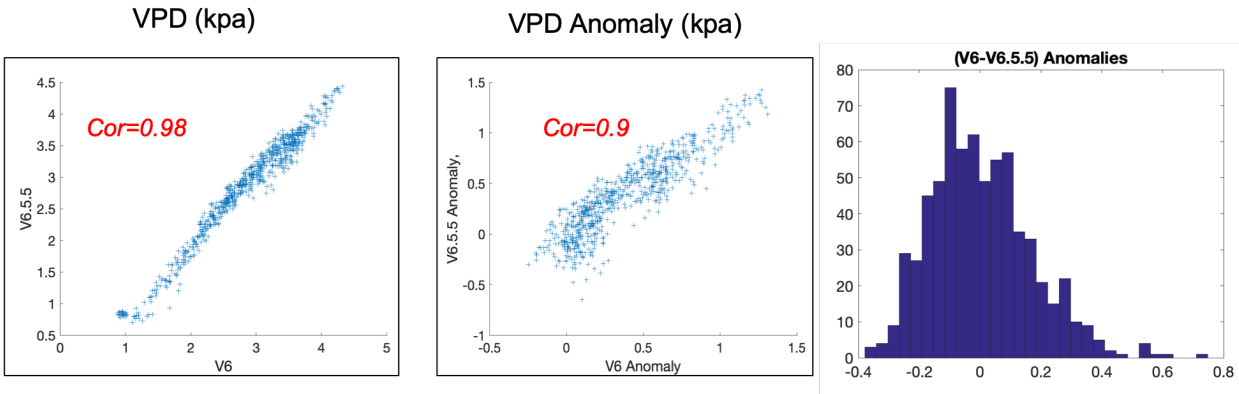


Figure 5.7.12: Comparisons on the monthly V6 vs V6.5.5 VPD actual (left), anomaly (middle), and the histogram of VPD anomaly differences between V6 and V6.5.5 (right) for July 2011 over Texas.

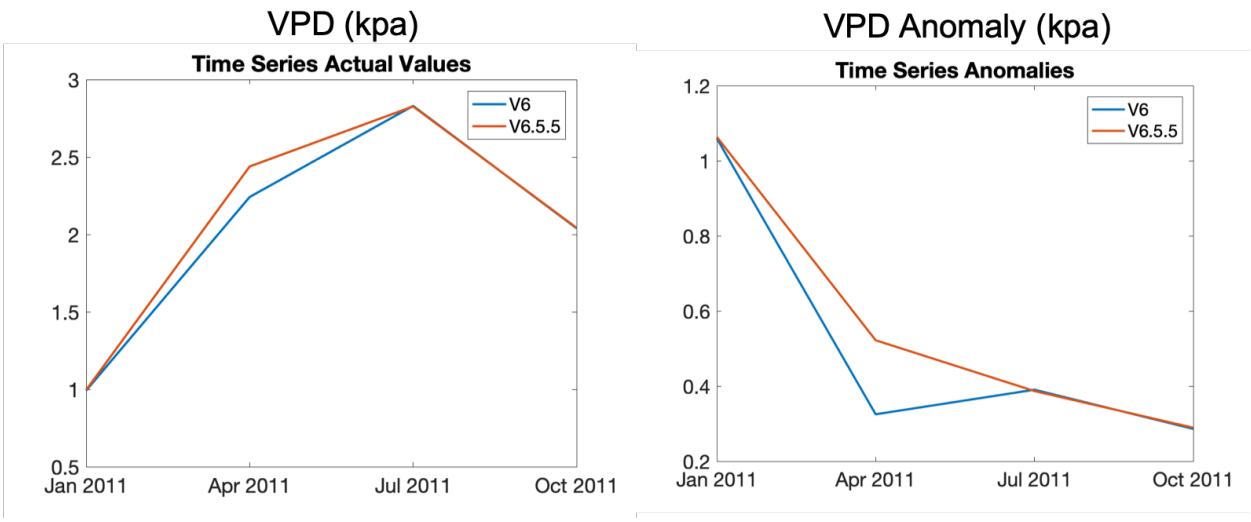


Figure 5.7.13: The time series of VPD actual and anomalies during the Texas drought event in 2011. Red: V.6.5.5. Blue: V6.

5.7.2 Applications of AIRS Data in a Prediction System for Seasonal Influenza

Contributor: Heidar Thor Thrastarson

Data

Based on the hypothesis that a low specific humidity environment is linked with increased influenza transmission, a prediction system for seasonal influenza has been built that uses AIRS humidity data. The AIRS data used is Level 3 daily near-surface water vapor mass mixing ratio, in this case for the ascending branch (H2O_MMR_Surf_A). Time series for four case studies are analyzed here, namely two flu seasons at two locations. The two flu seasons are fall to spring 2017-2018 and 2018-2019 and the two US locations are Los Angeles, CA and Dallas, TX (the 1° x 1° cell covering most of the city area). Results of the influenza prediction system using input near-surface water vapor from AIRS V6 and AIRS V6.55 are compared. No "truth" data is used but regional influenza surveillance data is shown for reference.

Methodology

The purpose of the analyses described here is to perform a sanity check and limited comparison between versions. The influenza prediction system is run with the aim of mimicking as closely as possible how it would be used in an operational setting regularly run throughout the given flu season. For a given "forecast launch date", an ensemble (100 members) calculation is initialized at an earlier time, here nine days earlier, for which influenza surveillance data would be available (using the surveillance data to adjust initial values). Daily AIRS water vapor is used until the launch date, after which NCEP humidity forecasts are used up to a given forecast range, which in this case is seven days. The seven-day forecasts of influenza activity (given in terms of a scaled "ILI+ %" value and with a seven-day running mean taken over ensemble mean values) are collected to form a time series over the flu season, and compared for runs using different AIRS version data sources. The input time series of AIRS near-surface water vapor from the different versions are also compared. A further description of the prediction system is given in Thrastarson & Teixeira (2020, in prep.). For a given location and time period, all model settings

are fixed except the input AIRS near-surface water vapor. No additional quality filtering is performed on the AIRS data.

Conclusions

- Based on the four cases studied, differences between AIRS V6 and V6.55 near-surface water vapor are not expected to have major impacts on the humidity-driven influenza prediction system.
- There are periods where lower humidity in V6.55 appears to lead to higher forecasted flu activity (e.g., for Los Angeles in December, 2017), which is consistent with the expected model behavior. But the resulting differences in flu activity do not have a major effect on how the overall flu seasons would be characterized (e.g., main peak timings, relative strength of season), and there are also examples of higher V6.55 humidity leading to lower forecasted flu activity (to a lesser degree).
- There are significant differences in individual mixing ratio values between the AIRS versions (up to ~50% at several points), but not a clear overall systematic bias for lower or higher values.
- There are more missing data points in V6.55 than V6 for all four cases of time series studied.

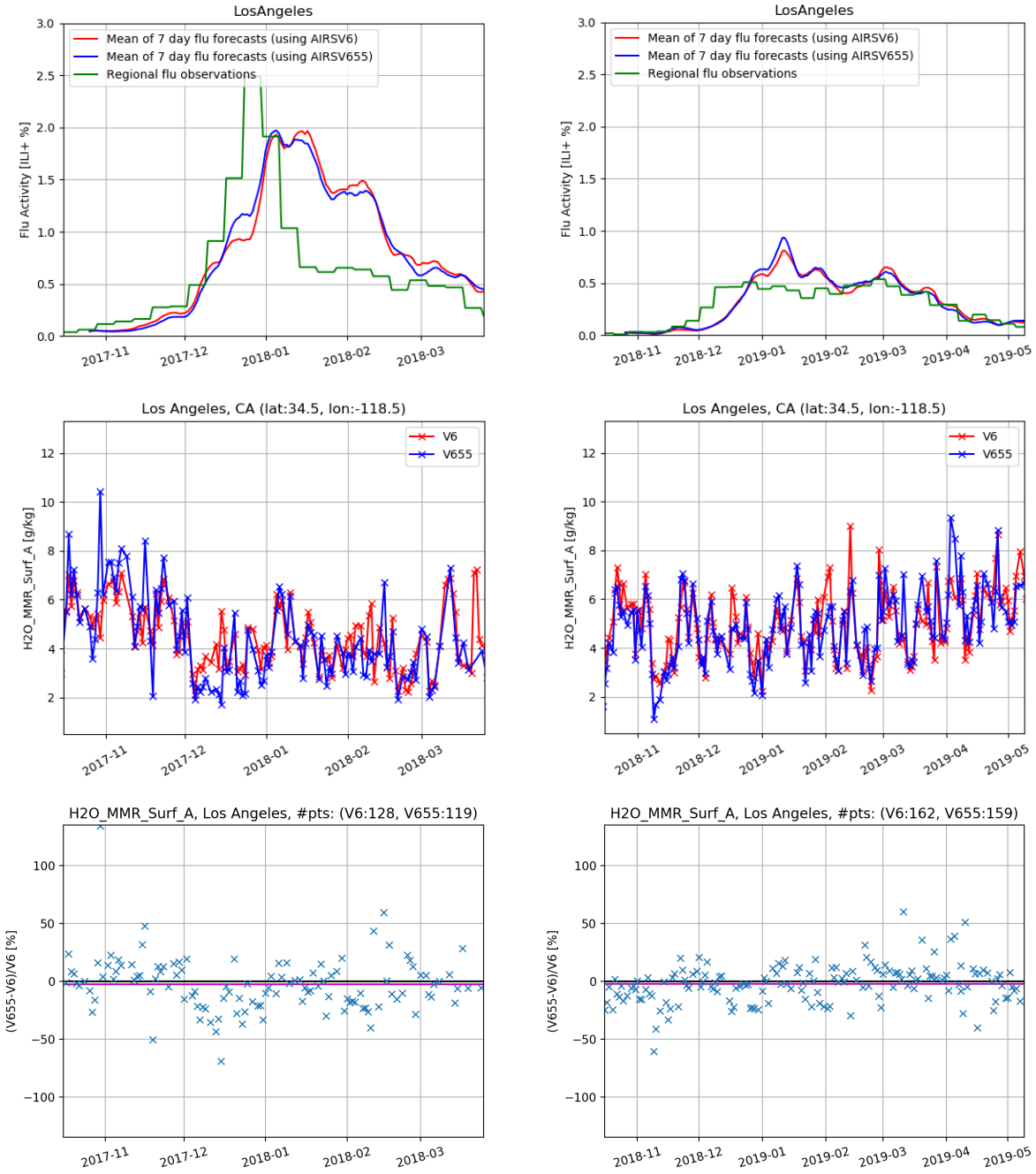


Figure 5.7.14: Comparison of time series for Los Angeles 2017-2018 (left) and 2018-2019 (right). Top: Forecasted influenza activity from the seasonal influenza prediction system driven by V6 (red) and V6.55 (blue) near-surface water vapor mass mixing ratio (H2O_MMR_Surf_A). The time series are composed of ensemble means of seven-day forecast values. The green line shows influenza surveillance data for the multi-state region. Middle: Time series of H2O_MMR_Surf_A for V6 (red) and V6.55 (blue) that are used as input for the influenza prediction system. Bottom: Relative difference of H2O_MMR_Surf_A values from V6 and V6.55 at each point in time. The magenta line shows the mean of the relative differences. The number of valid points in the time series for each version is shown in the plot titles.

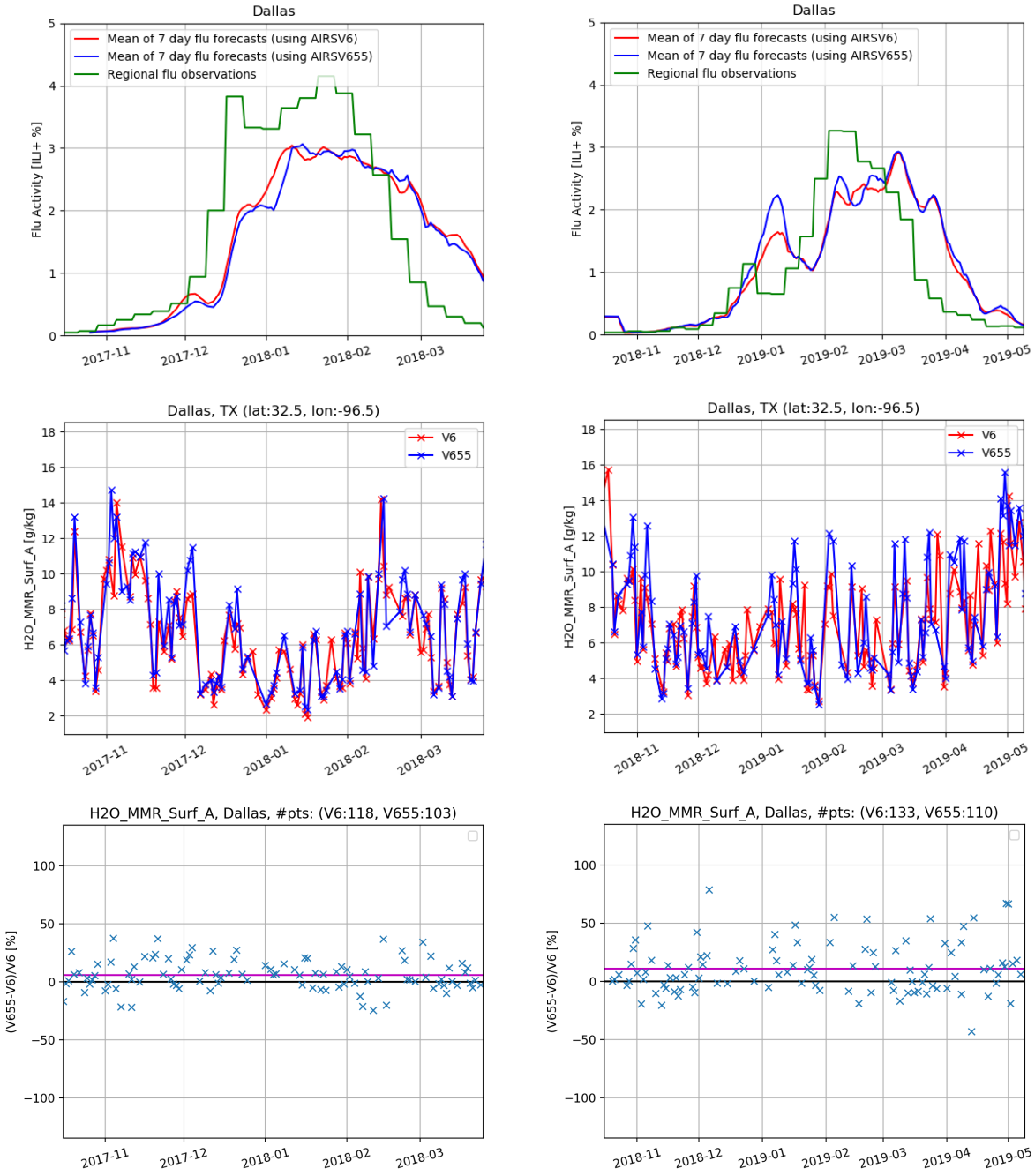


Figure 5.7.15: Comparison of time series for Dallas 2017-2018 (left) and 2018-2019 (right). Top: Forecasted influenza activity from the seasonal influenza prediction system driven by V6 (red) and V6.55 (blue) near-surface water vapor mass mixing ratio (H2O_MMR_Surf_A). The time series are composed of ensemble means of seven-day forecast values. The green line shows influenza surveillance data for the multi-state region. Middle: Time series of H2O_MMR_Surf_A for V6 (red) and V6.55 (blue) that are used as input for the influenza prediction system. Bottom: Relative difference of H2O_MMR_Surf_A values from V6 and V6.55 at each point in time. The magenta line shows the mean of the relative differences. The number of valid points for each version is shown in the plot titles.

6. Averaging Kernels and Other Information Content Metrics for Water Vapor and Temperature Profiles in the AIRS L2 Products

In this chapter we discuss averaging kernels (AK), degrees of freedom of information (DOF or DOFI), vertical resolution, and other measures of information content for AIRS water vapor and temperature profiles, which is reported in the AIRS L2 retrieval products.

The AKs for temperature and water vapor profile retrievals in the AIRS V6 and V7 L2 products represent knowledge arising from the physical retrieval step only, and is therefore not directly comparable with those from a single-step Optimal Estimation retrieval. Since SCCNN (which is used as a first guess for the physical retrieval, more discussion see section 3.1.1 and 4.3) has already extensively used the AIRS radiances, the temperature and water vapor averaging kernels reported in the L2 product only reflect the weighting between the SCCNN first guess and the physical retrieval, instead of the total information content of the AIRS spectra. Therefore the user should be cautious in using the temperature and water vapor AKs in AIRS V6 and V7 L2 product to interpret the information content of AIRS on these variables. Note that the SCCNN only produces first guesses for temperature profile, surface temperature, and water vapor profile. Other quantities from the physical retrieval (O_3, CO, CH_4) begin with a climatological guess, so the AKs and DOF values for these retrievals in the AIRS L2 products are more comparable to those from an optimal estimation algorithm. This section focuses on the information content metrics for water vapor and temperature profiles only. Readers can refer to Section 5.4.1 and 5.4.3 for discussions on AKs for AIRS composition product and analysis on O_3.

Section 6.1 describes how AKs in AIRS L2 products are derived and how they differ from the Optimal-Estimation AKs that many users are familiar with. The analysis in Sec. 6.1.3 shows although the AIRS AKs represent only the physical retrieval step, they can still be used to characterize the sensitivity of AIRS spectra and how the user should interpret the differences on AKs between V6 and V7. Section 6.2 discusses empirical averaging kernels, which do represent the complete retrievals, to quantify the changes in temperature and water vapor products from V6 and V7. Note that empirical AKs cannot be directly compared with the standard AKs and should be used with great caution. Section 6.2 describes a particular application of empirical AKs.

6.1 Degree of Freedom and Verticality for Water Vapor Profile Retrievals Shown by AIRS V7 Averaging Kernel Functions

Contributor: Fredrick Irion, Qing Yue, John Blaisdell, Tao Wang, Sun Wong, and Nadia Smith

6.1.1 Introduction to Averaging Kernel Matrix

The averaging kernel matrix, $\boldsymbol{A}$, describes the sensitivity of the retrieved vector, $\hat{\boldsymbol{x}}$, to changes in the true state, $\boldsymbol{x}$ (e.g, Rodgers 2000):

$$\boldsymbol{A} = \frac{\partial \hat{\boldsymbol{x}}}{\partial \boldsymbol{x}} = \frac{\partial \hat{\boldsymbol{x}}}{\partial \boldsymbol{F}} \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{x}} = \boldsymbol{G}\boldsymbol{K}, \tag{6.1.1}$$

where $\boldsymbol{F}$ is computed radiance for the retrieval state vectors using the forward model, the gain matrix $\boldsymbol{G} = \partial \hat{\boldsymbol{x}} / \partial \boldsymbol{F}$ is a measure of the sensitivity of the retrieval, $\hat{\boldsymbol{x}}$, to changes in the radiance, and the Jacobian $\boldsymbol{K} = \partial \boldsymbol{F} / \partial \boldsymbol{x}$ is the matrix of derivatives of the radiance to changes in each element of the state vector. $\boldsymbol{A}$ is a square matrix dimensioned $n \times n$, where n is the number of elements of the state vector $\boldsymbol{x}$, which can include scalar values such as surface temperature or

individual cloud properties, and vertical profiles such as trace gases and atmospheric temperature. Each element of $\boldsymbol{A}$ is a measure of the sensitivity for one retrieved member of a state vector ($\hat{x}_i$) to the changes in the true value of that member (x_i) or to the true value of a different member (x_j). Therefore, one row of $\boldsymbol{A}$ contains measures of the sensitivity of a single element of the state vector to changers in the true values of itself and one column of $\boldsymbol{A}$ contains measures of the sensitivity of the entire retrieval state vector to changes in a single true state vector element. The row sums (verticality) are indicative of how much information come from observation: a value near unity indicates that the retrieval at that layer relies mostly on the observed spectra data, while a value near zero indicates most reliance on the a priori data. Degree of freedom (DOF) is the trace of an averaging kernel matrix measuring the number of independent pieces of information in the retrieval. As discussed by Irion et al. (2018) using an optimal estimation based AIRS retrieval algorithm, an averaging kernel is scene dependent. Sensitivities at different layers, for example, can depend on the amounts of trace gases present, the temperature lapse rate, the cloud field, the view geometry, and the spectral channels employed.

6.1.2 Averaging Kernel in the AIRS V6 and V7 Retrieval

Since the V6 and V7 AIRS retrieval algorithms do not use an optimal estimation scheme, the averaging kernels in the AIRS L2 product are different from those described in Rodgers (2000). Below we present a summary of the AIRS retrieval scheme for temperature and trace gases, and calculation of the averaging kernel. For more detailed descriptions readers should refer to *Susskind et al*. (2003) and *Maddy et al*. (2008).

In the current AIRS algorithm, the retrieval of a profile quantity (temperature, water vapor, ozone, etc.) begins with a first guess profile, $\hat{\mathbf{x}}_0$, on the 100-level AIRS support grid[1]. The profile is then iteratively perturbed on a series of superimposed basis functions (in this case, trapezoids):

$$\begin{aligned}\hat{\mathbf{x}}_m &= \hat{\mathbf{x}}_{m-1} + \mathcal{T}\left(\Delta\hat{\boldsymbol{\alpha}}_m\right) \\ &= \hat{\mathbf{x}}_0 + \mathcal{T}\hat{\boldsymbol{\alpha}}_m\end{aligned} \tag{6.1.2}$$

where

$\hat{\mathbf{x}}_m$ is the retrieved profile vector at iteration *m*,
$\hat{\mathbf{x}}_0$ is the first guess for the profile vector,
$\mathcal{T}$ is a matrix describing the trapezoids,
$\hat{\boldsymbol{\alpha}}_m$ is a vector of scaling factors for the trapezoids at iteration *m*, and
$\Delta\hat{\boldsymbol{\alpha}}_m$ is a vector of the incremental changes in the scaling factors $\boldsymbol{\alpha}$, that is:

[1] The AIRS retrieval grid is 100 levels. The numbering of the grid begins at the lowest pressure, p_1 = 0.0016 mb, increasing in pressure to p_{100} = 1100 mb. In the case of temperature, $\mathbf{x}_j$ is defined as the temperature at level j. In the case of a gas, $\mathbf{x}_j$ is the integrated amount of gas between level *j* and level *j-1* (a slab column). This is converted to mean mixing ratios in the standard product.

$$\Delta\hat{\boldsymbol{\alpha}}_m = \hat{\boldsymbol{\alpha}}_m - \hat{\boldsymbol{\alpha}}_{m-1}. \tag{6.1.3}$$

The number of rows in $\mathcal{T}$ is equal to the number of levels on the support grid (100), while the number of columns is the number of trapezoids. The endpoints of the trapezoids are selected so that the profile is roughly Nyquist sampled in the vertical, with ~2 trapezoids per retrievable layer quantity. Initially, the sum along the horizontal of the trapezoids is unity at all levels except the very top or bottom. As the trapezoids scale with $\boldsymbol{\alpha}$, the retrieved profile changes as the superimposition of the trapezoids.

We define the spectral residual for channel *i* as the difference between the (cloud-cleared) observed radiance, L_i, and the forward model calculated radiance, $\left[\mathbf{F}(\mathbf{x})\right]_i$. As the radiances can vary by orders of magnitude within the frequencies of the AIRS bandpass, the residual is normalized by the inverse gradient of the Planck function radiance with respect to temperature. This puts the spectral residual in terms of brightness temperature, not radiance.

$$\delta\mathrm{y}_i = \left(\mathrm{L}_i - \left[\mathbf{F}(\mathbf{x})\right]_i\right)\left(\left.\frac{dB_i}{d\mathrm{T}}\right|_{\hat{\mathrm{T}}_b\left(\left[\mathbf{F}(\mathbf{x})\right]_i\right)}\right)^{-1} \tag{6.1.4}$$

where:

$\delta\mathrm{y}_i$ is the spectral residual at channel *i*,
L_i is the cloud-cleared radiance,
$\left[\mathbf{F}(\mathbf{x})\right]_i$ is the forward modeled radiance, and
$\left.\frac{dB_i}{d\mathrm{T}}\right|_{\hat{\mathrm{T}}_b\left(\left[\mathbf{F}(\mathbf{x})\right]_i\right)}$ is the gradient of the Planck function, *B*, with respect to temperature, T. The gradient is calculated at the brightness temperature, $\hat{\mathrm{T}}_b$, for the forward modeled radiance at channel *i*.

We define the Jacobian as the derivative of the modeled radiance by the trapezoid scaling factors, also normalized by the inverse gradient of the Planck function:

$$\begin{aligned}\mathrm{K}_{i,l} &= \frac{\partial\left[\mathbf{F}(\mathbf{x})\right]_i}{\partial\alpha_l}\left(\left.\frac{dB_i}{d\mathrm{T}}\right|_{\hat{\mathrm{T}}_b\left[\mathbf{F}(\mathbf{x})\right]_i}\right)^{-1} \\ &= -\frac{\partial\left(\delta\mathrm{y}_i\right)}{\partial\alpha_l}\end{aligned} \tag{6.1.5}$$

Note that for a gas, the (normalized) derivative of the radiance with respect to the scaling factor is equivalent to taking the derivative with respect to the natural logarithm of the gas amount. For temperature, the change in α is set to be equivalent to a linear change in temperature over a trapezoid. The iterative solution for the trapezoid scaling factor is:

$$\begin{aligned}\Delta\hat{\boldsymbol{\alpha}} &= \left[\mathbf{K}^T\mathbf{S}_\varepsilon^{-1}\mathbf{K} + \mathbf{H}\right]^{-1}\mathbf{K}^T\mathbf{S}_\varepsilon^{-1}\delta\mathbf{y} \\ &= \mathbf{G}\delta\mathbf{y}\end{aligned} \tag{6.1.6}$$

where:

$\mathbf{S}_\varepsilon$ is the radiance error covariance matrix,
$\mathbf{H}$ is a stabilization matrix, and
$\mathbf{G}$ is a gain matrix, that is:

$$\mathbf{G} = \left[\mathbf{K}^T\mathbf{S}_\varepsilon^{-1}\mathbf{K} + \mathbf{H}\right]^{-1}\mathbf{K}^T\mathbf{S}_\varepsilon^{-1} \tag{6.1.7}$$

Formulation of the error covariance matrix, $\mathbf{S}_\varepsilon$, is described in detail by *Susskind et al*. (2003), however we note here that the cloud-cleared radiance errors are correlated and, again, the individual matrix elements are normalized by the derivatives of the Planck function:

$$\mathbf{S}_{\varepsilon_{ij}} = \frac{\mathbf{S}^*_{\varepsilon_{ij}}}{\left(\left.\frac{dB_i}{d\mathrm{T}}\right|_{\hat{\mathrm{T}}_b[\mathbf{F}(\mathbf{x})]_i}\right)\left(\left.\frac{dB_j}{d\mathrm{T}}\right|_{\hat{\mathrm{T}}_b[\mathbf{F}(\mathbf{x})]_j}\right)} \tag{6.1.8}$$

where $\mathbf{S}^*_{\varepsilon_{ij}}$ is an error covariance term that incorporates uncertainties from radiative noise and uncertainties in temperature, water vapor, etc. The formulation of the gain matrix, $\mathbf{G}$, is similar to that of Equation 3.27 in *Rodgers* (2000), but unlike *Rodgers*, the stabilization matrix $\mathbf{H}$ is not a constraint from an *a priori* covariance. Constraint of the retrieval is determined through a singular value decomposition including $\mathbf{K}^T\mathbf{S}_\varepsilon^{-1}\mathbf{K}$, with the higher eigenvalued functions more weighted in the solution. We rewrite Equation A1:

$$\begin{aligned}\hat{\mathbf{x}}_m &= \hat{\mathbf{x}}_{m-1} + \mathcal{T}\left(\Delta\hat{\boldsymbol{\alpha}}_m\right)\\ &= \hat{\mathbf{x}}_{m-1} + \mathcal{T}\mathbf{L}\left(\Delta\hat{\boldsymbol{\xi}}_m\right)\end{aligned} \tag{6.1.9}$$

Substituting for $\Delta\boldsymbol{\alpha}$, Equation A7 then becomes[2]:

$$\Delta\hat{\boldsymbol{\xi}} = \left[\mathbf{L}^T\mathbf{K}^T\mathbf{S}_\varepsilon^{-1}\mathbf{K}\mathbf{L} + \mathbf{H}\right]^{-1}\mathbf{L}^T\mathbf{K}^T\mathbf{S}_\varepsilon^{-1}\delta\mathbf{y} \tag{6.1.10}$$

Noting that $\mathbf{S}_\varepsilon^{-1}$ is a symmetric matrix, so that $\mathbf{K}^T\mathbf{S}_\varepsilon^{-1}\mathbf{K}$ is also symmetric, $\mathbf{L}$ is chosen so that $\boldsymbol{\Lambda} = \mathbf{L}^T\mathbf{K}^T\mathbf{S}_\varepsilon^{-1}\mathbf{K}\mathbf{L}$, with eigenvalues $\boldsymbol{\lambda}$ on the diagonal matrix $\boldsymbol{\Lambda}$. Substitution into Equation 6.1.10 gives:

$$\Delta\hat{\boldsymbol{\xi}} = \left[\boldsymbol{\Lambda} + \mathbf{H}\right]^{-1}\mathbf{L}^T\mathbf{K}^T\mathbf{S}_\varepsilon^{-1}\delta\mathbf{y} \tag{6.1.11}$$

To limit instabilities, $\mathbf{H}$ is constructed as a diagonal matrix such that no diagonal element of $\boldsymbol{\Lambda} + \mathbf{H}$ approach zero. A "minimum damping" is thus applied to $\mathbf{L}^T\mathbf{K}^T\mathbf{S}_\varepsilon^{-1}\delta\mathbf{y}$ in (Eq. 6.1.11) such

[2] In practice, $\delta\mathbf{y}$ is modified slightly after the first iteration to account for non-linearities in the radiative transfer (see *Susskind et al.* (2003)), but no modification is made in calculating the averaging kernel.

that any element of the transformed step-size vector $\Delta\hat{\boldsymbol{\xi}}$ is limited to an empirically determined maximum. The untransformed scaling factors for the trapezoids are then calculated:

$$\Delta\hat{\boldsymbol{\alpha}} = \mathbf{L}\left(\Delta\hat{\boldsymbol{\xi}}\right) \tag{6.1.12}$$

Calculation of the averaging kernel, **A**, is defined on the reduced measurement space of the trapezoids, $\mathcal{T}$, and scale factors, $\boldsymbol{\alpha}$:

$$\begin{aligned}\mathbf{A} &= \frac{\partial\hat{\boldsymbol{\alpha}}}{\partial\boldsymbol{\alpha}} \\ &= \frac{\partial\left(\Delta\hat{\boldsymbol{\alpha}}\right)}{\partial(\delta\mathbf{y})}\frac{\partial(\delta\mathbf{y})}{\partial\left(\Delta\boldsymbol{\alpha}\right)} \\ &= \frac{\partial\left(\Delta\hat{\boldsymbol{\alpha}}\right)}{\partial(\delta\mathbf{y})}\mathbf{K} \\ &= \mathbf{L}\left[\boldsymbol{\Lambda}+\mathbf{H}\right]^{-1}\mathbf{L}^{T}\mathbf{K}^{T}\mathbf{S}_{\varepsilon}^{-1}\mathbf{K}\end{aligned} \tag{6.1.13}$$

Multiplying both sides by $\mathbf{L}\mathbf{L}^{T}$ (= **I** since $\mathbf{L}^{T} = \mathbf{L}^{-1}$):

$$\begin{aligned}\mathbf{A} &= \mathbf{L}\left[\boldsymbol{\Lambda}+\mathbf{H}\right]^{-1}\mathbf{L}^{T}\mathbf{K}^{T}\mathbf{S}_{\varepsilon}^{-1}\mathbf{K}\mathbf{L}\mathbf{L}^{T} \\ &= \mathbf{L}\left[\boldsymbol{\Lambda}+\mathbf{H}\right]^{-1}\boldsymbol{\Lambda}\mathbf{L}^{T}\end{aligned} \tag{6.1.14}$$

Converting the averaging kernel from a trapezoid basis to one on the AIRS 100-level gridding can be done by the following. Recalling the trapezoid matrix, $\mathcal{T}$, the least-squares inverse is calculated:

$$\mathcal{T}' = \left[\mathcal{T}^{T}\mathcal{T}\right]^{-1}\mathcal{T}^{T} \tag{6.1.15}$$

The averaging kernels can be used with "truth" data, $\mathbf{x}_T$, on the 100-level AIRS support grid as follows:

$$\mathbf{x}_{est} = \mathbf{x}_0 + \mathcal{T}\mathbf{A}\mathcal{T}'(\mathbf{x}_T - \mathbf{x}_0) \tag{6.1.16}$$

In the following discussion, we show the averaging kernels of vertical profiles of AIRS water vapor ($\partial\ln[\widehat{H_2O}]/\partial\ln[H_2O]$) and temperature ($\partial\hat{T}/\partial T$) with the aim to quantify the changes between V6 and V7 and to explain the physical meaning of the averaging kernels reported in the current AIRS L2 support products.

6.1.3 The Change of Averaging Kernels and DOF for Temperature and Water Vapor in AIRS V7 L2 Product

The averaging kernels and DOF output for temperature and water vapor profile retrievals in the AIRS V6 and V7 L2 products represent knowledge arising from the physical retrieval step only, and is therefore not directly comparable with a DOF output from a single-step Optimal Estimation retrieval. Since SCCNN (which is used as a first guess for the physical retrieval, more discussion see section 3.1.1 and 4.1.3) has already extensively used the AIRS radiances, the temperature and water vapor averaging kernels reported in the L2 product and calculated DOF only reflect the weighting chosen between the SCCNN first guess and the physical retrieval, instead of the total information content of the AIRS spectra.

In the V7 algorithm, several changes have been made on the water vapor retrieval (See section 3.1.4) and briefly summarized here: 1) increasing damping (which gives the SCCNN first guess greater weight relative to the physical retrieval than in V6) to avoid large degradation of near surface water vapor profiles relative to SCCNN results, 2) removal of shortwave channels to remove the spurious day-night performance differences in V6, 3) other channel changes including removal of strongest water lines and the addition of water vapor continuum channels, and 4) a new V7 SCCNN first guess and changes in the humidity trapezoids. As a result, it is expected the water vapor averaging kernels reported in the AIRS L2 support product will be different in V7 and V6. Three days of AIRS data in January 14-16, 2015 are used to highlight the version differences. These days are chosen as previous studies have shown AIRS retrievals have reduced yield and larger bias in cold season over midlatitude land region. As shown in Figure 6.1.1-6.1.2 there is small difference in the DOF for temperature between V6 and V7 with V7 is slightly larger than V6, while water vapor DOF show a clear reduction in V7. Such reduction shows the change of weight between humidity SCCNN and the physical retrieval step, and is a result of increased damping and removal of shortwave channels used in the humidity retrieval.

In order to attribute the changes due to different factors, three additional test runs are carried out: 1) V7 algorithm but with V6 damping factor (V655 less damping), 2) V7 algorithm with shortwave channels added and V6 damping (V655 less damping + SW), 3) V7 algorithm with full water vapor retrieval channels in V6 and V6 damping (V655 less damping + allV6chnl). It is clear that both the removal of shortwave channels and increased damping cause the decrease of water vapor DOF and changes in the verticality in V7, in which larger damping factor implemented in V7 water vapor retrieval is the dominant reason.

The SCCNN does not produce an averaging kernel or DOF equivalent variable to examine the information from AIRS spectra used in SCCNN. As a result, the temperature and water vapor averaging kernels in the V6 and V6 AIRS L2 support products do not characterize the full information content and sensitivity of the AIRS instrument.

Note that the SCCNN only produces values for temperature profile, surface temperature, and water vapor profile. Other quantities from the physical retrieval (O_3, CO, CH_4) begin with a climatological guess, so the averaging kernels and DOF values for these retrievals are more comparable to those from an optimal estimation algorithm.

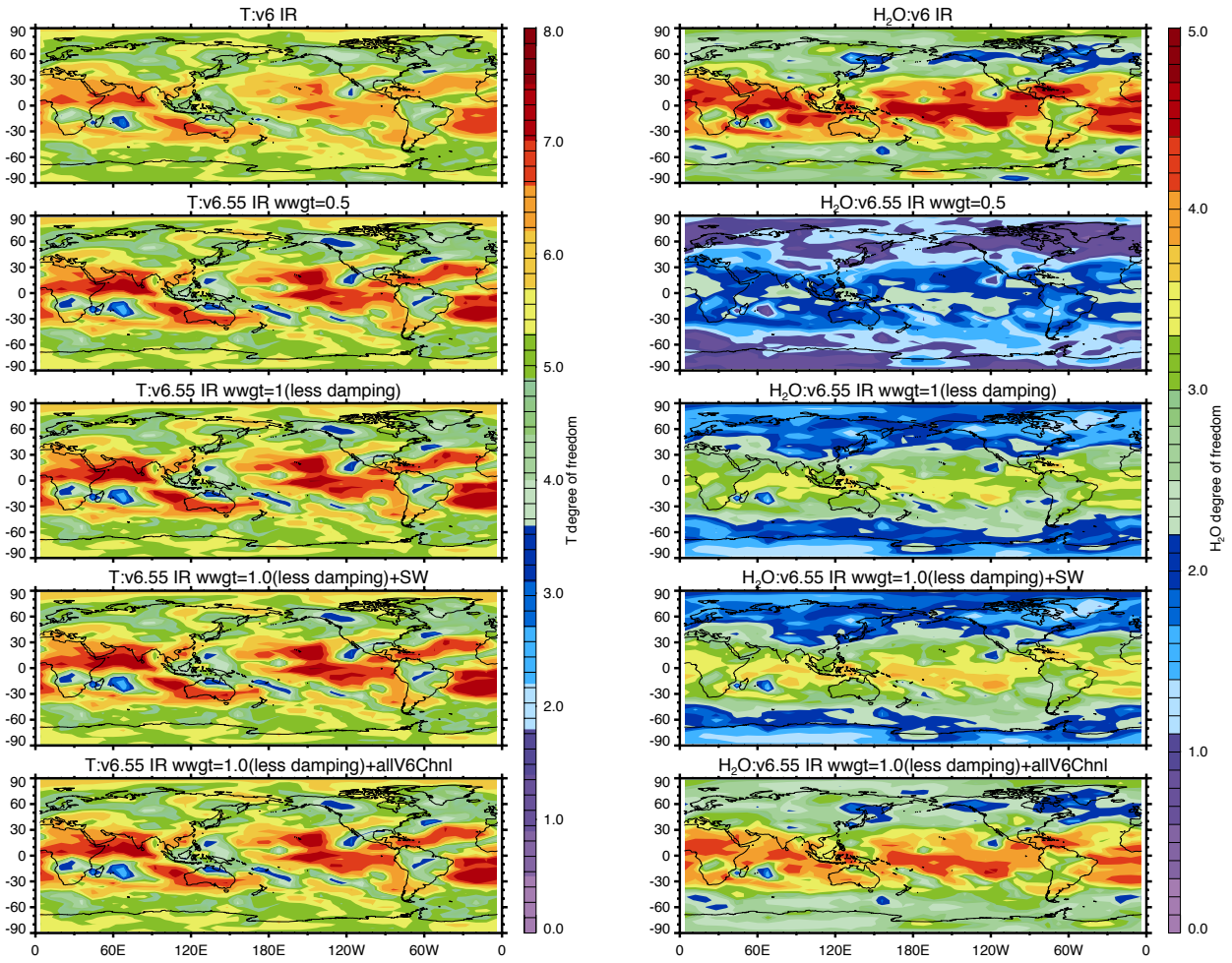


Figure 6.1.1. Maps of degree of freedom (DOF) calculated from temperature (left) and water vapor (right) averaging kernels in AIRS IR-only L2 retrievals. Data that pass the AIRS quality control (QC=0 or 1) from January 14-16, 2015 is used in the analysis. V6 and V6.55 results are shown in the top two rows. The bottom three tows show results from the three test runs.

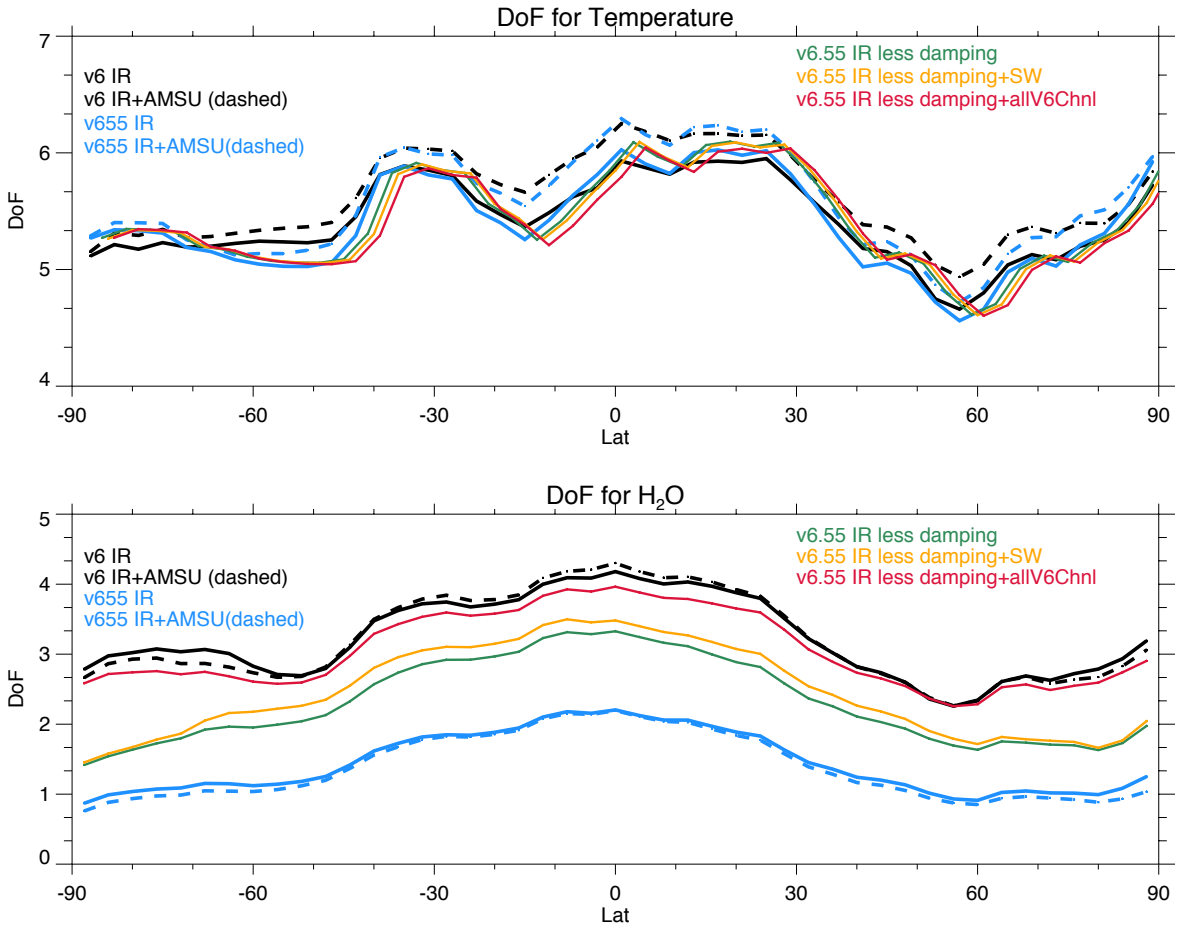


Figure 6.1.2. Zone mean of degree of freedom (DOF) calculated from temperature (top) and water vapor (bottom) averaging kernels in AIRS L2 retrievals. Data that pass the AIRS quality control (QC=0 or 1)from January 14-16, 2015 is used in the analysis. Dashed lines show IR+MW results and solid lines show IR only results. Colors correspond with different versions of AIRS retrievals: black for V6, blue for V6.55, and green, orange, and red for three different test cases.

6.2 AIRS Temperature and Water Vapor Vertical Structures Examined Using the Empirical Averaging Kernel Method and PREPQC Radiosonde Observations

Contributor: Yixin Wen, Yuliya Marchetti, Fredrick (Bill) Irion, and Erica Griffin

A linear mathematical model for the statistical estimate of the bias and noise of AIRS L2 T and q profiles is presented in this section. For two different measurement systems with spatial and temporal noncoincidence on the same state of the atmosphere, a sensible comparison cannot be reduced to a simple straightforward point-by-point analysis (Rodgers and Connor, 2003). A proper statistical methods should be used to transform the correlative data to account for different spatial and temporal resolution between two measurements, as well as for different weighting functions, averaging kernels, and noise (Pougatchev, 2008). One proper statistical method used in this study is named Empirical Averaging Kernels (EAKs) in Pougatchev (2008).

Data and Methodology

AIRS L2 T and q profiles were matched to the National Centers for Environmental Prediction (NCEP) quality controlled final observation data files (PREPQC). Radiosonde observations were used as the "true" state to characterize the AIRS retrievals. The EAKs were then computed as a linear coefficient between radiosonde and AIRS measurements for temperature and water vapor:

$$EAKs = C_{x,r}^{-1} C_{x\hat{x}} \quad , \tag{6.2.1}$$

where $\hat{x}$ is AIRS measurements of atmosphere states, x is radiosonde measurements of atmospheric states, C_x is auto-correlation matrix for radiosonde, and $C_{x\hat{x}}$ is cross-correlation matrix for AIRS and radiosonde measurements.

When using the full rank matrix, C_x, the averaging kernel is too noisy, thus the following procedure was used to reduce the rank of C_x and choose the averaging kernel based on the asymptotic behavior of its trace as a function of rank:

1. Perform the singular value decomposition of C_x:
$$C_x = UDV^T \tag{6.2.2}$$
2. Reduce the rank of C_x and calculate its pseudo-inverse:
$$C_{x,r}^{-1} = \mathrm{V} D_r^+ U^T \tag{6.2.3}$$
3. Calculate the EAKs for all ranks, r = 1,…,R:
$$EAKs = C_{x\hat{x}} C_{x,r}^{-1} \tag{6.2.4}$$
4. Choose the EAKs based on the appropriate rank.

Rows of the EAKs were obtained with each row being a smoothing or regression function for the altitude corresponding to that row. The width of the peak of the function is defined as the vertical resolution of the retrieval. The area under each row, or the sum of the row elements, indicate whether the retrieval system is sensitive to the true profile. It should be unity for those altitudes that are sensitive and accurate. Degrees of freedom, or the trace of the EAKs, were also computed to obtain the number of independent sources of information or signal in the retrieval.

The EAKs were computed for temperature and water vapor using the matched observations for January in 2011. In order to calculate the EAKs of complete profiles, the quality control is performed to remove radiosonde profiles with significant missing data. Any matchups of temperature profiles with missing data in one profile greater than 25% are discarded. Any matchups of water vapor profiles with missing data in one profile greater than 75% are discarded.

Results

Radiosonde observations and AIRS retrievals for temperature and water vapor were separated by latitude into the polar, mid-latitude, and tropic regions. The numbers of these matched profiles with quality control were calculated for January 2011 and are provided in Table 6.2.1. Expectedly, the greatest number of matched temperature profiles occurred in the mid-latitudes and the least occurred in the tropics.

Table 6.2.1: Numbers of matched temperature and water vapor profiles in the polar, mid-, and tropic latitudes, for January 2011.

	Polar		Mid-Latitude		Tropics	
	V6	**V655**	**V6**	**V655**	**V6**	**V655**
Temperature	7468	18462	45599	59100	2246	2621

Water vapor	19025	19063	80309	80515	6940	6949

Examples of vertical profiles of the temperature for all radiosonde and AIRS observations at each of the latitude regions are displayed in Fig.6.2.1. These data compare closely with each other, providing validation and confidence in the accuracy of the AIRS retrievals. Data for the polar, mid-, and tropic latitude regions compare well at all levels except for a relatively shallow layer of increased temperature at the top of the AIRS profiles, which does not appear in the radiosonde profiles.

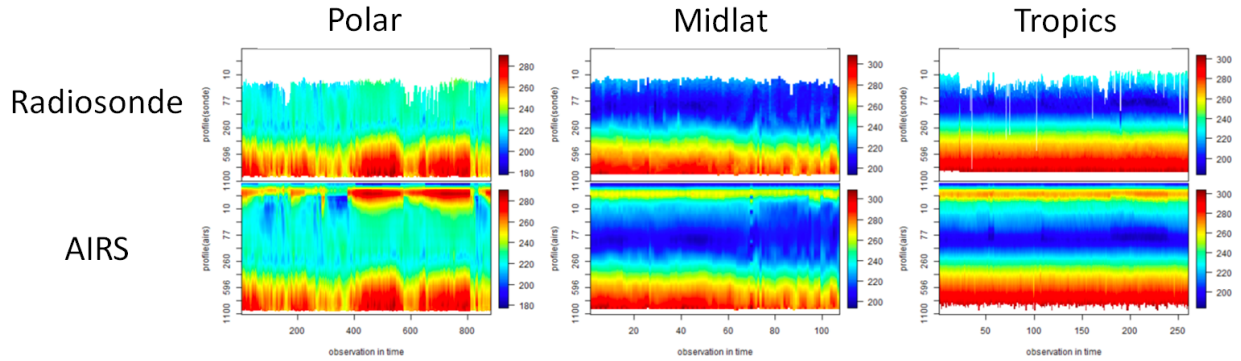
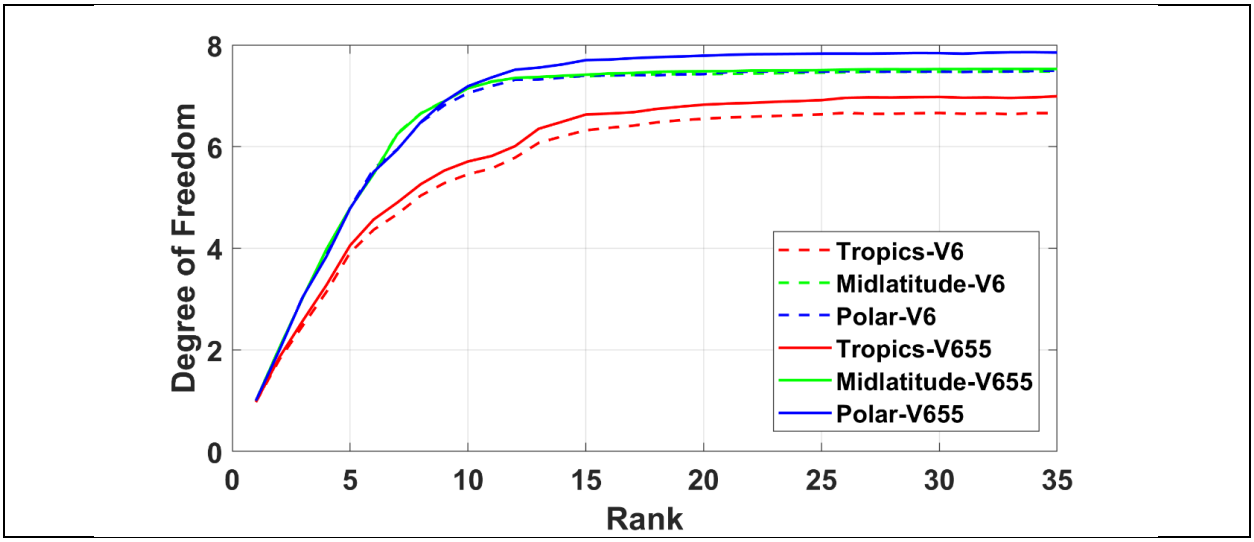


Fig. 6.2.1: Examples of matched radiosonde observations and AIRS retrievals for temperature in the polar, mid-, and tropic latitudes, for January 2011. Color scale indicates temperature in Kelvin.

Degrees of freedom (DOF) with different ranks for Version 6 compared to that of Version 655 for different latitude regions are presented in Fig. 6.2.2. For temperature in the upper figure, DOFs of V655 is slightly higher than of V6 for tropic region (red line) and polar region (blue line), while no difference between two versions in mid-latitude region (green line). For water vapor, the DOFs of new V655 are all lower than that of V6. All DOFs increase dramatically with an increasing rank until rank of 12 for temperature and rank of 7 for water vapor. Rank 12 and Rank 7 therefore have been selected to calculate EAKs in this study.



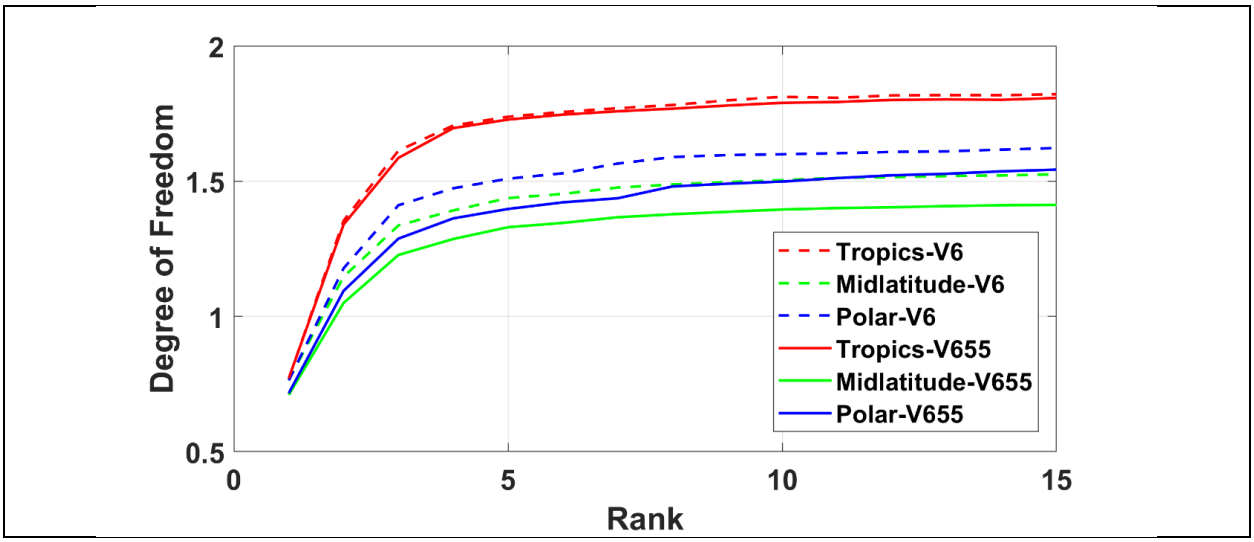
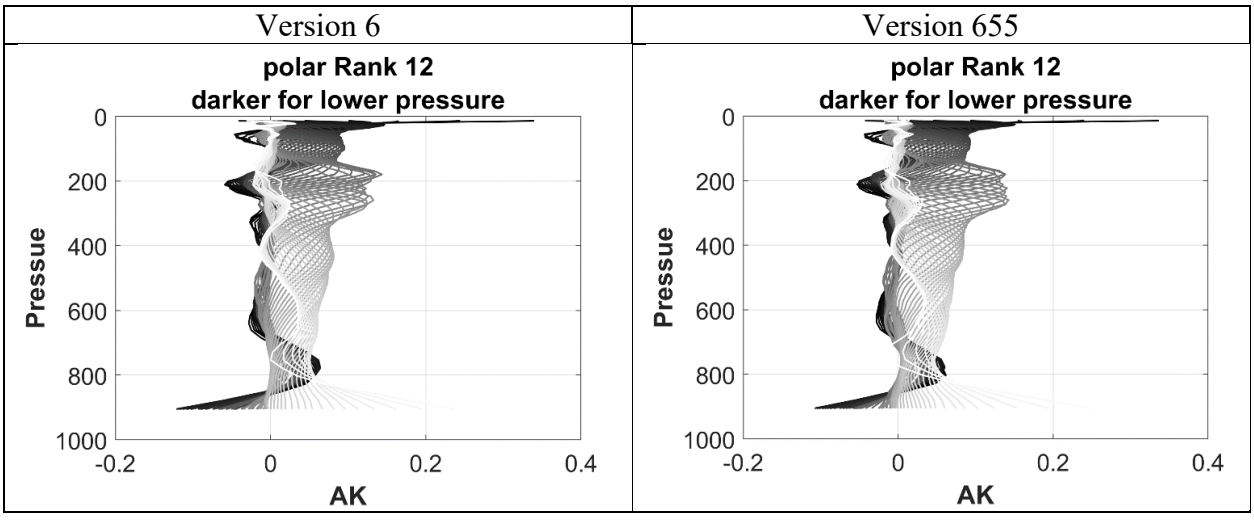


Fig. 6.2.2. Degree of freedom for temperature (upper panel) and water vapor (low panel) in the tropic (green lines), mid- (blue lines), and polar (red lines) latitudes. Dashed lines represent data for V655 and solid lines represent data for V6.

EAKs of temperature for V6 and V655 at Rank 12 are shown in Fig. 6.2.3 in the polar, mid-, and tropic latitude regions, respectively. The data for the two versions compare well and exhibit averaging kernels between approximately -0.5 and 0.2. In polar and mid latitude regions, high EAKs are from lower pressure lever (higher latitude) around 300 hPa. EAKs in tropic region lower EAK values. This is because EAKs indicates the potential of AIRS measurement accuracy, and the homogeneous temperature distribution in tropic region makes the auto-correlation matrix, C_x is relatively smaller compared to other regions. Thus the lower EAKs in tropic regions do not mean worse AIRS measurements in tropics.



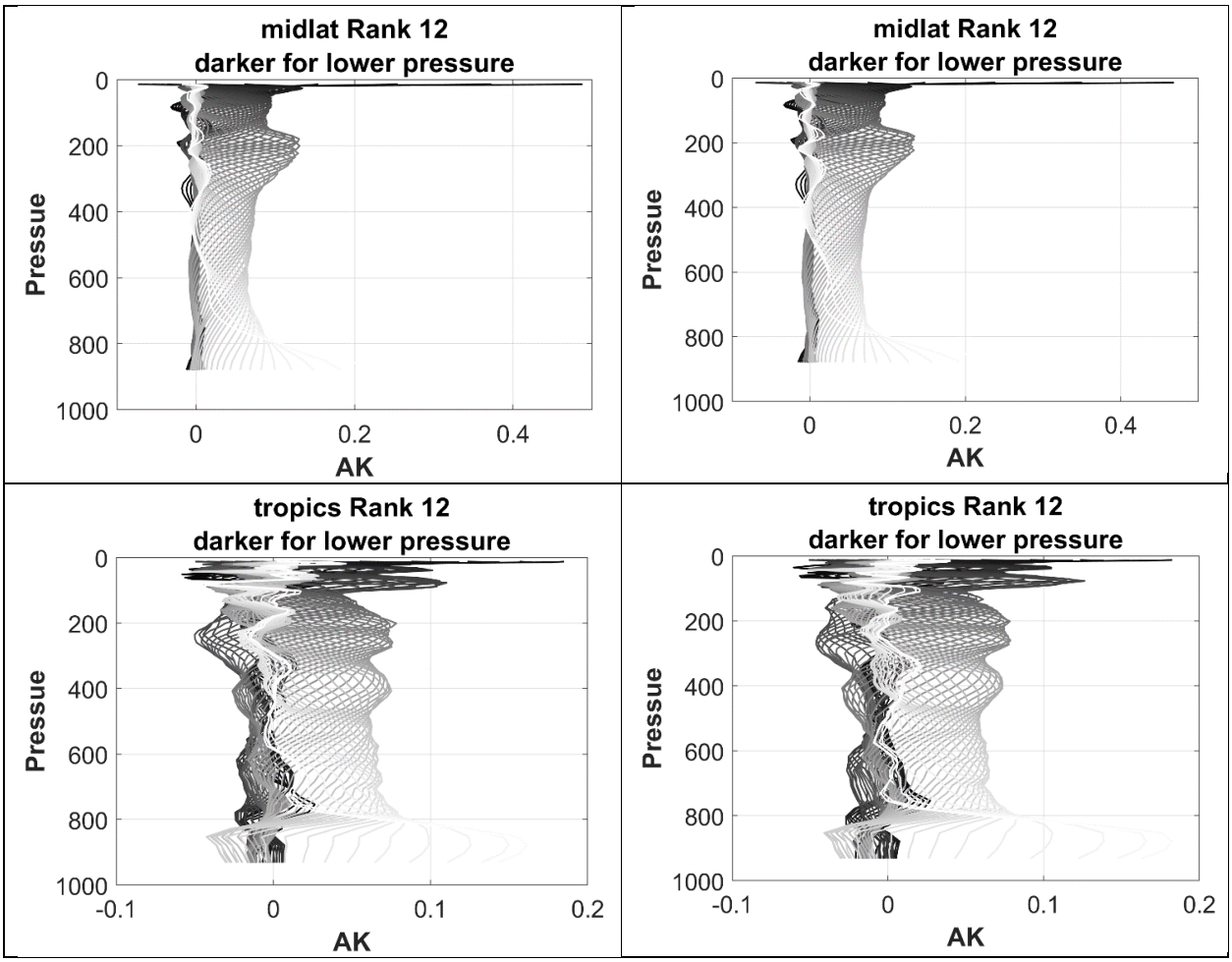


Fig. 6.2.3. EAKs for temperature in the polar, mid-, and tropic latitude regions in January 2011, comparing V6 (left column) and V655 (right column) for rank 12. Empirical averaging kernel profiles with darker grey-colored lines corresponding with lower pressure.

Fig. 6.2.4 shows the EAKs of water vapor resulting from V6 and V655 for three latitude regions. Unlike temperature EAKs, water vapor EAKs do not show reasonable results at each pressure level.

V6	V655

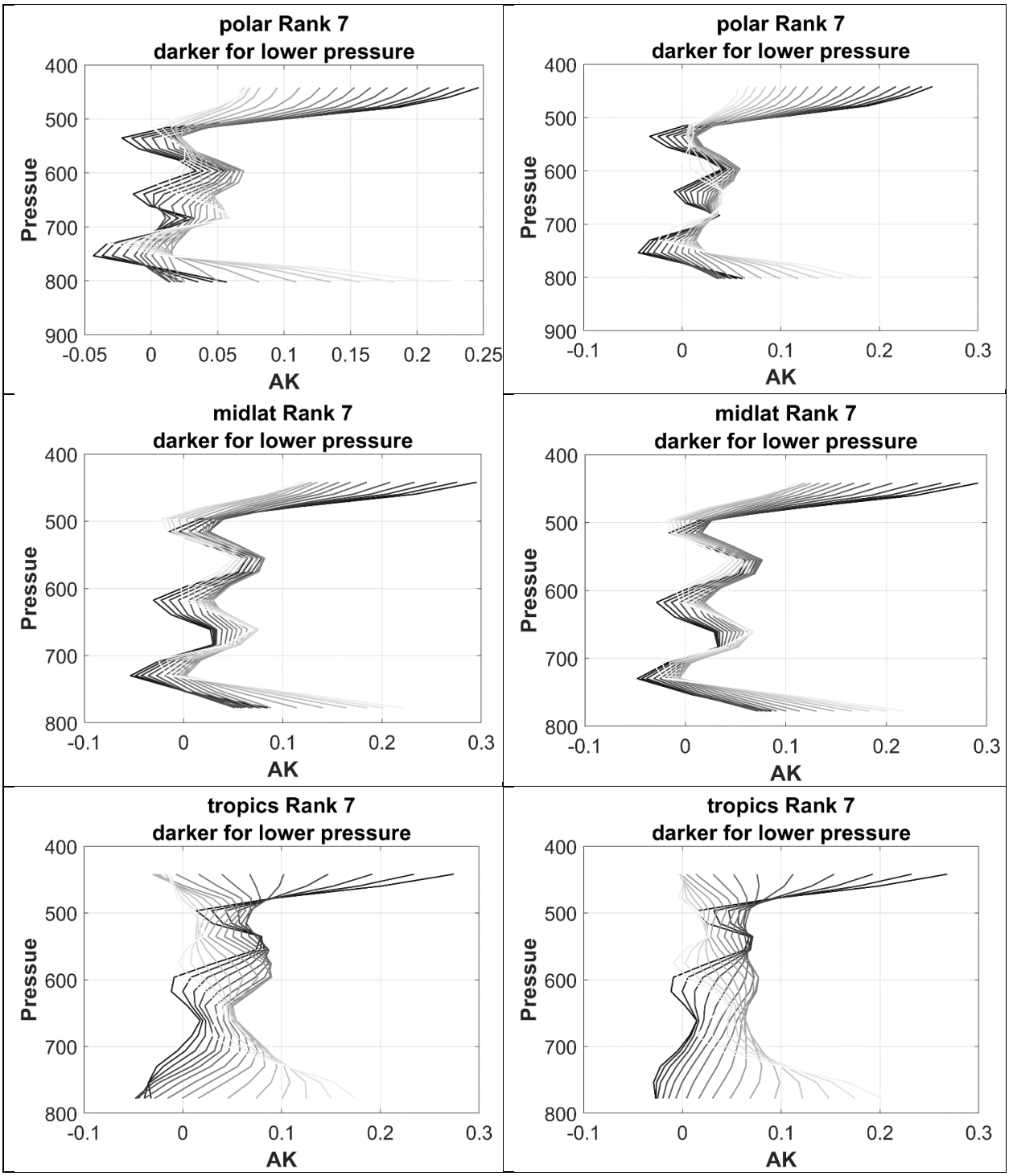


Fig. 6.2.4. Similar to Fig. 6.2.2, but for water vapor.

Fig. 6.2.5 demonstrates row sums of the EAKs for V6 and V655 for temperature and water vapor. Row sums of temperature EAKs (left column) for V6 and V655 compare well overall, particularly for the polar and mid- latitude regions with row sums of approximately unit. Row sums of temperature for the tropics were more variable. Again, this is caused by the homogeneous horizontal distribution of temperature in tropics. Row sums of water vapor EAKs (right column)

for V6 and V655 shows V6 is generally closer to the unit line than V655 in polar and mid- latitude regions. In tropic region, V655 is better than V6 at pressures above approximately 600 hPa.

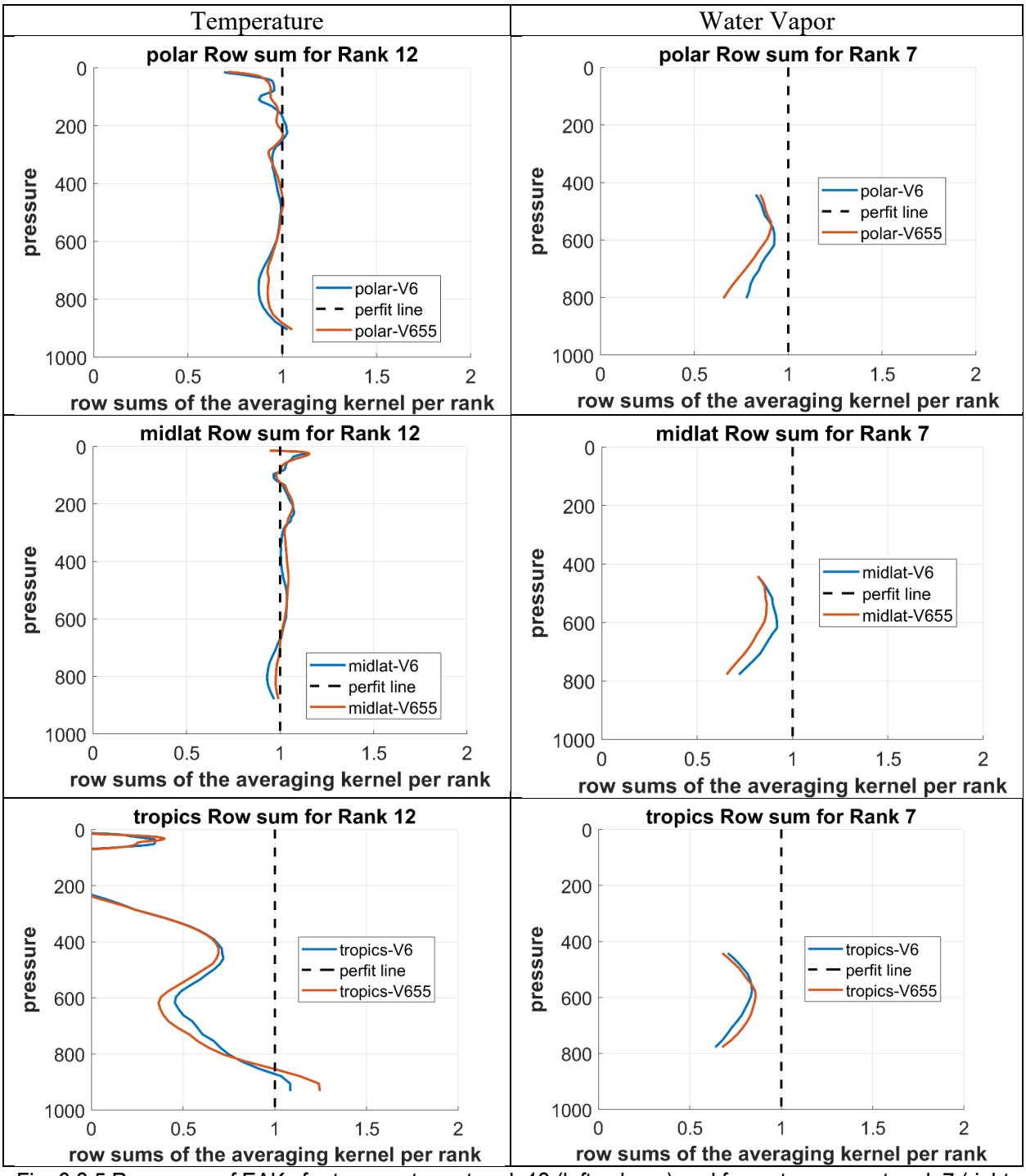


Fig. 6.2.5 Row sums of EAKs for temperature at rank 12 (left column) and for water vapor at rank 7 (right column). Blue lines represent data for V6 and red lines represent data for V655.

Conclusions

Empirical averaging kernels are in general agreement for V6 and V655. For temperature, DOF of temperature of V655 is higher than that of V6, while water vapor of V655 is lower than that of V6. Row sums of temperature EAKs for the polar and mid- latitude regions are approximately unit, but row sums of temperature for the tropics were more variable. Row sums of water vapor EAKs for V6 and V655 show V6 is generally closer to the unit line than V655 particularly in polar and mid- latitude regions.

7. References

Behrangi, A., P. Loikith, E. Fetzer, H. Nguyen, and S. Granger, 2015: Utilizing Humidity and Temperature Data to Advance Monitoring and Prediction of Meteorological Drought. *Climate*, **3,** 999-1017.

Dang, V. H., and coauthors, ed., 2017, AIRS/AMSU/HSB Version 6 Level 2 Performance and Test Report, available from: https://docserver.gesdisc.eosdis.nasa.gov/repository/Mission/AIRS/3.3_ScienceDataProductDocumentation/3.3.5_ProductQuality/V6_L2_Performance_and_Test_Report.pdf.

Edwards, D., Anderson, G., Oakley, T., and Gault, P., Met Office intercomparison of Vaisala RS92 and RS41 radiosondes, U. K. Met Office, Exeter, U.K., 2014.

Farahmand A., AghaKouchak A., Teixeira J., 2015, A Vantage from Space Can Detect Earlier Drought Onset: An Approach Using Relative Humidity, Scientific Reports, 5, 8553

Fetzer, E. J., et al, ed., 2020, AIRS V7 Algorithm Theoretical Basis Document, Jet Propulsion Laboratory, California Institute of Technology.

Fishbein, E, 2018, IDL tool: ECMWF_INTERP_LLT.

Granskog, M. A., P. Assmy, S. Gerland, G. Spreen, H. Steen and L. H. Smedsrud, 2016: Arctic research on thin ice: Consequences of Arctic sea ice loss, Eos, 97, doi:10.1029/2016EO044097.

Horel, J., et al. "Weather support for the 2002 winter olympic and paralympic games." Bulletin of the American Meteorological Society 83.2 (2002): 227-240.

Irion, F. W., Kahn, B. H., Schreier, M. M., Fetzer, E. J., Fishbein, E., Fu, D., Kalmus, P., Wilson, R. C., Wong, S., and Yue, Q.: Single-footprint retrievals of temperature, water vapor and cloud properties from AIRS, Atmos. Meas. Tech., 11, 971–995, https://doi.org/10.5194/amt-11-971-2018, 2018.

Kahn, B. H., Chahine, M. T., Stephens, G. L., Mace, G. G., Marchand, R. T., Wang, Z., Barnet, C. D., Eldering, A., Holz, R. E., Kuehn, R. E., and Vane, D. G.: Cloud type comparisons of AIRS, CloudSat, and CALIPSO cloud height and amount, Atmos. Chem. Phys., 8, 1231–1248, https://doi.org/10.5194/acp-8-1231-2008, 2008.

Keating, G.M., Pitts, M.C., Young, D.F., 1990. Ozone referencemodels for the middle atmosphere. Advances in SpaceResearch 10 (12), 317–355.

Loeb, Norman & R. Doelling, David & Wang, Hailan & Su, Wenying & Nguyen, Cathy & Corbett, Joseph & Liang, Lusheng & Mitrescu, Cristian & Rose, Fred & Kato, Seiji. (2017). Clouds and the Earth'S Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) top-of-atmosphere (TOA) edition-4.0 data product. Journal of Climate. 31. 10.1175/JCLI-D-17-0208.1.

Maddy, E. S. and C. D. Barnet, 2008, Vertical Resolution Estimates in Version 5 of AIRS Operational Retrievals, IEEE Trans. Geosci. Remote Sens., 46 (8), 2375-2384, 2008.

Miloshevich, L. M., Vömel, H., Whiteman, D. N., Lesht, B. M., Schmidlin, F. J., and Russo, F., Absolute accuracy of water vapor measurements from six operational radiosonde types launched during AWEX-G and implications for AIRS validation, J. Geophys. Res., 111(d9), doi:10.1029/2005JD006083, 2006.

Milstein, A. B., and Blackwell, W. J. (2016), Neural network temperature and moisture retrieval algorithm validation for AIRS/AMSU and CrIS/ATMS, *J. Geophys. Res. Atmos.*, 121, 1414–1430, doi:10.1002/2015JD024008.

OMI Team, Ozone Monitoring Instrument (OMI) Data User's Guide, 2012. (Available at https://acdisc.gesdisc.eosdis.nasa.gov/data/s4pa/Aura_OMI_Level2G/OMTO3G.003/doc/README.OMI_DUG.pdf)

Peterson, C. A., Yue, Q., Kahn, B. H., Fetzer, E. J., & Huang, X. (2020). Evaluation of AIRS Cloud Phase Classification over the Arctic Ocean Against Combined CloudSat-CALIPSO Observations, submitted to *Journal of* Applied Meterology and Climatology, under revision.

Pougatchev N.S., 2008, Validation of atmospheric sounders by correlative measurements, *Appl. Opt*. 47(26), pp. 4739-4748 (2008).

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding: The-ory and Practice, World Scientific Publishing, Singapore, 2000.

Rodgers C.D., and B.J. Conner, 2003, Intercomparison of remote sounding instruments, *J. Geophys. Res.* 108(D3), 4116 (2003).

Sotiropoulou, G., M. Tjernström, J. Sedlar, P. Achtert, B. J. Brooks, I. M. Brooks, P. O. G. Persson, J. Prytherch, D. J. Salisbury, M. D. Shupe, P. E. Johnston and D. Wolfe, 2016: Atmospheric Conditions during the Arctic Clouds in Summer Experiment (ACSE): Contrasting Open Water and Sea Ice Surfaces during Melt and Freeze-Up Seasons, J. Clim., 29, 8721-8744, doi:10.1175/JCLI-D-16-0211.1.

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

Tjernström, M. and coauthors, 2014: The Arctic Summer Cloud Ocean Study (ASCOS): overview and experimental design, Atmos. Chem. Phys., 14, 2823-2869, doi:10.5194/acp-14-2823-2014.

Thrastarson, H. Th. & Teixeira, J. (2020). A Humidity-Driven Prediction System for Influenza Outbreaks. Manuscript in preparation.

Thrastarson, H. Th., ed. (2020), AIRS/AMSU/HSB Level 2 Product User Guide – Product Version 7. Jet Propulsion Laboratory, California Institute of Technology.

Tian, B., ed. (2020), AIRS/AMSU/HSB Level 3 Product User Guide – Product Version 7. Jet Propulsion Laboratory, California Institute of Technology.

Veefkind, J.P., J.F. de Haan, E.J. Brinksma, M. Kroon and P.F. Levelt, "Total Ozone from the Ozone Monitoring Instrument (OMI) Using the DOAS Technique," IEEE Trans. Geo. Rem. Sens., 44(5), 1239-1244, 2006.

Wong, S., E. J. Fetzer, M. Schreier, G. Manipon, E. F. Fishbein, B. H, Kahn, Q. Yue, and F. W. Irion (2015), Cloud-induced uncertainties in AIRS and ECMWF temperature and specific humidity, *J. Geophys. Res.*, 120, doi:10.1002/2014JD022440.

Yue, Q., E. J. Fetzer, B. H. Kahn, S. Wong, G. Manipon, A. Guilaume, and B. Wilson, 2013: Cloud-state-dependent sampling in AIRS observations based on CloudSat cloud classification, *J. of Climate*, 26, 8357–8377, doi: http://dx.doi.org/10.1175/JCLI-D-13-00065.1.

Yue, Q., and B. Lambrigtsen, ed., 2017, AIRS V6 Test Report Supplement: Performance of AIRS+AMSU vs. AIRS-only Retrievals, available at: https://docserver.gesdisc.eosdis.nasa.gov/repository/Mission/AIRS/3.3_ScienceDataProductDocumentation/3.3.5_ProductQuality/V6_Test_Report_Supplement_Performance_of_AIRS+AMSU_vs_AIRS-Only_Retrievals.pdf.