Test Report of Performance of Version 1 RAMSESII-SNPP Retrievals

Prepared by Jacola Roman, Tao Wang, Qing Yue, and Sun Wong

Contributors: Bjorn Lambrigtsen, Mathias Schreier, Ruth Monarrez, Evan Manning, Eric Fetzer

Jet Propulsion Laboratory, California Institute of Technology

RAMSES II Product Version v01_41_00 (algorithm version V1.1.0)

This assessment was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004).

June, 2021

Goddard Earth Sciences Data and Information Services Center (GES DISC)
http://disc.gsfc.nasa.gov
NASA Goddard Space Flight Center
Code 610.2
Greenbelt, MD 20771 USA
© 2021. All rights reserved.

Table of Contents

1. Technical Summary

This report assesses the general quality of the Level 2 (L2) products from the Version 2 RAMSES II (**R**etrieval **A**lgorithm for **M**icrowave **S**ounders in **E**arth **S**cience) retrieval system for ATMS (**A**dvanced **T**echnology **M**icrowave **S**ounder) onboard the SNPP (**S**uomi-**N**ational **P**olar-orbiting **P**artnership) satellite (referred to in this report as *RAMSESII-SNPP*). The specific algorithm version tested is version 01-41-00, herein referred to as V1. This first release of RAMSES II is an early research version intended to solicit user comments. It has a number of known issues but will soon be followed by a revised and improved version.

RAMSES II is a single-footprint MW-Only retrieval system for ATMS, which applies an optimal estimation algorithm performing "all-sky" retrievals based solely on Microwave (MW) sounder data independent of cloud coverage. RAMSES II uses background information from MERRA2 (**M**odern-**E**ra **R**etrospective Analysis for **R**esearch and **A**pplications version 2) along with a radiative transfer algorithm and a non-linear solver to achieve an optimized retrieval of profiles of atmospheric states. Uncertainty estimates and quality flags are based on optimization and convergence between observed and calculated brightness temperatures.

ATMS onboard the SNPP satellite is a cross-track scanner with 22 channels in spectral bands from 23 GHz to 183 GHz, providing MW measurements in both clear and cloudy conditions. ATMS provides both temperature (T) soundings (between the surface and the upper stratosphere) and humidity (q) soundings (between the surface and upper troposphere). ATMS has better sampling and two more channels than its predecessor AMSU (**A**dvanced **M**icrowave **S**ounding **U**nit). SNPP is a polar orbiting satellite launched in 2011 that crosses the equator about 14 times daily with a 13:30 local time ascending node, providing full global coverage twice a day. Acting as the bridge between NASA's Earth Observing System and the Joint Polar Satellite System (JPSS), SNPP is the first in a series of five next generation U.S. weather satellites of the JPSS, and is a result of a partnership between NOAA, NASA and the Department of Defense (DoD).

This report aims to test the performance of V1 RAMSESII-SNPP L2 retrievals. The data can be downloaded from the Goddard Distributed Active Archive Center (GDAAC). The testing report aims to summarize the general quality of several core products of the RAMSESII-SNPP L2 retrievals by the following analyses: 1) evaluating the retrieval yield for atmospheric T and H2O profiles; 2) quantifying the biases and root mean square errors (RMSE) of T and H2O retrievals by pixel-scale comparisons with collocated radiosonde measurements from the Integrated Global Radiosonde Archive (IGRA, over land), Alfred Wegener Institute (AWI) Polarstern laboratory (over ocean), as well as the 5th generation global atmospheric reanalysis from the European Centre for Medium-Range Weather Forecasts (ERA5, both land and ocean); 3) evaluation of surface and 2-meter temperature, tropopause temperature and pressure, and total column water vapor retrievals using pixel-scale collocations to ERA5; and 4) comparisons of daily mean gridded L2 retrievals of these core variables with daily fields from MERRA2.

2. Data and Methodology

2.1 RAMSESII-SNPP V1 Retrieval Products

This test version of RAMSES II, applied to ATMS onboard SNPP, is v01_41_00, herein V1. More details on the retrieval algorithm can be found in the RAMSES II Retrieval Algorithm for Microwave Sounders in Earth Science The NASA ATMS Retrieval System Level 2 Algorithm Theoretical Basis Document (ATBD).

RAMSES II final retrievals of T and H2O profiles are reported at 100 and 66 pressure levels, respectively. The H2O profiles are specific humidity defined as mass fraction of water vapor in moist air in units of kg/kg. In addition, both T and H2O profiles are also reported on 72 sigma pressure levels. The MERRAII temperature and water vapor first guesses are not reported in this version.

For both T and H2O retrievals, RAMSES II L2 provides associated quality control indicators with names "_qc" appended to each variable, where QC=0 indicates the best retrievals that meet the accuracy requirements, QC=1 indicates good retrievals, and QC=2 indicates the use of such data is not recommended. The quality flag for RAMSES II is based on a threshold for the convergence using a two-step process. For more information on this two-step process, please see the ATBD.

T and H2O analyses are aggregated by spatial coverage as well as the associated surface type, surface temperature and total column water vapor from the RAMSESII-SNPP L2 product. This will provide insight on the quality of RAMSESII-SNPP retrievals in different climate regimes and physical conditions. Single level variables such as 2-meter air temperature, surface skin temperature, and sea ice fraction are also retrieved. These variables along with total column water vapor (TCWV) will be examined and assessed. 2-meter air temperature, surface skin temperature, and TCWV have corresponding quality control flags that follow the same format as those for the T and H2O profiles. Surface temperature does have a first guess from MERRA2, but other regression estimates, such as rain rates, 2-meter surface temperature, and ice/snow coverage, are solely estimated on channel differences. All variables presented here are reported on the 135 Along-track Field Of View (FOV) and 96 Cross-Track FOV spatial dimension. The detailed file format and definitions of variables in the RAMSESII-SNPP files can be found in the product user guide [*Monarrez et al.* 2021].

2.2 IGRA radiosondes

In order to evaluate the performance of RAMSESII-SNPP retrieval products, 8 months (Jan/Apr/Jul/Oct of 2013 and 2015) of RAMSESII-SNPP retrieved L2 T and q profiles were collocated to radiosonde observations of the version 2 Integrated Global Radiosonde Archive (IGRA) [*Durre and Yin*, 2008] using a nearest neighbor approach with temporal and spatial tolerances of 3 hours and 50 km, respectively. The v2 IGRA consists of radiosonde and pilot balloon observations at over 2,700 globally distributed stations - mostly over land. The time period availability of the IGRA archive varies from station to station. Previously, the IGRA dataset has been used to investigate the cloud-induced uncertainties in AIRS version 6 data [*Wong et al.*, 2015].

2.3 AWI radiosondes

RAMSESII-SNPP retrievals were also collocated to radiosondes available from the Alfred Wegener Institute (AWI) Polarstern laboratory [*König-Langlo and Marx*, 1997; *Driemel et al.*, 2016]. The AWI Polarstern has been operated since 1982, and it was the first research ship ever to across the Atlantic Ocean in a meridional section twice a year. It is therefore a useful validation dataset for data-sparse regions such as over the oceans and at the poles.

2.4 ERA5 reanalysis

The ERA5 is the 5th generation global atmospheric reanalysis from the European Centre for Medium-Range Weather Forecasts (ECMWF), replacing the ERA-Interim reanalysis which stopped being produced on August 31st, 2019. Produced using a 4D-Var data assimilation, the ERA5 has a much higher spatial and temporal resolution than its predecessor, the ERA-Interim. In addition, newly reprocessed datasets along with recent instruments have been assimilated into the ERA5 that could not be ingested into the ERA-Interim [*Hennermann* 2020].

The ERA5 has output at various temporal resolutions. This study made use of hourly output of temperature and specific humidity profiles with 37 pressure levels, two-meter temperature, surface skin temperature, and sea ice all on a latitude-longitude grid of 0.25°x0.25°. These gridded products were extracted from the Copernicus Climate Change Service (C3S) Climate Data Store over the period January 1st-5th 2013 and July 1st-5th 2013 [*Copernicus* 2017].

Profiles and single levels products were collocated to RAMSESII-SNPP using a nearest neighbor approach with the following requirements:

- 1 hour temporal differences
- 1° latitude/longitude radius

Temperature and specific humidity profiles that were collocated to RAMSESII-SNPP were vertically log/log interpolated to the RAMSESII-SNPP pressure levels. The matchups were aggregated by month (January or July 2013), quality control flags (QC), and spatial coverage (global, zonal, and maps).

2.5 Testing metric

The bias and root mean square error (RMSE) will be used to assess the differences of RAMSESII-SNPP retrievals to reference datasets. RAMSESII-SNPP L2 profiles of T and H_2O will be compared to IGRA and AWI (section 3.2) and to ERA5 (section 3.3). For temperature, the bias is calculated using the following formula:

$$T_{bias} = mean(T_{retrieval} - T_{reference}) \qquad (1)$$

Similarly, the RMSE for temperature is defined as:

$$T_{RMSE} = \sqrt{mean((T_{retrieval} - T_{reference})^2)} \qquad (2)$$

Because of the large variability of H_2O from the surface to the upper troposphere, in most cases the H_2O bias is calculated relative to the mean state of the reference using the following equation:

$$H_2O_{bias} = \frac{mean(H_2O_{retrieval} - H_2O_{reference})}{mean(H_2O_{reference})} \times 100 \qquad (3)$$

Similarly, the RMSE will be normalized for specific humidity using the following formula:

$$H_2O_{RMSE} = \frac{\sqrt{mean((H_2O_{retrieval} - H_2O_{reference})^2)}}{mean(H_2O_{reference})} \times 100 \qquad (4)$$

Both temperature and H_2O will be conditioned by the quality control flags (QC) and other variables. Global maps at selected pressure levels, global means, as well as zonal means will be presented to highlight spatial differences. Similar definitions of bias and RMSE are used for other retrieved variables, please see details discussed in following sections.

2.6 Skill Score of The Final Retrieval Against Reference Datasets

A skill score test was calculated to assess whether the RAMSESII-SNPP retrieval outperforms AIRS v7 – when comparing to the common reference datasets from IGRA or ERA5. Skill score is a forecasting metric that allows a user to test whether one forecast has more skill than another [*Murphy*, 1988; *WMO*, 2012]. The formula for the mean squared error skill score is as follows:

$$Skill\ Score = 1 - \frac{RMSE_{RAMSESII}}{RMSE_{AIRS\ v7}} = 1 - \frac{\sqrt{mean((T_{RAMSESII} - T_{reference})^2)}}{\sqrt{mean((T_{AIRS\ v7} - T_{reference})^2)}} \qquad (5)$$

The skill score for H2O uses the same formula. Please note that this metric depends highly on what is considered the reference, which in this case is either IGRA or ERA5.

If the skill score is greater than 0, RAMSESII-SNPP has more skill than AIRS v7 (the RMSE of the RAMSESII-SNPP is smaller than the RMSE of AIRSv7); and vice versa. The value of the skill score represents how much the RMSE was reduced or increased. For example, a skill score value of 0.25 means RAMSESII-SNPP reduced the RMSE observed in AIRS v7 by 25%.

3. Results

RAMSESII-SNPP retrieval yield (section 3.1) will be presented below. In section 3.2, L2 RAMSESII-SNPP profiles collocated to sondes will be analyzed while those collocated to ERA5 will be discussed in section 3.3. Results of L2 single level variables collocated to the ERA5 will be discussed in section 3.4 while gridded L3 results compared to MERRA2 will be discussed in section 3.5.

3.1 L2 RAMSESII-SNPP T and H2O Profile Retrieval Yield

Fig. 1 compares the zonal mean vertical cross section of RAMSESII-SNPP retrieval yields (%) for QC=0 (best retrievals), QC=1 (good retrievals), QC=0 or 1 for scientific retrievals when combining the two. The percentage of QC = 2 cases (not suggested for scientific use) is also shown. The top row is for temperature and the bottom row is for specific humidity (H2O). The yield is the percentage of retrievals that are flagged by QC divided by the total observation count; therefore, the yields of QC=0 or 1 plus the yields of QC=2 equal 100%.

RAMSESII-SNPP retained about 30% of retrievals with the highest quality control (QC =0). Less than 20% of all retrievals were flagged as poor (QC = 2) between 60°N and 60°S. More retrievals were flagged as poor at the poles, particularly in the southern hemisphere where upwards of 60% of all retrievals were rejected. Yields were constant throughout the profile.

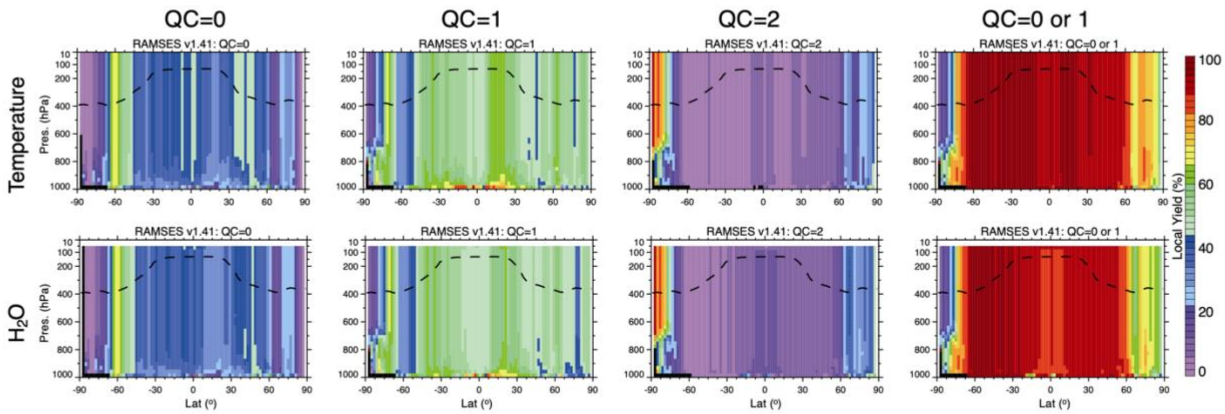

Figure 1. Zonal mean vertical cross section of RAMSESII-SNPP retrieval yields as a percentage for retrievals with quality flag QC=0 (best), QC=1 (good), QC=2 (not for scientific use), and QC=0 or 1 (good for scientific use). The data depicts one day (Jan. 1, 2015) of global results. The dashed lines are the local tropopause using the WMO definition.

3.2 Evaluation of L2 T and H2O Profiles and TCWV Using Collocated In-situ Radiosonde Observations

Two sets of independent in-situ field campaign datasets are used to calculate the biases and RMSE of T and H2O profiles as well as TCWV retrieved by RAMSESII-SNPP using the definitions in eqns. (1-4). They are radiosonde measurements obtained during the IGRA (predominantly over land) and AWI (over ocean) experiments.

3.2.1 Comparison with IGRA Data

Fig. 2a shows one month (January 2015) of RAMSESII-SNPP profiles (red dots, total 15212) collocated to all IGRA stations (black dots) with a temporal tolerance of ±3 hours and a spatial tolerance of 50 km radius from the sonde sites. The most frequent collocations were centered in Europe, where the launching times of the radiosondes were more frequently within the tolerance from the SNPP passing times. Fig. 2b-c shows the histograms of the distances (black) and time differences (blue) of the collocated matchups. Most of the collocated profiles were within 20 km radius. Over Europe, the IGRA launching time matched the SNPP passing time, with the sample size largest at time differences of 0 min and slowly decaying towards the tolerance limits. Over the tropical broad band all collocated profiles have time differences equally distributed within the ±3 hour window.

Fig. 3 shows the T bias (left), RMSE (middle), and skill score with respect to AIRS v7 (right) by geographical region (rows) for RAMSESII-SNPP (black), AIRS v7 IR Only (blue) and AIRS v7 IR+MW (red). *Note that AIRS v7 retrievals were independently collocated to the IGRA sondes by following the QC flags and satellite sampling of AIRS products***.** Regardless of the geographical aggregation, there was a strong and persistent cold bias in RAMSESII-SNPP throughout the troposphere that ranged from -1 Kelvin near the surface to more than -6 Kelvin near the tropopause. This bias was substantially larger than the bias in the AIRS v7 retrievals. Similarly, the RAMSESII-SNPP RMSE was quite large compared to AIRS v7 ranging from 2 Kelvin near the surface to more than 8 Kelvin near the tropopause. Furthermore, the vertical structure of the RMSE and bias were different between RAMSESII-SNPP and AIRS v7. In particular, the RMSE grew with altitude up to the tropopause for RAMSESII-SNPP while the RMSE was largest near the surface in AIRS v7. RAMSESII-SNPP had a smaller RMSE than AIRS v7 near the surface over Europe and the northern high latitudes. This resulted in a positive skill suggesting that the RAMSESII-SNPP final retrieval has skill near the surface in these regions. However, whether that skill comes from the retrieval itself or the first guess remains to be determined.

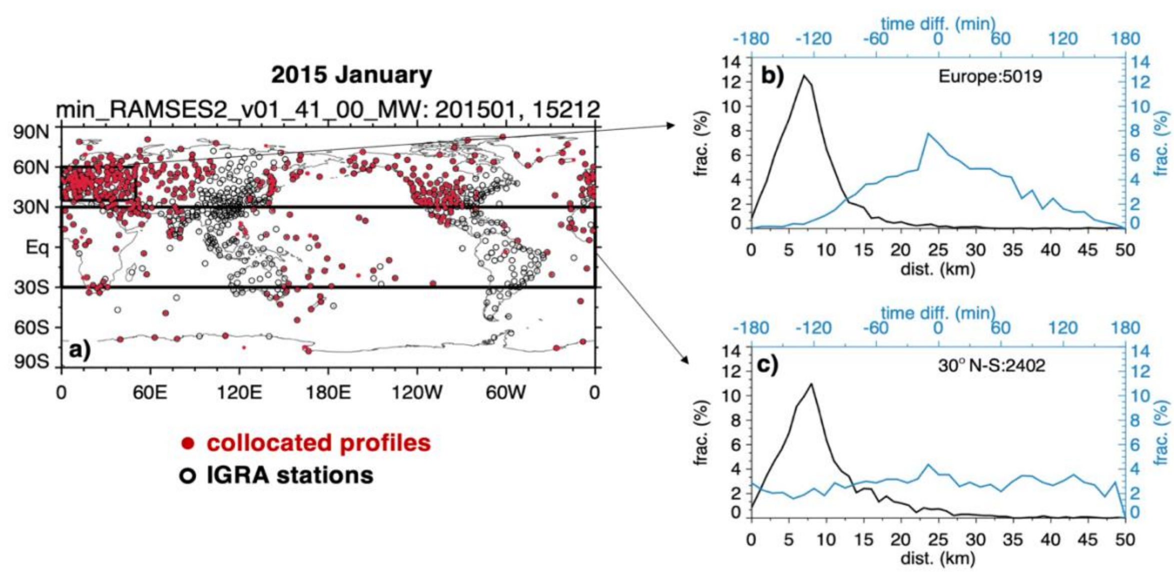

Figure 2. (a) The IGRA stations (black dots) and successfully collocated RAMSESII-SNPP profiles (red dots, total 15212 profiles) within 3-hour, 50-km searching criteria for January 2015; (b) histogram of the distance (lower x-axis, black) and time differences in minutes (upper x-axis, blue) from the collocated records over the Europe; (c) over the tropics (30° N–30° S).

Fig. 4 is the same as Fig. 3 but for H2O profiles in units of g/kg and Fig. 5 shows these results in relative differences in units of %. AIRS v7 and RAMSESII-SNPP had comparable bias values in the lower part of the troposphere over Europe, the northern high latitudes and the southern mid-latitudes. Generally, RAMSESII-SNPP had a dry bias with a magnitude larger than that in AIRS v7. RMSE values were more comparable than for temperature between AIRS v7 and RAMSESII-SNPP. However, RAMSESII-SNPP still had a larger RMSE than AIRS v7. Notable exceptions included Europe between the surface and 700 hPa, where the skill score reached 0.25 suggesting RAMSESII-SNPP reduced the RMSE by 25%, and over the tropics and mid-latitudes near the surface, where a positive skill score was observed.

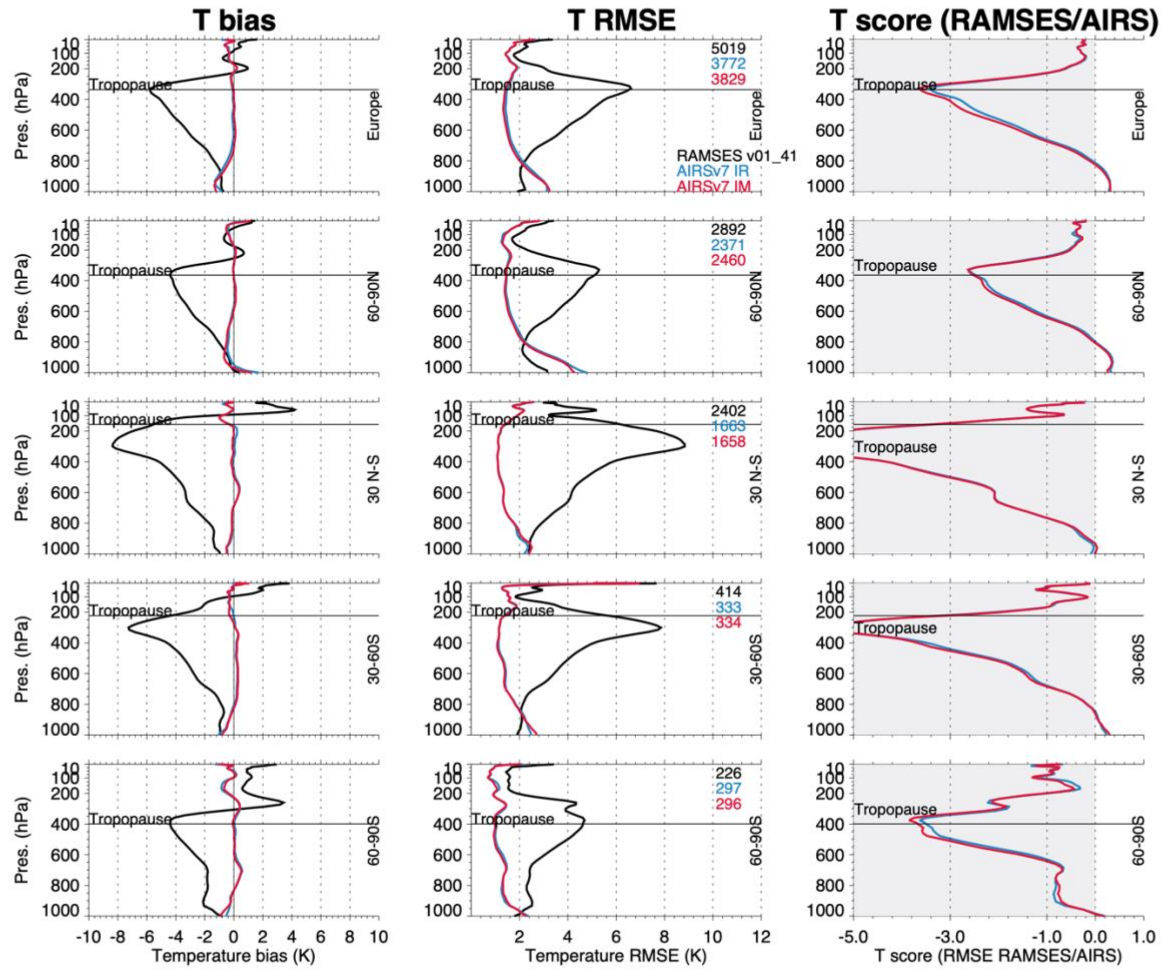

Figure 3. Compared to IGRA radiosondes, RAMSES (black) and AIRS v7 (blue for IR and red for IR+MW) temperature bias (left column) and RMSE (middle column) in Europe (top row), 60-90ºN, the tropics (30º N–S), 30-60ºS, and 60-90ºS (bottom row). The right column shows the retrieval score defined as $1 - RMSE_{RAMSES}/RMSE_{AIRS}$, so that negative values (shaded in gray) indicate RAMSESII-SNPP retrieval has larger RMSE compared to AIRS; and positive values indicating better performances of RAMSES compared to AIRS.

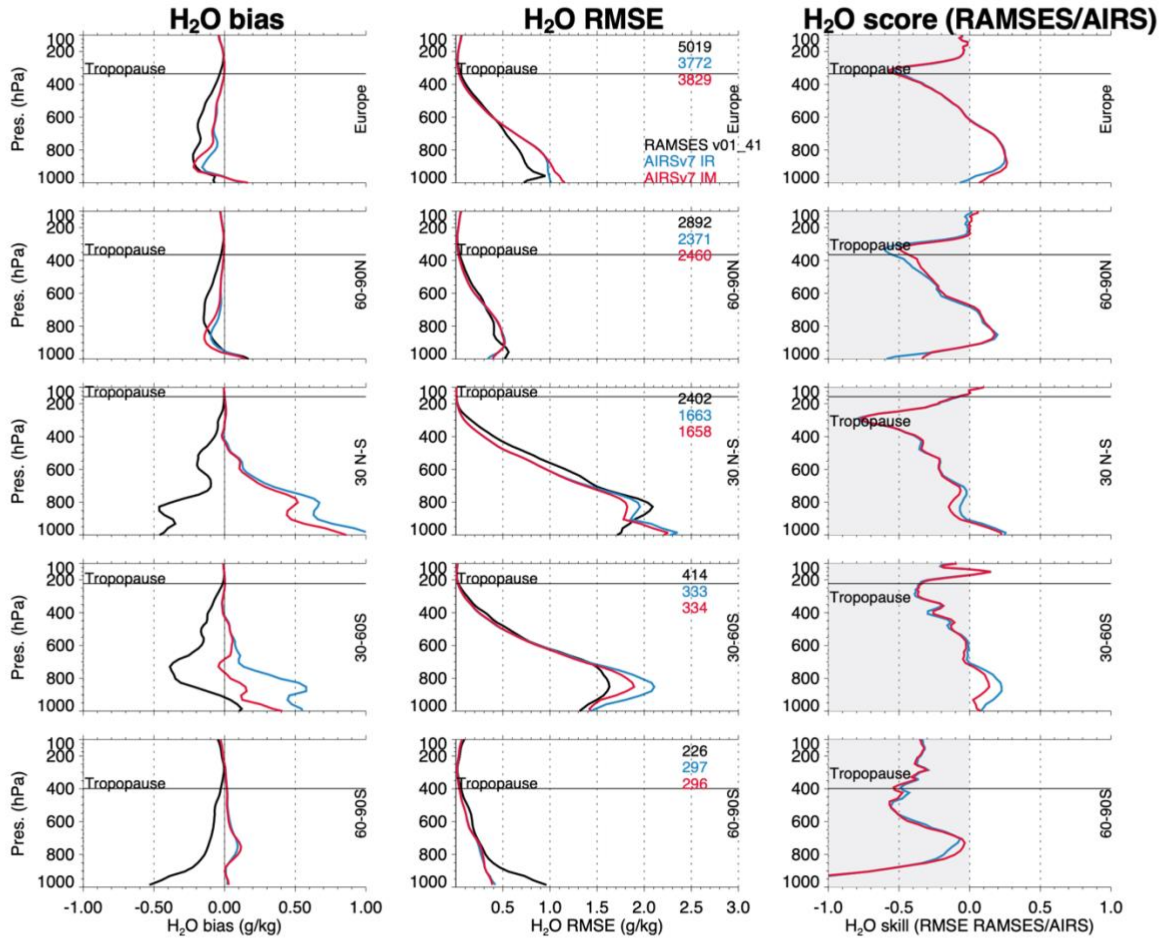

Figure 4. Compared to IGRA radiosondes, RAMSESII-SNPP (black) and AIRS v7 (blue for IR and red for IR+MW) H_2O bias (left column) and RMSE (middle column) in Europe (top row), 60-90^{o}N, the tropics (30^o N–S), 30-60^{o}S, and 60-90^{o}S (bottom row). The right column shows the retrieval score defined as 1 – $RMSE_{RAMSES}/RMSE_{AIRS}$, so that negative values (shaded in gray) indicate RAMSESII-SNPP retrieval has larger RMSE compared to AIRS; and positive values indicating better performances of RAMSES compared to AIRS.

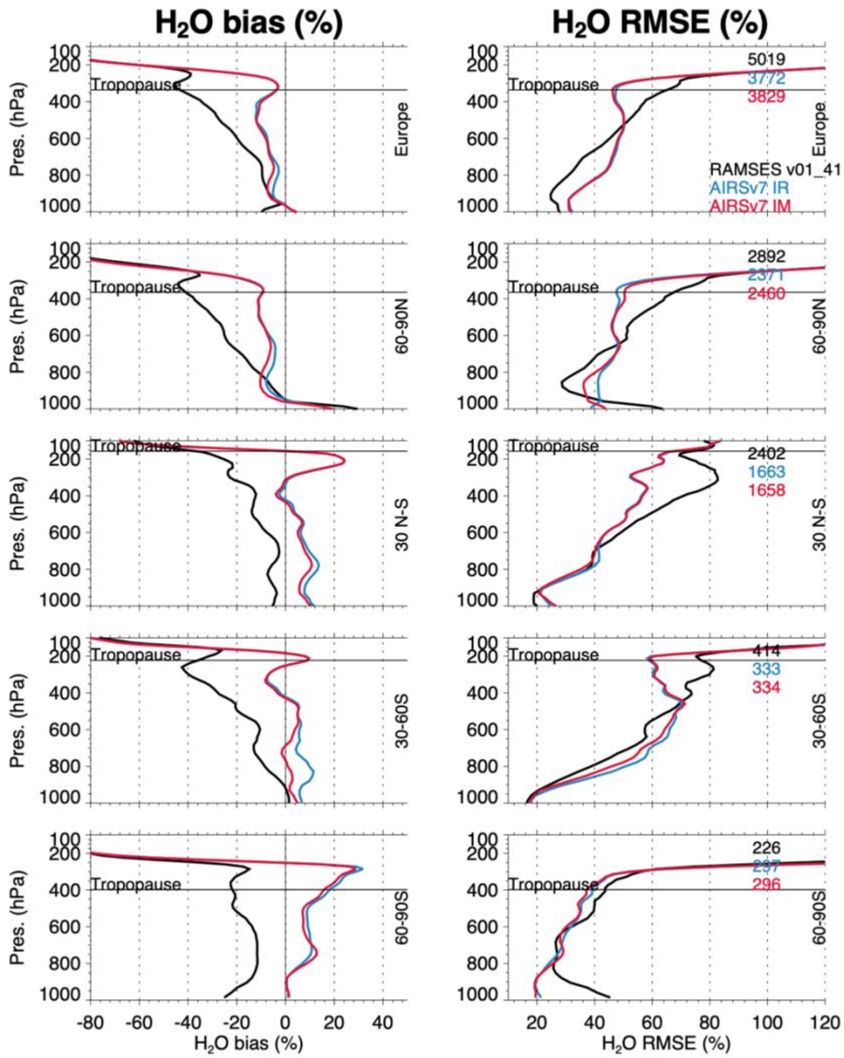

Figure 5. Same as Fig. 4, but with bias and RMSE reported in percentages compared to IGRA.

3.2.2 Comparison with AWI Data

In this section, RAMSESII-SNPP is compared to radiosondes from the AWI Polarstern ship. Fig. 6a shows all collocated RAMSESII-SNPP profiles (color-coded by surface types, total 76) compared to all radiosondes from AWI Polarstern 2013-2019 (black circles), with a temporal tolerance of ±3 hours and a spatial tolerance of 50 km radius. Fig. 6b shows the histograms of the distances (black) and time differences (blue) of collocated datasets. Most of the collocated profiles were within 10 km radius.

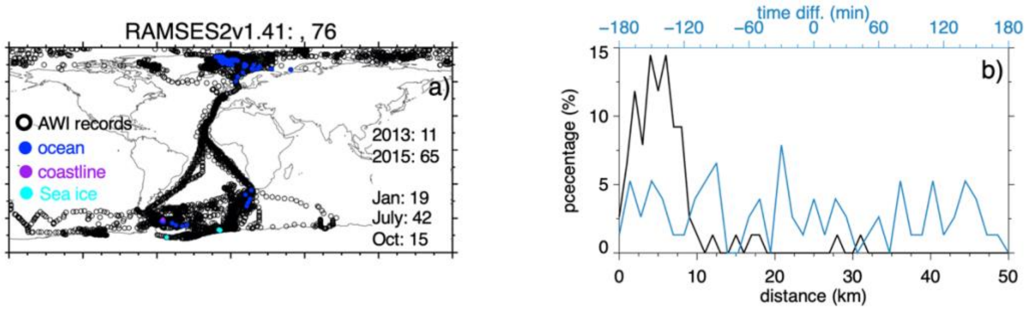

Figure 6. (a) The AWI Polarstern records 2013-2019 (black dots) and successfully collocated RAMSESII-SNPP profiles (color-coded by surface types, total 76 profiles) within 3-hour, 50-km searching criteria for Jan/Apr/Jul/Oct of 2013 and 2015; (b) histogram of the distance (lower x-axis, black) and time differences in minutes (upper x-axis, blue) from the collocated records.

Fig. 7 shows the T and H_2O biases and RMSE compared to radiosondes from the AWI Polarstern cruises for RAMSESII-SNPP (black), AIRS v7 IR Only (blue) and AIRS v7 IR+MW (red). Similar to the IGRA results, RAMSESII-SNPP had a consistent cold bias throughout the troposphere that peaked in the tropopause. The RAMSESII-SNPP temperature RMSE was substantially larger than AIRS v7, particularly in the upper troposphere. However, near the surface, the RAMSESII-SNPP RMSE is smaller than the RMSE observed in AIRS v7 IR Only.

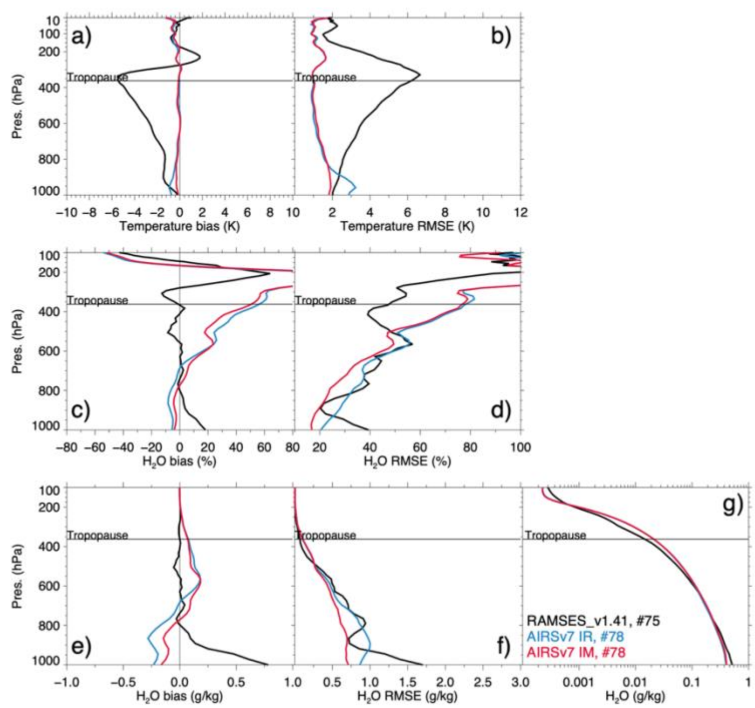

Figure 7. Compared to collocated radiosondes from the AWI Polarstern cruises during Jan/Apr/Jul/Oct of 2013 and 2015, temperature (top row) and H_2O (middle row in relative percentage and bottom row in g/kg) bias and RMSE for RAMSES and ARIS v7 (IR in blue and IR+MW in red).

The RAMSESII-SNPP H2O bias was smaller than AIRS v7 in the mid-upper troposphere but predominantly larger (and wet) in the lower troposphere. This is different to the IGRA results in which the H_2O bias was consistently larger in RAMSESII-SNPP than AIRS v7. These differences could be due to IGRA sondes that are not good at measuring specific humidity in freezing conditions as well as geolocation differences (land for IGRA and ocean for AWI).

Fig. 8 compares the TCWV between collocated RAMSESII-SNPP retrievals and radiosondes from the AWI Polarstern cruises. Note that only limited data points over high latitude oceans are available after applying collocation, the results for which may not represent the total precipitable water retrieval performance in other regions. The total precipitable water was highly correlated between RAMSESII-SNPP and AWI. Generally, RAMSESII-SNPP underestimated the low total precipitable water values and overestimated high values.

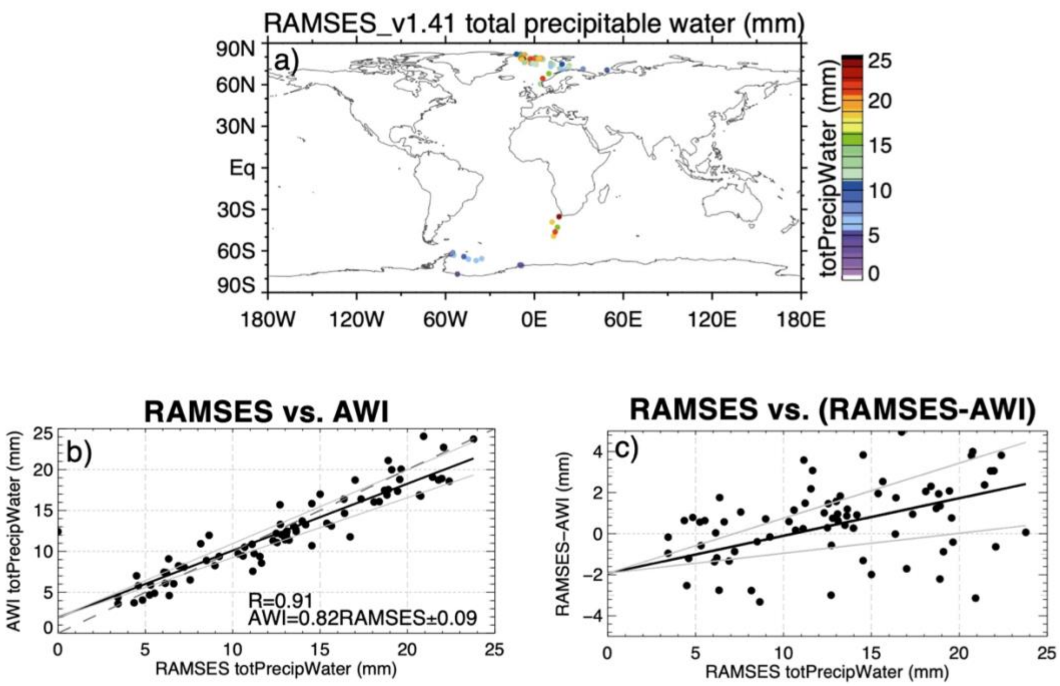

Figure 8. Compared to radiosondes from the AWI Polarstern cruises, the collocated RAMSES total precipitable water (panel a) and the comparisons to AWI total precipitable water (panel b) and to the differences of RAMSES-AWI (panel c).

3.3 Evaluation of L2 T and H_2O Profiles Using Collocated ERA5 Reanalysis

3.3.1 Global Mean Retrieval Performance

Results presented in this section highlight the global mean bias and RMSE, with respect to the ERA5, and the yield by quality control flag for temperature and specific humidity profiles. AIRS v7 IR+MW results will be presented to provide a comparison. The AIRS v7 IR+MW temperature and specific humidity profiles were independently collocated to ERA5; thus, the spatial and temporal matchup was slightly different compared to RAMSESII-SNPP. Furthermore, the QC was with respect to each individual retrieval system. Since AIRS v7 was not collocated to RAMSESII-SNPP the results presented here highlight the bias, RMSE, and yield one would expect if using the retrievals systems independently.

Fig. 9 shows the global mean bias (left), RMSE (middle), and yield (right) for temperature aggregated by month (rows) and QC (color) for RAMSESII-SNPP (solid lines) and AIRS v7 IR+MW (dashed). Regardless of the quality control, RAMSESII-SNPP had a strong and persistent cold bias throughout the troposphere ranging from -1.5 Kelvin near the surface to -7.5 kelvin at 300 hPa. Although the bias was slightly smaller when QC = 2 for January, this result may be spurious and does not necessarily mean the QC flags failed. The AIRS v7 bias stayed close to the 0 line throughout the profile, further emphasizing the strong and persistent cold bias in RAMSESII-SNPP. The RAMSESII-SNPP RMSE was 2.5 Kelvin near the surface and more than 7 Kelvin around 300 hPa. The RAMSESII-SNPP RMSE was largest when QC = 2 throughout the troposphere and smallest when QC = 0 or QC = 0 or 1. The vertical structures of AIRS v7 and RAMSESII-SNPP RMSE were quite different. The AIRS v7 RMSE ranged from 1 to 2.5 kelvin and was typically larger near the surface. RAMSESII-SNPP, however, had a smaller RMSE near the surface that grew with altitude up to 300 hPa. The AIRS v7 and RAMSESII-SNPP yield was smallest when QC = 2 suggesting most retrievals were retained. The RAMSESII-SNPP high quality flag had a yield of 25-30% while the relaxed QC had a yield 50% larger. The AIRS v7 yield structure was different to RAMSESII-SNPP. In particular, the AIRS v7 yield for QC = 0 or QC = 0 or 1 grew steadily with altitude and more retrievals were rejected near the surface while RAMSESII-SNPP had a steady yield throughout the profile. In addition, AIRS v7 threw out more retrievals near the surface than RAMSESII-SNPP.

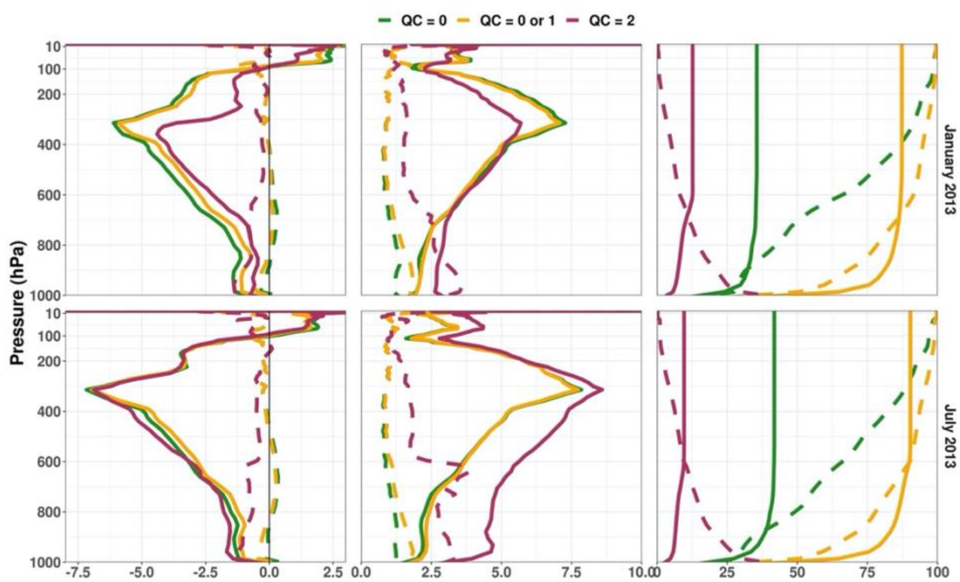

Figure 9. Global mean temperature bias (left), RMSE (middle), and yield (right) for RAMSESII-SNPP (solid line) and AIRS v7 (dashed line) aggregated by QC (QC = 0 is green, QC = 0 or 1 is orange, and QC = 2 is red)) and month (January 2013 top and July 2013 bottom).

Fig. 10 is the same as Fig. 9 but for Specific Humidity (as a percent). RAMSESII-SNPP had a consistent dry bias through the troposphere that ranged from -2% near the surface up to -40% near 300 hPa. The difference between AIRS v7 and RAMSESII-SNPP was not as significant for specific humidity as it was for temperature. However, AIRS v7 had a smaller bias by about 5-20% compared to RAMSESII-SNPP. The AIRS v7 bias was predominantly dry, like RAMSESII-SNPP, except in the lower to mid-troposphere. The RAMSESII-SNPP RMSE grew from 5% near the surface to more than 100% near the tropopause. In addition, the RMSE was smallest when QC = 0 and largest when QC = 2. Unlike temperature, the AIRS v7 RMSE vertical structure was similar to the RAMSESII-SNPP RMSE vertical structure. The specific humidity yield results were similar to the temperature results. RAMSESII-SNPP was more conservative with QC=0 (smaller yield than AIRS v7). In addition, the RAMSESII-SNPP yield did not fluctuate with pressure while the AIRS v7 yield for QC = 0 and QC = 0 or 1 grew with altitude. AIRS v7 flags fewer retrievals as poor (QC = 2) than RAMSESII-SNPP except near the surface.

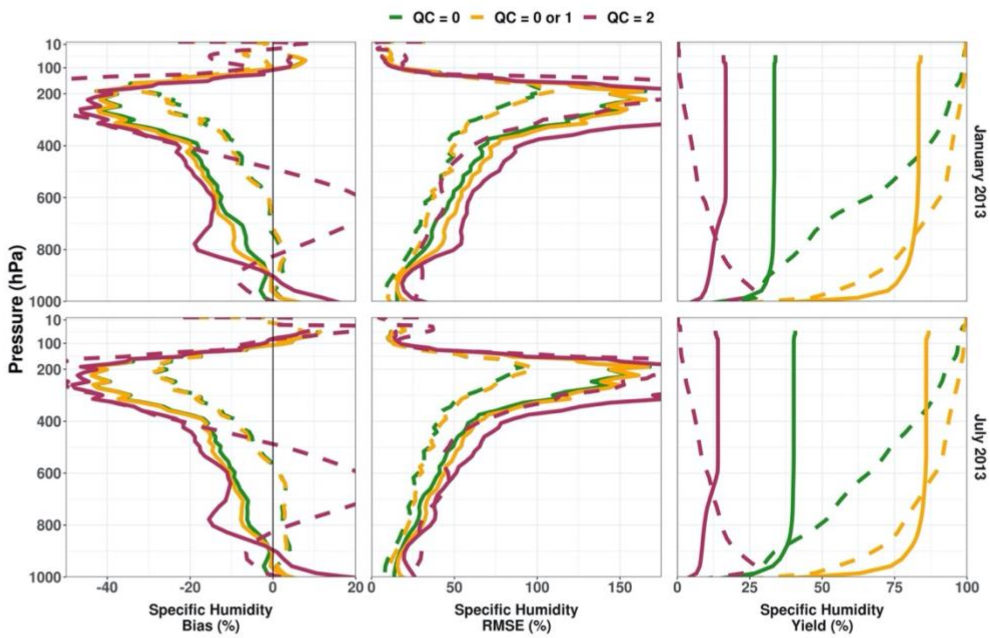

Figure 10. Global mean specific humidity bias (left), RMSE (middle), and yield (right) for RAMSESII-SNPP (solid line) and AIRS v7 (dashed line) aggregated by QC (QC = 0 is green, QC = 0 or 1 is orange, and QC = 2 is red)) and month (January 2013 top and July 2013 bottom).

3.3.2 Zonal Mean Retrieval Performance

Results presented here will be aggregated by 5° latitude bins (x-axis) along with month (rows) and QC (columns). In addition to bias, RMSE, and yield, the skill score will be presented. Since AIRS v7 was not collocated to RAMSESII-SNPP, the skill score represents the general retrieval skill if using each system independently and does not necessarily represent the system's ability to retrieve under the same conditions.

Fig. 11 shows the temperature bias for each retrieval system (rows) aggregated by QC (columns), and month (rows). Similar to the global results, RAMSESII-SNPP had a consistent cold bias throughout the troposphere regardless of the latitude zone. Between 800 and 400 hPa the RAMSESII-SNPP bias exceeded 3 Kelvin while closer to the surface the bias ranged from 0.5 to 1.5 Kelvin. There was a warm bias throughout most of the stratosphere that was strongest in the tropics (exceeding 3 Kelvin). There was little difference between the QC = 0 and QC = 0 or 1 results for RAMSESII-SNPP. The AIRS v7 bias was generally smaller in magnitude. Furthermore, the direction of the bias was different between 800 – 400 hPa. Indeed, AIRS v7 had a slight warm bias of 0.5 Kelvin throughout this pressure range while RAMSESII-SNPP had a cold bias. AIRS v7 quality control differences (QC = 0 vs QC = 0 or 1) occurred predominantly near the surface over the high latitudes. There were little seasonal differences in the bias structure for both RAMSESII-SNPP and AIRS v7. However, RAMSESII-SNPP had slight warm bias near the surface in the northern high latitudes during the month of July that was cold in January.

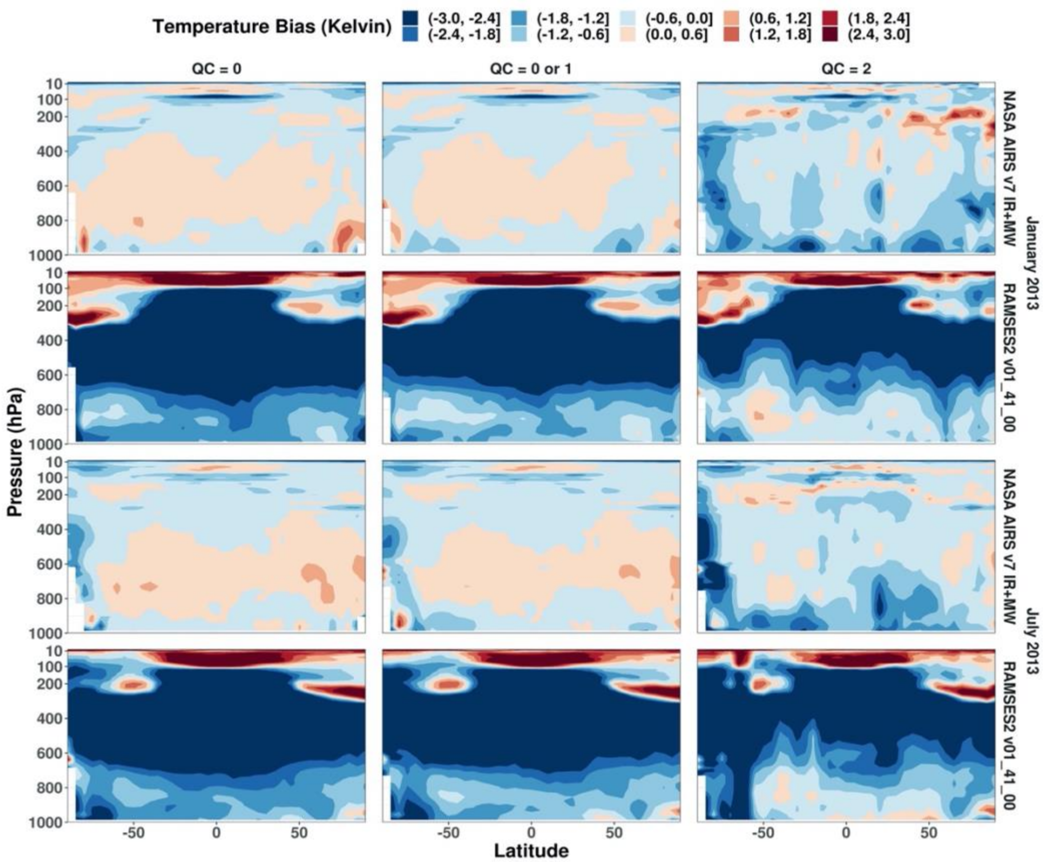

Figure 11. Zonal mean temperature bias for RAMSESII-SNPP and AIRS v7 (rows) aggregated by QC (columns) and month (rows).

Fig. 12 Is the same as Fig. 11 but for RMSE. The AIRS v7 RMSE was small, usually less than 2 Kelvin, and there was little change between season. The RMSE was largest near the surface for AIRS v7. RAMSESII-SNPP had a different RMSE structure. The RMSE was largest, exceeding 8 Kelvin, around 200 hPa over the tropics and was smallest near the surface, regardless of latitude. For both AIRS v7 and RAMSESII-SNPP the RMSE was largest when QC = 2 suggesting the quality flags were appropriately throwing out poor retrievals.

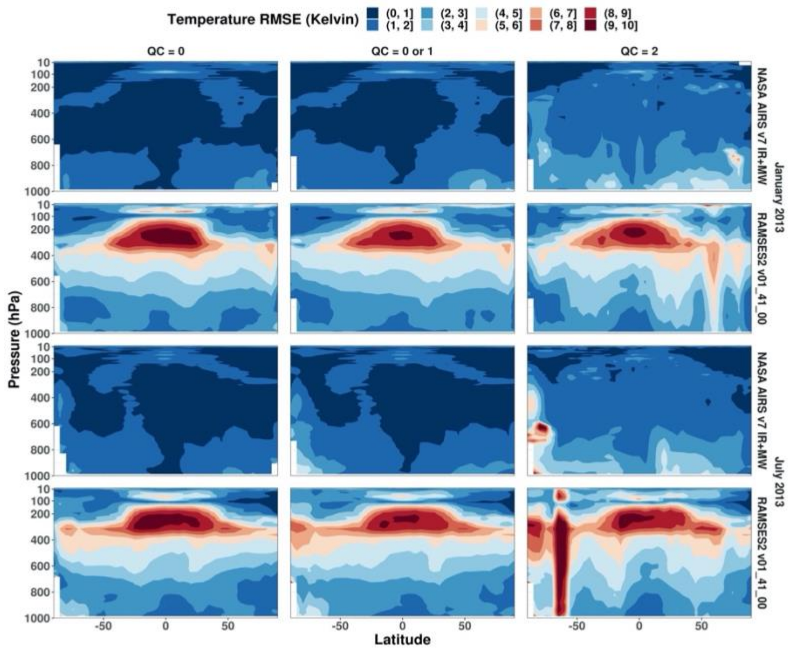

Figure 12. Same as Fig. 11 but for RMSE

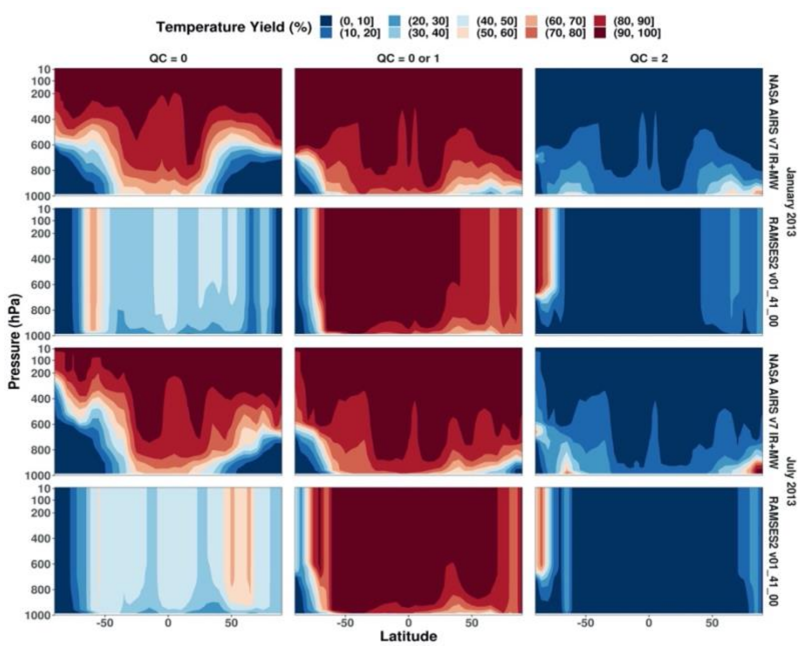

Figure 13. Same as Fig. 11 but for yield

Fig. 13 is the same as Fig. 11 but for yield. RAMSESII-SNPP had a larger yield in the summer polar region. When QC = 0, the RAMSESII-SNPP yield was 60-70% during January around 60-70^{o}S while the yield was about 10-20% around 60-70^{o}N. In July the yield dropped to 10-20% in the southern hemisphere and increased to 60-70% in the northern. A similar seasonal yield shift was observed when QC = 0 or 1. There were no seasonal yield differences in AIRS v7. AIRS v7 flagged more retrievals near the surface as poor in the mid-high latitudes than RAMSESII-SNPP while RAMSESII-SNPP flagged more retrievals as bad over the southern pole in both January and July.

Fig. 14 is the same as Fig. 11 but for Skill Score. A negative score (blue) means RAMSESII-SNPP increased the RMSE relative AIRS v7. ERA5 was considered the truth in this example. Throughout most of the profile, regardless of the latitude zone, the score was generally negative; that is RAMSESII-SNPP had little skill compared to AIRS v7. However, there were a few regions in which RAMSESII-SNPP had skill. In both January and July, RAMSESII-SNPP had skill near the surface in the mid-high latitudes. In particular, RAMSESII-SNPP was able to reduce the RMSE in AIRS v7 by 20-40% between 50^{o}N and 70^{o}N in January. RAMSESII-SNPP also had some skill near the tropopause in the tropics. When QC = 2 the skill score was positive for RAMSESII-SNPP. However, the interpretation of this result is difficult. In particular, the positive score means that the RMSE for bad retrievals was smaller in RAMSESII-SNPP than AIRS v7. Although it is possible that RAMSESII-SNPP is throwing away 'good' profiles when QC = 2, it is also possible that the RMSE was simply smaller in RAMSESII-SNPP because RAMSESII-SNPP throws away far fewer profiles (e.g. the sample size was smaller in RAMSESII-SNPP than AIRS v7 when QC = 2 near the surface). Interpretation of the skill score results when QC = 2 should be done with care. Furthermore, since AIRS v7 and RAMSESII-SNPP were not collocated the skill comparison presented here comes from different samples which may impact the results.

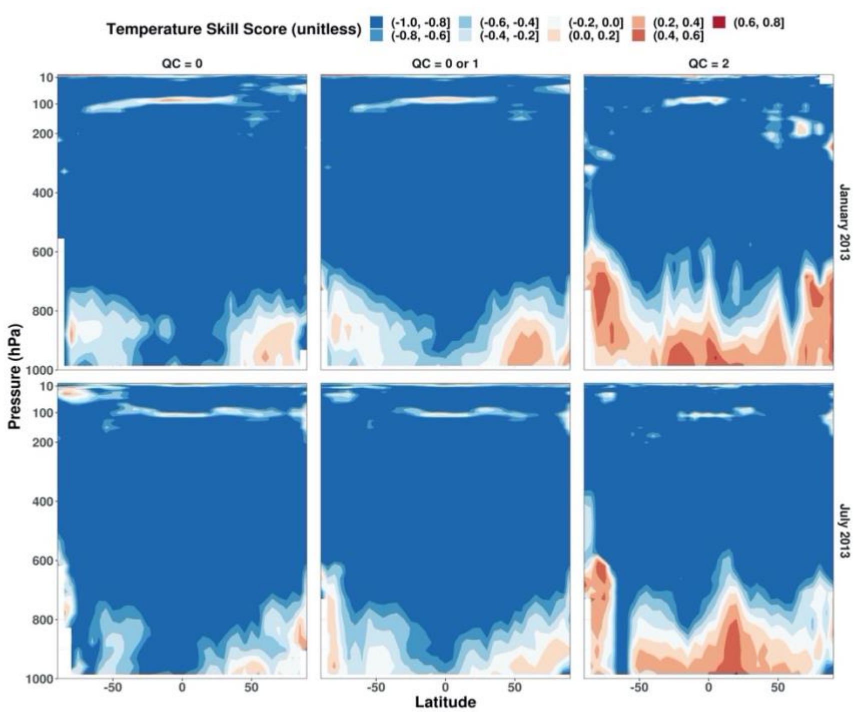

Figure 14. Same as Fig. 11 but for Skill Score

Fig. 15 shows the specific humidity bias for each retrieval system (rows) aggregated by QC (columns), and month (rows). RAMSESII-SNPP had a dry bias throughout most of the troposphere. However, near the mid-high latitudes there was a slight wet bias between the surface and 800 hPa in both January and July. The was little difference in the bias for RAMSESII-SNPP and AIRS v7 when using QC = 0 or QC = 0 or 1. Unlike RAMSESII-SNPP, the AIRS v7 bias was predominantly wet between the surface and 800 hPa, regardless of latitude, and then dry between 800 and 300 hPa. The magnitude of the dry bias was smaller in AIRS v7 than RAMSESII-SNPP. The magnitude of the bias when QC = 2 was generally larger for both AIRS v7 and RAMSESII-SNPP suggesting the quality flags were throwing out poor retrievals.

Fig. 16 is the same as Fig. 15 but for RMSE. The latitudinal RMSE structures for AIRS v7 and RAMSESII-SNPP were quite similar but the overall magnitude was larger in RAMSESII-SNPP. The RMSE was highest for both AIRS v7 and RAMSESII-SNPP near the tropopause, especially in the tropics and smallest near the surface. The RMSE was generally larger when QC = 2.

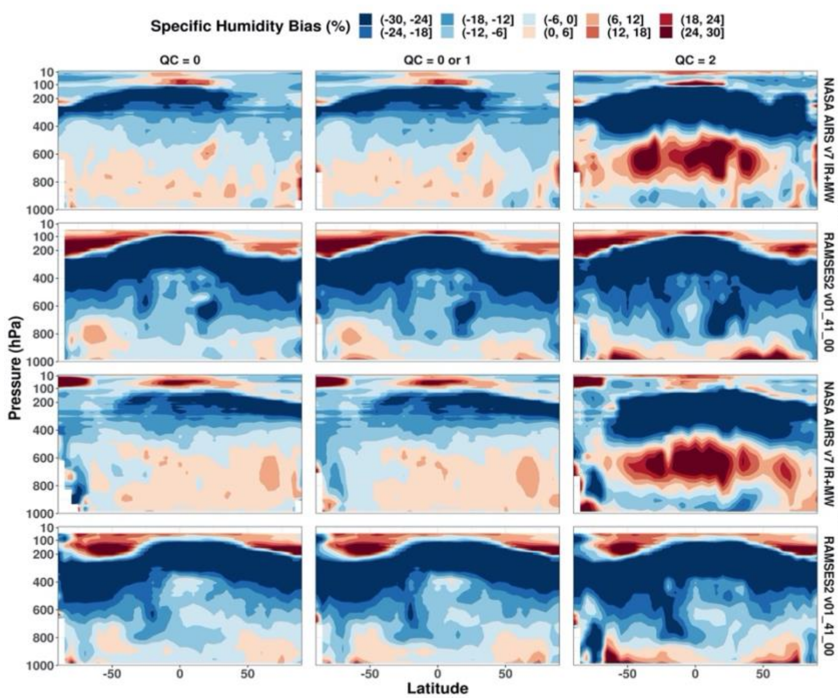

Figure 15. Zonal mean specific humidity bias for RAMSESII-SNPP and AIRS v7 (rows) aggregated by QC (columns) and month (rows).

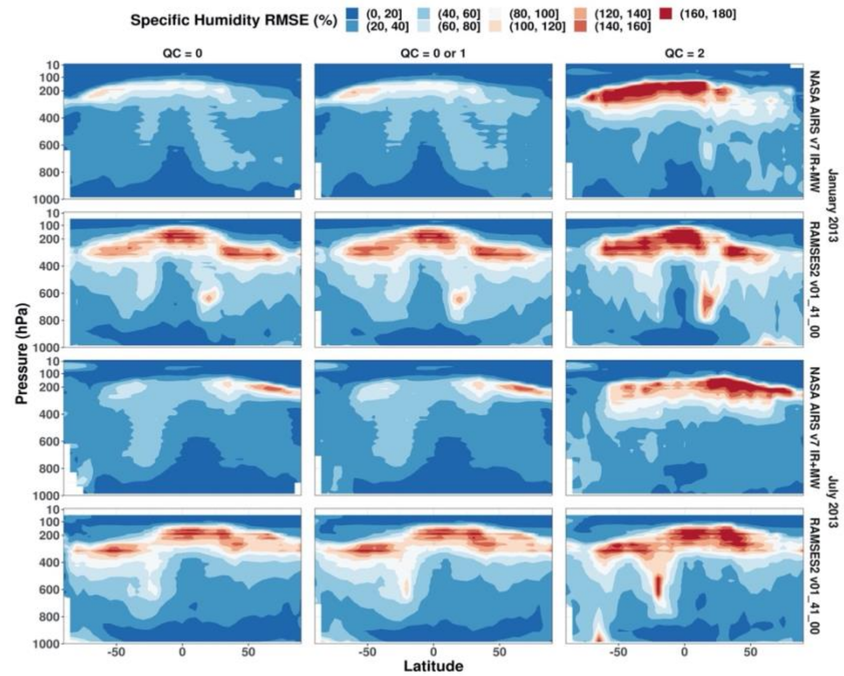

Figure 16. Same as Fig. 15 but for RMSE.

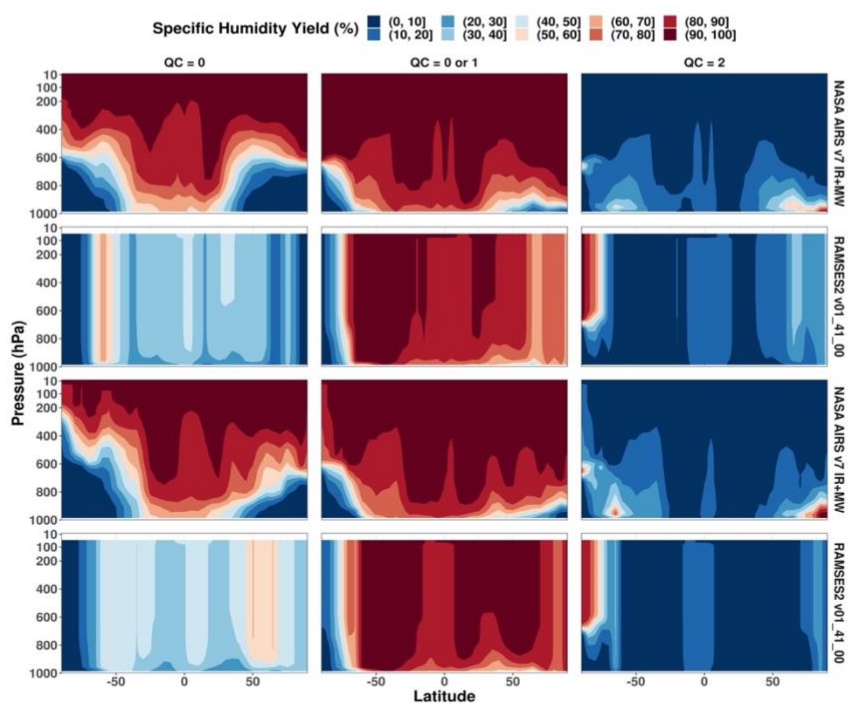

Figure 17. Same as Fig. 15 but for yield.

Fig. 17 shows the results for yield. The yield results were quite similar to what was discussed for temperature. However, the RAMSESII-SNPP yield was slightly higher for specific humidity than temperature when QC = 2. In particular, RAMSESII-SNPP flagged more retrievals of specific humidity as poor than temperature in the tropics. More retrievals in the northern mid to high latitudes were flagged as poor for specific humidity during January and the southern hemisphere during July (particularly between 800 hPa and the TOA).

Fig. 18 shows the zonal results for skill score. RAMSESII-SNPP had skill in the mid-high latitudes between 900 and 600 hPa as well as near the tropopause; reducing the AIRS v7 RMSE by more than 40%. RAMSESII-SNPP had no skill in the upper troposphere (600-300 hPa) in the mid-high latitudes. The skill appeared to be 'better' when looking at QC =2 compared to QC = 0 or QC =0 or 1. However, the interpretation is difficult and the emphasis should be made on the skill when using retrievals that passed quality control. Furthermore, the samples could be very different as AIRS v7 and RAMSESII-SNPP were not collocated for this analysis.

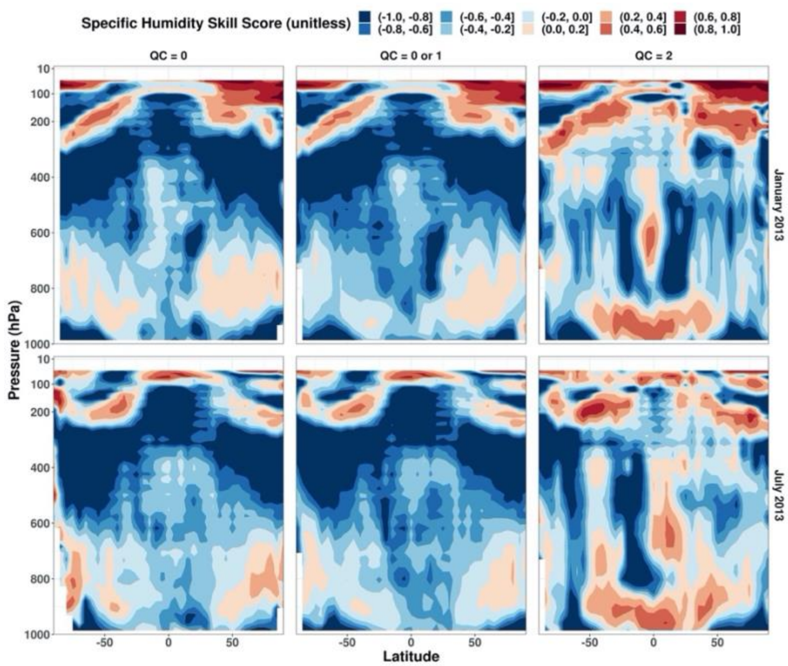

Figure 18. Same as Fig. 15 but for Skill Score.

3.3.3 Global Maps of T and H2O

Results presented here show the 5-day mean bias and RMSE as a global map for RAMSESII-SNPP using QC = 0 or 1 at 300, 500, 700 and 850 hPa.

Fig. 19 shows the global map of bias (left) and RMSE (right) for Temperature for QC = 0 or 1 by month (column) and pressure (row). Generally, the magnitude of the cold bias reduced from the Top of the Atmosphere (TOA) to the surface regardless of the geographical location. At 300 hPa, there was a warm bias over Antarctica during January. The tropics and mid-latitudes had a consistent strong cold bias exceeding 6 Kelvin for all longitudes. At 500 hPa there was a consistent global cold bias of about 4 Kelvin. At 700 hPa the cold bias was not as strong but still cold over the entire globe. Near the surface, at 850 hPa, there was still a persistent cold bias, however, there were regions of warm biases like off the coast of Brazil, over the ocean off the coast of Angola in western Africa, and off the coast of Chile. Most of the warm biases occurred over the ocean instead of land, except in January where regions of warm biases were observed over Canada and north eastern Russia. At 300 hPa, the RMSE was highest in the tropics regardless of month. At 500 hPa, the RMSE was about 3-5 Kelvin with some regions of high RMSE over land (e.g. Egypt in July and the middle east in January). The RMSE was smaller, relative to upper pressure levels, at 700 hPa and generally continued to decrease closer to the surface. Indeed, the overall RMSE decreased from the TOA to the surface which is consistent with the global mean results.

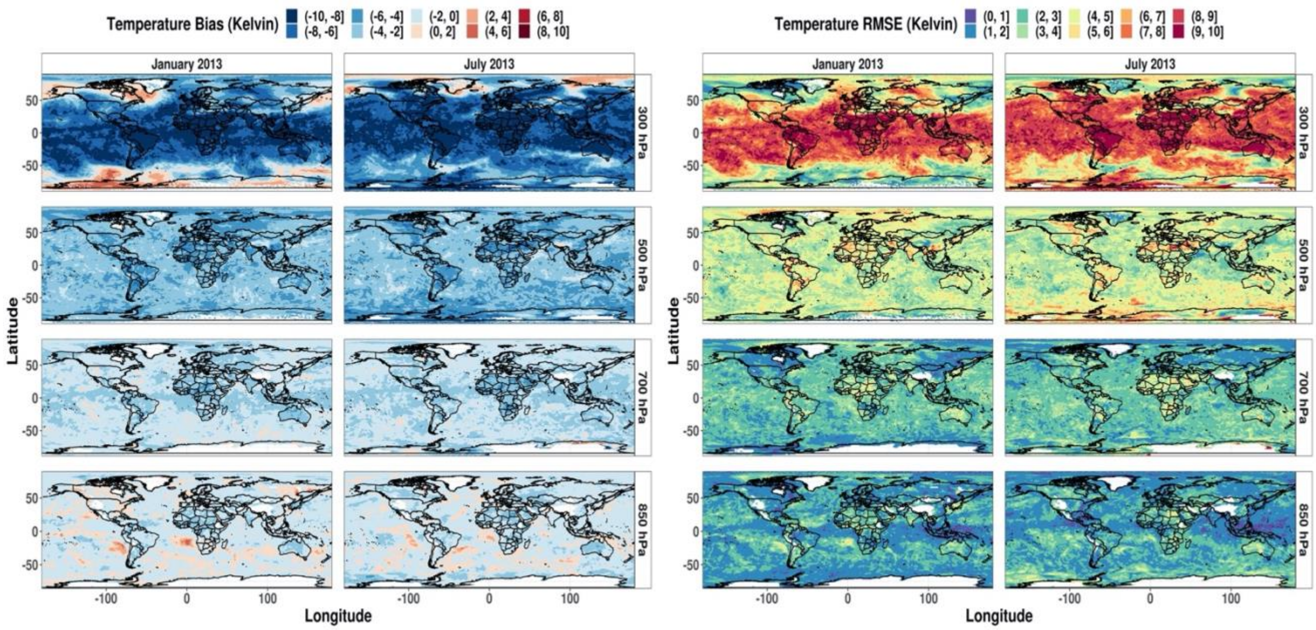

Figure 19. Global maps of bias (left) and RMSE (right) for temperature with QC = 0 or 1 aggregated by pressure level (rows) and month (columns).

Fig. 20 is the same as Fig. 19 but for specific humidity. Generally, the bias turned from dry to wet from the 300 hPa to the surface. At 300 hPa the bias was predominantly dry near the poles and wet in the Tropics. Similarly, at 500 hPa, the bias was predominantly dry except in the tropics where the bias was wet. At 700 hPa the magnitude of the bias was smaller with regions of extreme dry values over the oceans such as in the Atlantic in the Tropics. At 850 hPa the bias was predominantly wet. Like temperature, the RMSE decreased from the 300 hPa to the surface. At 300 hPa, the RMSE pattern was sporadic with no real patterns, however, the RMSE was somewhat higher over the ocean than land but exceptions included western Australia in January and Mongolia in July. At 500 hPa the RMSE ranged from 20-80%. Regions of high RMSE occurred over Australia while low RMSE was observed over India and Southeast Asia. At 700 hPa, areas of high RMSE occurred predominantly over the ocean. The RMSE was smallest at 850 hPa and was generally smaller over land than Ocean, especially in July.

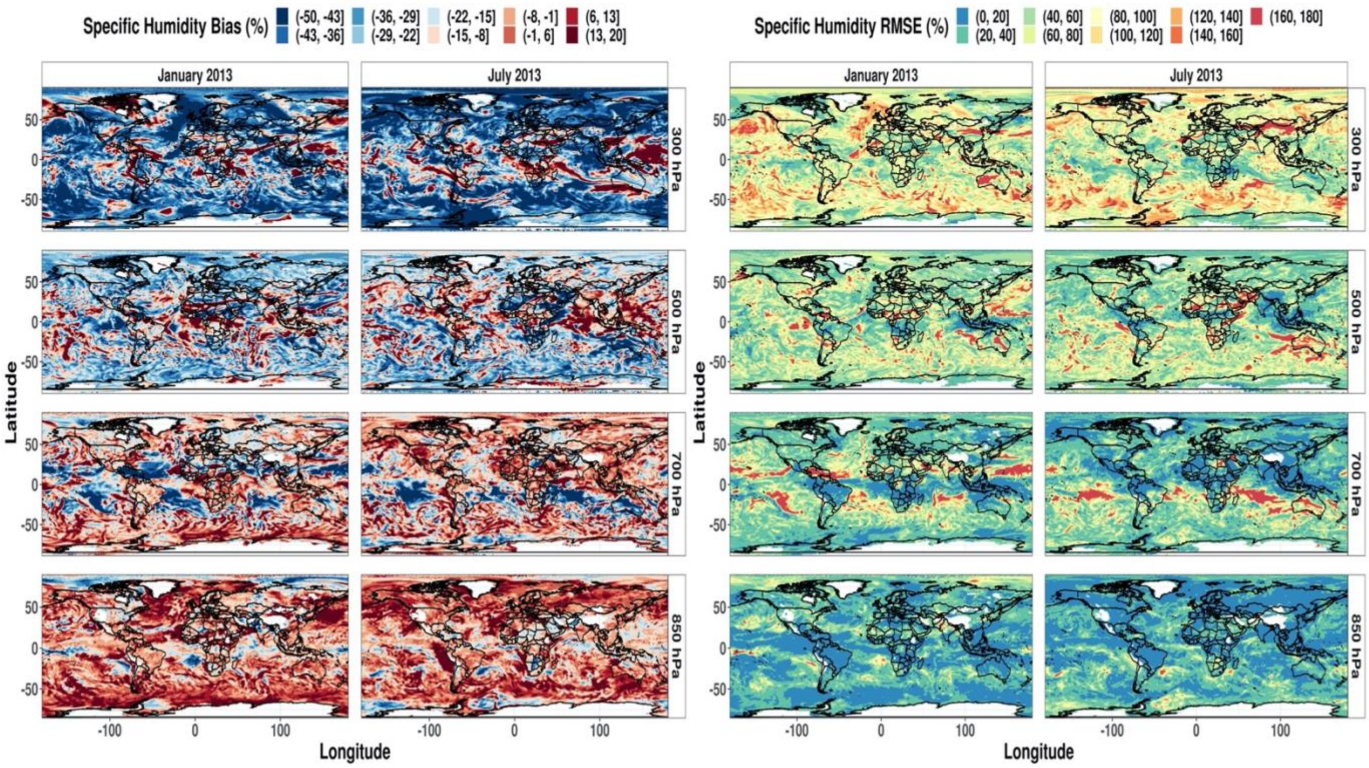

Figure 20. Same as Fig. 19 but for Specific Humidity.

3.3.4 T and H2O Retrieval Performance Dependence on Climate Regimes

Results presented here illustrate the dependence of RAMSESII-SNPP T and H2O retrieval performance on climate regimes. Specifically, the instantaneous retrievals are aggregated by scene types, including land versus ocean, surface temperature, and total column water vapor concentrations.

Fig. 21 shows the temperature bias, RMSE, and yield aggregated by land and ocean (color), QC (column), and month (row). 'Land' was defined as mw_land_frac = 1 and 'Ocean' was defined as mw_sea_frac = 1. During January, there was a clear difference between land and ocean. In particular, the cold bias was stronger over land than ocean and the RMSE was larger over land than ocean. This pattern was somewhat evident in July. The cold bias was about 0.5 to 1 Kelvin larger over land through most of the troposphere. The pattern was similar regardless of QC, except for QC = 2 where the RMSE was smaller over land than ocean. About 35% of retrievals over ocean pass the high-quality control flag (QC = 0) while about 60-65% pass for land.

Fig. 22 shows the temperature bias, RMSE, and yield aggregated by surface skin temperature (color), QC (column), and month (row). The purple color represents cooler scenes while the red color represents warmer scenes and the orange represents scenes in between (moderate). The bias was larger when the surface skin temperature was warmer throughout out the entire profile, especially in the upper troposphere where the temperature bias was almost 5 Kelvin larger when viewing warmer scenes compared to colder scenes. This pattern was true regardless of the month or quality control flag. The RMSE was smallest throughout the entire atmosphere when the scene was colder compared to warmer. Again, this was most evident in the upper troposphere where the difference was more than 5 Kelvin. Colder scenes had smaller yields when QC = 0 (~25% pass the strict flag) while 40% passed for the hotter scenes and about 50% for moderate scenes.

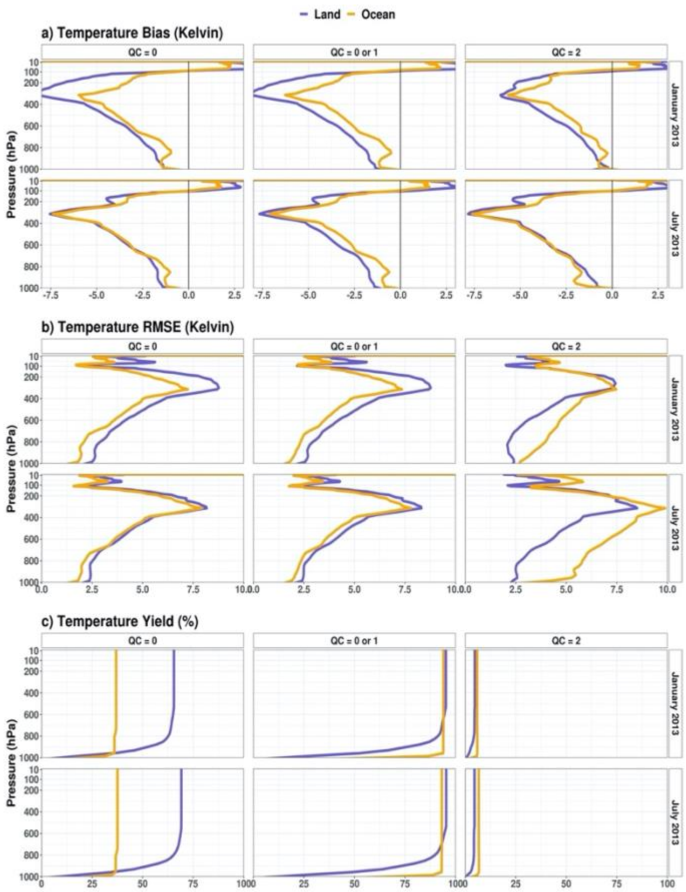

Figure 21. Global mean temperature bias (a), RMSE (b), and yield (c) aggregated by land versus ocean. Rows represent months and columns represent the quality control flag.

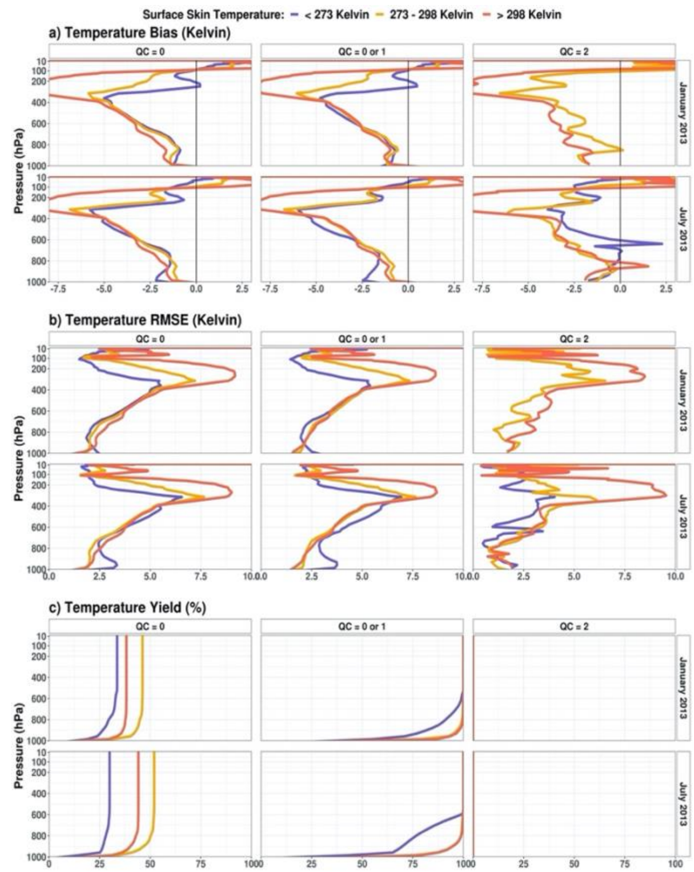

Figure 22. Global mean temperature bias (a), RMSE (b), and yield (c) aggregated by surface skin temperature. Rows represent months and columns represent the quality control flag.

Fig. 23 shows the temperature bias, RMSE, and yield aggregated by TCWV (color), QC (column), and month (row). Drier environments are in purple, wetter environments are in red, and those in-between are in orange. The strongest pattern occurred in the upper troposphere lower stratosphere (UTLS). Here, the cold bias was about 5 Kelvin smaller in drier environments as opposed to wetter environments. This held true regardless of QC and month, although the pattern was stronger in January than in July. RMSE was smallest in the UTLS when the TCWV was low compared to high. In July, the RMSE was larger in the lower-mid troposphere when the TCWV was low. The yield for QC = 0 was smallest when TCWV was low, followed by high and moderate.

Fig. 24 shows the specific humidity bias, RMSE, and yield aggregated by land versus ocean (color), QC (column) and month(row). With respect to bias, there was little difference when retrieving over land versus ocean. During July, the RMSE was slightly smaller when retrieving over land than ocean in the low-mid troposphere. 70% of all retrievals over land passed the highest quality control while only 30-40% over ocean passed the highest quality control. Overall, about 10% of all ocean retrievals were thrown out (QC = 2) while only about 5% were thrown out over land.

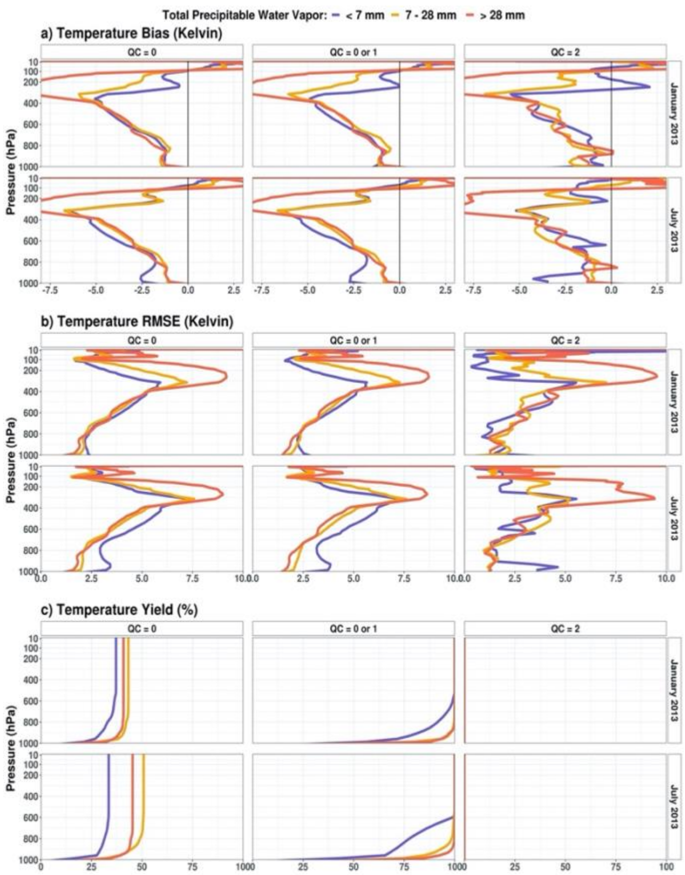

Figure 23. Global mean temperature bias (a), RMSE (b), and yield (c) aggregated by total column water vapor. Rows represent months and columns represent the quality control flag.

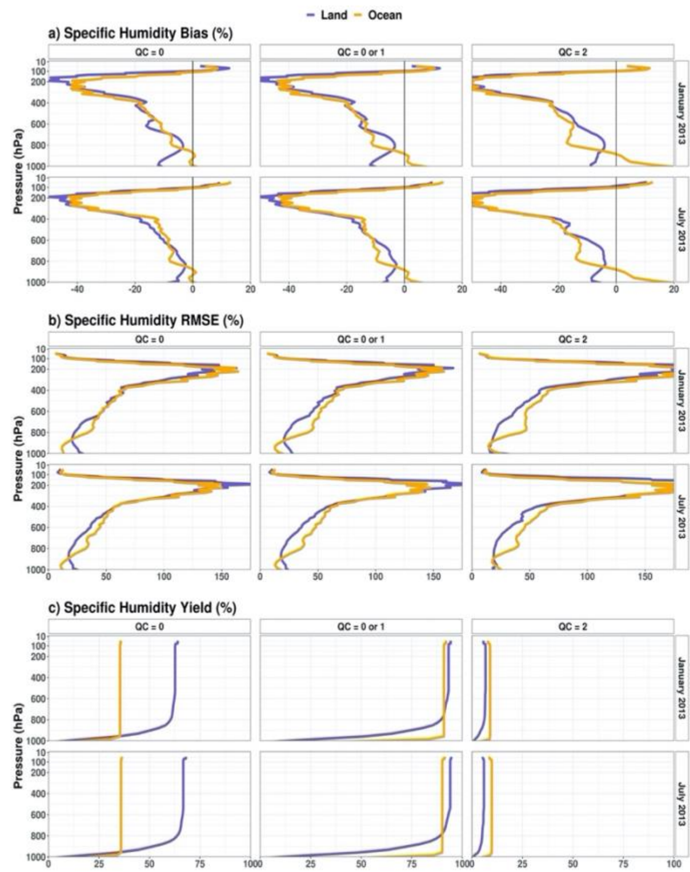

Figure 24. Global mean specific humidity bias (a), RMSE (b), and yield (c) aggregated by land and ocean. Rows represent months and columns represent the quality control flag.

Fig. 25 shows the specific humidity results aggregated by surface skin temperature. In the low to mid troposphere, the specific humidity bias was smallest when retrieving over warm scenes compared to colder scenes regardless of the QC and the month. In addition, the RMSE was largest for colder scenes compared to warmer scenes in the lower-mid troposphere and the pattern was stronger in July than January. About 25% of cold scenes passed the highest QC while about 35% of the warm scenes and 50% of the moderate passed.

Fig. 26 shows the specific humidity results aggregated by total column water vapor. The specific humidity bias between the surface and 300 hPa was smallest in moist environments compared to dry. This pattern was most evident in July. Similarly, between the surface and 300 hPa, the smallest RMSE occurred when retrieving over moist scenes instead of dry. Overall, 10-15% of retrievals were rejected in dry environments.

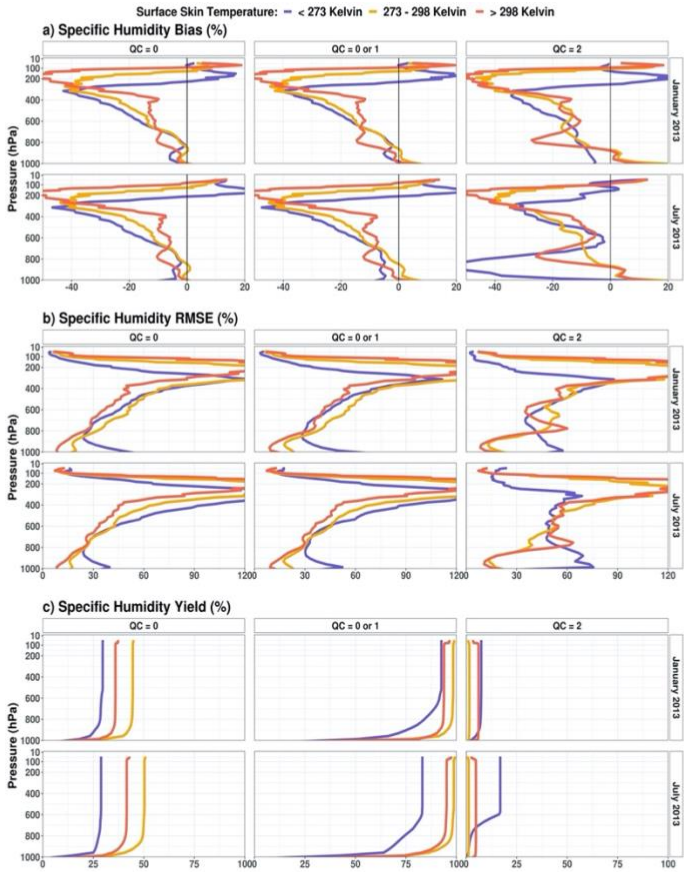

Figure 25. Global mean specific humidity bias (a), RMSE (b), and yield (c) aggregated by surface skin temperature. Rows represent months and columns represent the quality control flag.

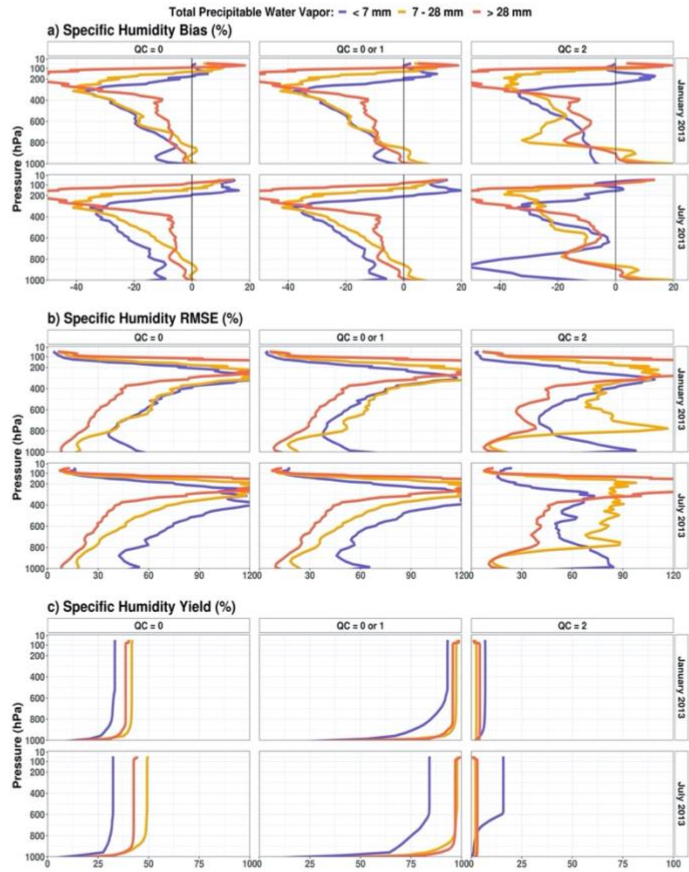

Figure 26. Global mean specific humidity bias (a), RMSE (b), and yield (c) aggregated by total column water vapor. Rows represent months and columns represent the quality control flag.

3.4 Evaluation of L2 2-Meter Air Temperature, Surface Skin Temperature and Sea Ice Using Collocated ERA5 Reanalysis

Results will be presented that compare the single level variables of 2-Meter Air Temperature, Surface Skin Temperature, and Sea Ice from RAMSESII-SNPP to collocated ERA5.

Fig. 27 shows the 5-day mean bias (RAMSESII-SNPP – ERA5) for 2-meter air temperature for January (left) and July (right) 2013. Generally, RAMSESII-SNPP had a warm bias over land and a cold bias over ocean. This was especially evident when QC = 0 and was true regardless of the month. Table 1 shows the corresponding coefficient of determination (R^2) value, global mean bias and global mean RMSE by month and quality control. RAMSESII-SNPP and ERA5 were highly correlated with R^2 values ranging from 0.97 to 0.98. On average, RAMSESII-SNPP was cooler than ERA5 by ~1 Kelvin. The global mean RMSE was about 2 to 2.5 Kelvin when QC = 0 or QC = 0 or 1 and was largest for QC = 2.

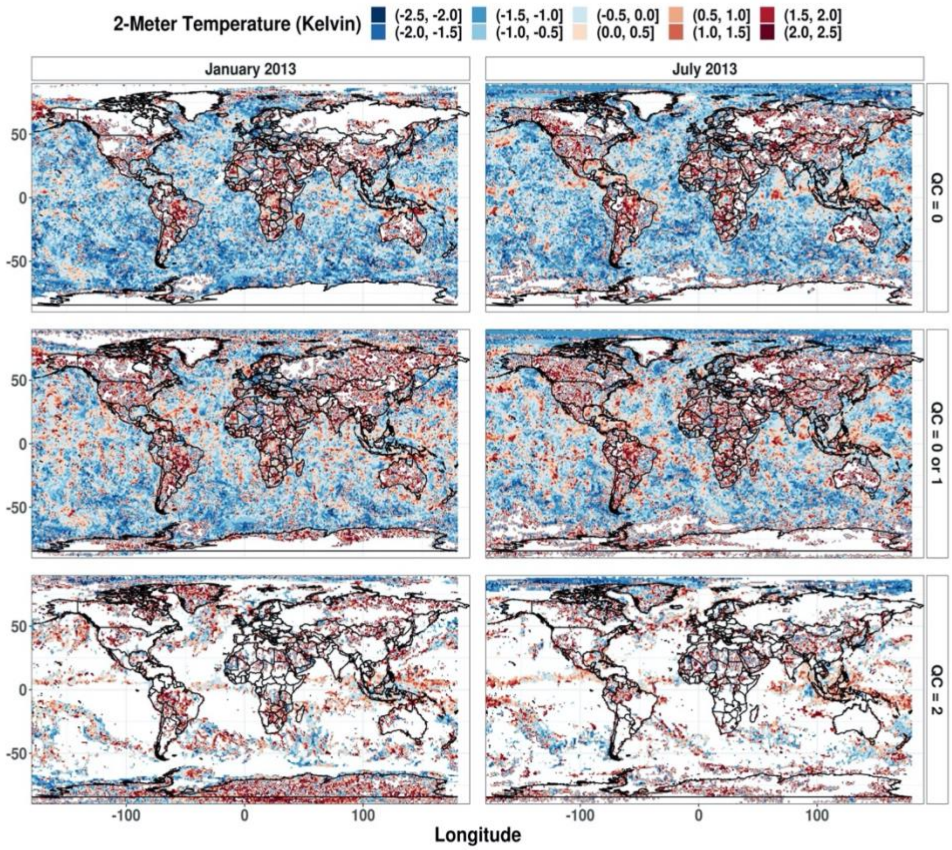

Figure 27. Global maps of 2-meter temperature bias (RAMSESII-SNPP - ERA5) aggregated by month (column) and QC (rows).

Fig. 28 is the same as Fig. 27 but for surface skin temperature. During January, RAMSESII-SNPP had a warm bias in surface skin temperature compared to ERA5, especially in the northern high latitudes. In July, RAMSESII-SNPP had a cold bias in the northern high latitudes. Generally, the bias was larger in magnitude over land than ocean. RAMSESII-SNPP and ERA5 surface skin temperatures were highly correlated, as observed in Table 1. The mean bias ranged from -0.04 to -0.5 Kelvin when QC = 0 or 0 or 1; smaller than for 2-meter temperature which may be due the use of MERRA2 as a first guess for surface temperature.

Table 1. Summary statistics for RAMSESII-SNPP single level variables compared to ERA5

Variable	Month	QC/Requirment	R^2	Global Mean Bias	Global Mean RMSE
2-Meter Temperature (Kelvin)	Jan	QC = 0	0.980	−0.948	2.259
		QC = 1	0.983	−0.636	2.331
		QC = 2	0.976	0.408	3.240
	July	QC = 0	0.971	−0.954	2.471
		QC = 1	0.983	−0.938	2.621
		QC = 2	0.983	−2.093	4.907
Surface Skin Temperature (Kelvin)	Jan	QC = 0	0.984	−0.279	1.998
		QC = 1	0.987	−0.054	2.109
		QC = 2	0.983	0.746	3.010
	July	QC = 0	0.981	−0.507	2.035
		QC = 1	0.987	−0.588	2.542
		QC = 2	0.989	−2.166	4.956
Sea Ice Fraction	January	with SI =0	0.855	-0.024	0.122
		SI > 0	0.394	−0.094	0.197
	July	with SI =0	0.857	−0.029	0.128
		SI > 0	0.494	−0.094	0.174

The RMSE was also smaller for surface skin temperature than 2-meter temperature, ranging from 2 to 5 Kelvin.

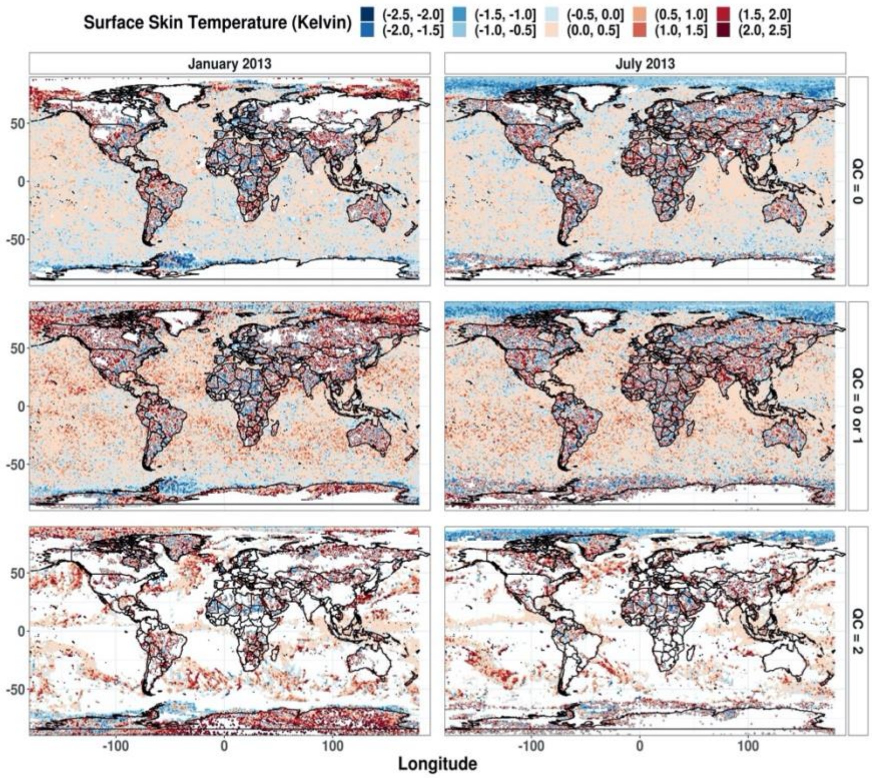

Figure 28. Global maps of surface skin temperature bias (RAMSESII-SNPP - ERA5) aggregated by month (column) and QC (rows).

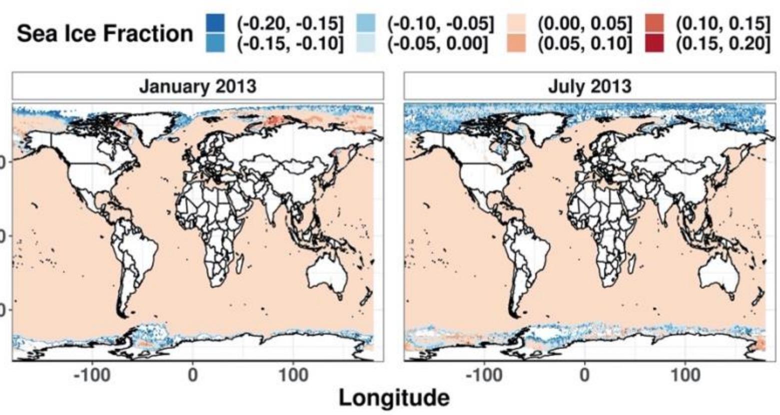

Figure 29. Global maps of sea ice fraction for bias (RAMSESII-SNPP – ERA5) aggregated by month (column).

Fig. 29 is the same as Fig. 27 but for sea ice fraction. There were no quality control flags for sea ice fraction. The location of the sea ice was quite comparable between RAMSESII-SNPP and ERA5, however, the magnitude was different. RAMSESII-SNPP sea ice fraction was generally smaller in magnitude to ERA5 for both January and July and at both poles. There was one exception, that was during January in an area north of Russia, particularly in the Kara

Sea; RAMSESII-SNPP had higher sea ice fraction values by almost 0.2 compared to ERA5. Table 1 showed that the sea ice fractions were highly correlated when sea ice fractions of 0 were included in the estimate. However, when only comparing sea ice fractions greater than 0 the R^2 value dropped from 0.8 to 0.4. On average, RAMSESII-SNPP had a smaller sea ice fraction by 0.02 to 0.09 and an RMSE of about 0.1 to 0.2.

3.5 Daily-Mean RAMSESII-SNPP Retrievals Compared with the Daily Gridded Fields in MERRA2 Reanalysis

In this section the deviations of the final retrievals from gridded MERRA2 for H_2O, T, tropopause pressure/temperature, surface temperature, 2-meter temperature, and total precipitable water on a daily basis will be examined. MERRA2 is the first guess for the specific humidity and temperature profile retrievals, however, because the first guess was not reported in the RAMSESII-SNPP output, daily gridded MERRA2 products were used instead. RAMSESII-SNPP and MERRA2 were placed on the same latitude/longitude/pressure grids for each day to offer a quick sanity check on the quality of RAMSESII-SNPP retrievals as well as provide a global quality of the RAMSESII-SNPP product. Since the gridded products were not collocated in space and time the MERRA2 output presented here is not representative of the true first guess. Similar to the bias calculations, (eqn. 1 and eqn. 3), the reference dataset will be MERRA2.

Fig. 30 shows the H_2O retrievals from RAMSESII-SNPP and MERRA2, as well as the deviations of RAMSESII-SNPP from MERRA2 in magnitude (g/kg) and in relative differences (%). RAMSESII-SNPP had similar H_2O vertical structure to MERRA2 but was broadly drier in most of the atmosphere except near the surface. The fractional deviations (right column) highlighted that the differences were pronounced from ~400 hPa to near the tropopause (blue). Deviations close to 0 mean differences between MERRA2 and the final retrieval are small suggesting the final retrieval may be sticking to the FG.

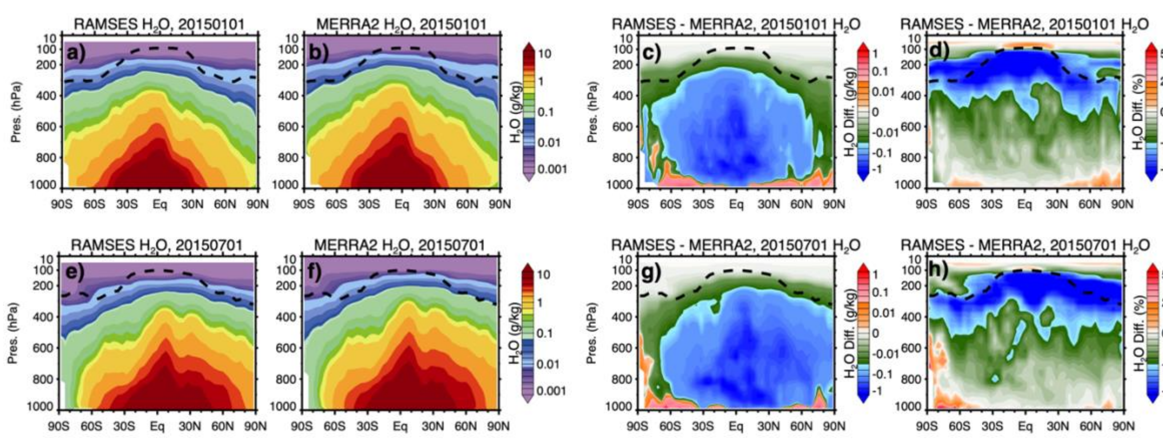

Figure 30. RAMSESII-SNPP H_2O retrievals (a, e) compared to MERRA2 (first guess proxy) (b, f) and the deviations from the MERRA2 in magnitude (g/kg, c, g) and in relative percentage change (%, d, h) for January 1, 2015 (top row) and July 1, 2015 (bottom row). The dashed lines are the local tropopause using the WMO definition.

Fig. 31 shows the temperature retrievals from RAMSESII-SNPP compared to MERRA2, as well as the deviations of RAMSESII-SNPP from MERRA2. Again, RAMSESII-SNPP had a similar vertical structure to MERRA2 but was broadly colder throughout the atmosphere, except near the polar surface and above the tropopause in the tropics.

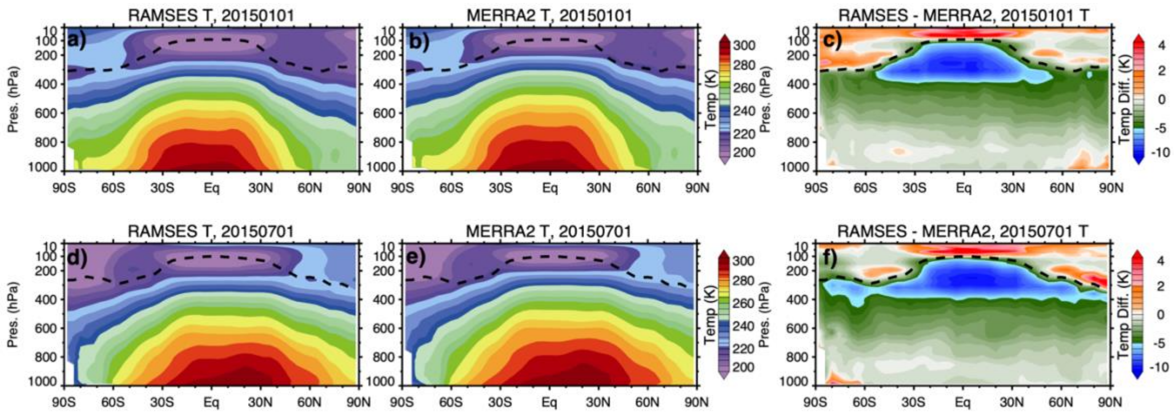

Figure 31. RAMSESII-SNPP temperature retrievals (a, d) compared to MERRA2 (proxy first guess) (b, e) and the deviations from the MERRA2 (c, f) for January 1, 2015 (top row) and July 1, 2015 (bottom row). The dashed lines are the local tropopause using the WMO definition.

Large differences in the tropopause pressure and temperatures are expected due to the differences observed in Fig. 31 between RAMSESII-SNPP and MERRA2. Fig. 32 shows the tropopause pressure (top) and temperature (bottom) for RAMSESII-SNPP (left), MERRA2 (middle), and the bias (right). RAMSESII-SNPP typically had a tropopause pressure 50 hPa higher (lower in altitudes) in the tropics and over 100 hPa higher in the high-latitudes compared to MERRA2, which inevitably brought higher tropopause temperatures. Generally the tropopause temperature was more than 2 Kelvin warmer in RAMSESII-SNPP than MERRA2.

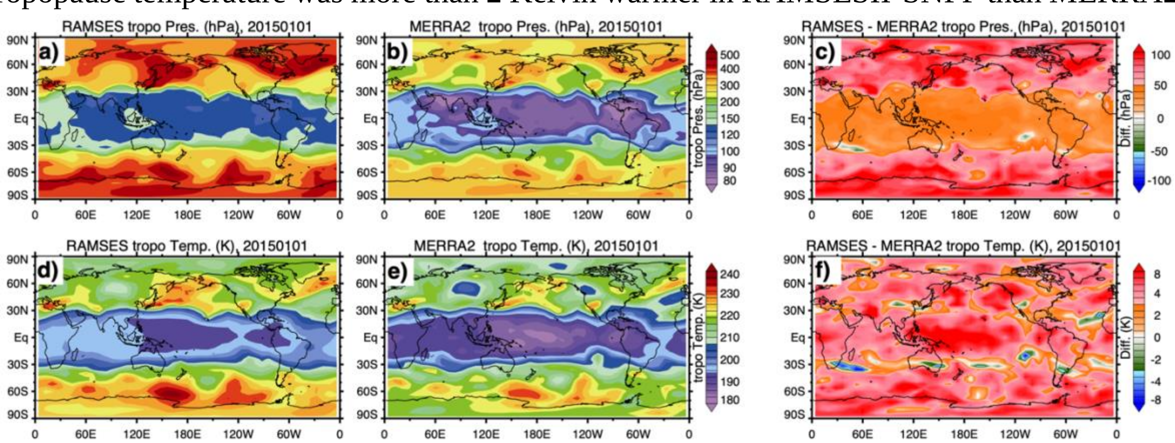

Figure 32. RAMSESII-SNPP tropopause pressure (a) and tropopause temperature (d) compared to MERRA2 (b, e) and the differences of the two (c, f) for January 1, 2015.

Fig. 33 shows the comparisons of RAMSESII-SNPP surface temperature, 2-m temperature, and total precipitable water to MERRA2. Generally, RAMSESII-SNPP captured the spatial pattern observed in MERRA2 for surface and 2-meter temperatures but

RAMSESII-SNPP was biased warm over land. Over the ocean, RAMSESII-SNPP 2-meter temperature was generally colder but the surface temperature was slightly warmer. Since MERRA2 is used as a first guess for surface temperature, it is possible that RAMSESII-SNPP sticks to MERRA2 over the ocean more hence smaller differences. However, there is a strong diurnal cycle over land for surface skin temperature which may impact a gridded daily comparison, such as what is presented here. Therefore, the persistent warm bias over land for surface skin temperature may be due to sampling differences. RAMSESII-SNPP produced smaller total precipitable water values, especially in the tropics. However, since this is a comparison on a daily grid; it is possible that the SNPP overpasses do not capture the true 24 -hour mean total precipitable water values. That is the RAMSESII-SNPP differences could be due to spatial and temporal sampling errors.

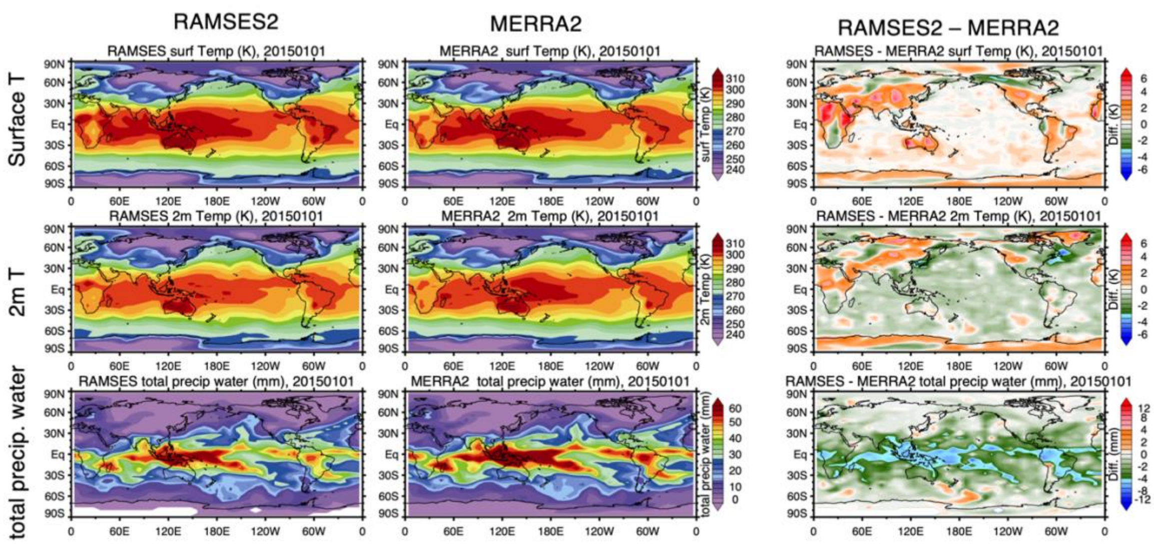

Figure 33. RAMSESII-SNPP surface temperatures (top row), 2-meter temperatures (middle row), and total precipitable water (bottom row) compared to MERRA2 (middle column) and the deviations from MERRA2 (right column) for January 1, 2015.

4. Summary and Conclusions

The goal of this report is to provide an initial testing analysis aimed at evaluating the general quality of the retrieval including the retrieval yields, the bias and RMSE of key parameters such as temperature (T) and water vapor/specific humidity (H_2O), and initial assessments of single level and total column variables. The possible connections of T/H_2O bias and RMSE to other parameters such as land fraction, surface temperature and total column water vapor were explored. IGRA/AWI radiosondes and ERA5 provided truth data for the comparison of retrievals from RAMSESII-SNPP. Below is a brief summary of the key results.

Temperature results were similar regardless of using IGRA/AWI or ERA5. Results discussed here are generalized across all truth data sets. RAMSESII-SNPP had a strong and persistent cold bias throughout the troposphere ranging from -1 to -7 Kelvin; substantially larger than AIRS v7. Similarly, RAMSESII-SNPP RMSE was quite large throughout the troposphere; especially when compared to AIRS v7. RAMSESII-SNPP bias and RMSE was smallest in the UTLS when retrieving over dry or cold scenes compared to wet or warm. There was a reduction in the bias and RMSE when retrieving over the ocean compared to land. RAMSESII-SNPP had a more conservative approach for quality control compared to AIRS v7. Indeed, the yield was smaller for QC = 0 compared to QC = 0 or 1 and the yield for QC = 0 was smaller than AIRS v7. However, RAMSESII-SNPP retained more retrievals near the surface than AIRS v7. Even with yield differences, the bias and RMSE from RAMSESII-SNPP was comparable regardless of using QC = 0 or QC = 0 or 1. Even with the high bias and RMSE, RAMSESII-SNPP had skill near the surface in mid-high latitudes and was able to reduce the AIRS v7 RMSE by 20-40%.

Specific humidity results were similar regardless of using IGRA/AWI or ERA5. Again, results discussed here are generalized across all truth data sets. RAMSESII-SNPP had a consistent dry bias throughout the troposphere ranging from -5 to -40%. The difference in bias and RMSE between RAMSESII-SNPP and AIRS v7 was much smaller compared to temperature. Furthermore, the RAMSESII-SNPP and AIRS v7 vertical structure of bias and RMSE were similar for specific humidity. There was no difference in bias and RMSE when retrieving over land versus ocean. The bias and RMSE were largest in the low to mid-troposphere when retrieving over dry or cold scenes compared to those that were wet or warm. RAMSESII-SNPP had similar bias and RMSE results regardless of using QC = 0 or QC = 0 or 1, even though the yields were different. Vertical structures of yield were different in RAMSESII-SNPP and AIRS v7. In particular, the yield did not change by pressure in RAMSESII-SNPP and remained constant throughout most of the profile. RAMSESII-SNPP had skill over AIRS v7 between 900 and 600 hPa, especially in the mid-high latitudes.

Single level and total column variables were assessed by comparing them to reanalysis (ERA5 and MERRA2). 2-meter temperature and surface skin temperature were quite comparable to the reanalyses. RAMSESII-SNPP 2-meter temperature had a warm bias over land and a cold bias over the ocean. ERA5 consistently had higher sea ice fraction compared to RAMSESII-SNPP with an RMSE of 0.1 to 0.2. RAMSESII-SNPP tropopause pressure was usually too high, by about 50 to 100 hPa which inevitably brought higher tropopause temperatures. RAMSESII-SNPP produced smaller total precipitable water values, especially in the tropics, when compared to MERRA2. When compared to AWI, RAMSESII-SNPP showed a consistent underestimation for low total precipitable water values and an overestimation for high values.

References

Copernicus Climate Change Service (C3S) (2017): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS), *date of access*. https://cds.climate.copernicus.eu/cdsapp#!/home

Driemel, A., Loose, B., Grobe, H., Sieger, R., and König-Langlo, G.: 30 years of upper air soundings on board of R/V POLARSTERN, Earth Syst. Sci. Data, 8, 213–220, https://doi.org/10.5194/essd-8-213-2016, 2016.

Durre, I., and X. Yin (2008), Enhanced radiosonde data for studies of vertical structure. Bulletin of the American Meteorological Society, 89, 1257-1262.

Hennermann, Karl. (2020). *What are the changes from ERA-Interim to ERA5?* - Copernicus Knowledge Base - ECMWF Confluence Wiki. From https://confluence.ecmwf.int/pages/viewpage.action?pageId=74764925

König-Langlo, G., and B. Marx, (1997): The meteorological in- formation system at the Alfred-Wegener-Institute. Climate and Environment Database Systems, M. Lautenschlager and M. Reinke, Eds., Kluwer, 117–126.

Lambrigtsen et al. RAMSES II Retrieval Algorithm for Microwave Sounders in Earth Science The NASA ATMS Retrieval System Level 2 Algorithm Theoretical Basis Document (ATBD), July 2021: https://doi.org/10.5067/K9LCXRFH92P

Monarrez, Ruth, et al, NASA RAMSES II Level-2 Products User Guide: File Format and Definition Version 1, July 2021 :https://doi.org/10.5067/K9LCXRFH92P

Murphy, A. H., Skill scores based on the mean square error and their relationship to the correlation coefficient, *Mon. Weather Rev.*, **116**, 2417–2424, 1988.

WMO. Manual on the Global Data-Processing and Forecasting system, Volume I - Global Aspects. 2012; Geneva, Switzerland: World Meteorological Organization. II. 7-36 - II. 7-38. Online at: http://www.wmo.int/pages/prog/www/DPFS/documents/485_Vol_I_en_colour.pdf .

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.