Remote Sensing of Environment 252 (2021) 112146

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Global evaluation of terrestrial near-surface air temperature and specific humidity retrievals from the Atmospheric Infrared Sounder (AIRS)

Jing Sun[a], Kaighin A. McColl[b,c,*], Yan Wang[a], Angela J. Rigden[b], Hui Lu[a], Kun Yang[a,d], Yishan Li[a], Joseph A. Santanello Jr[e]

[a] Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing 100084, China
[b] Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA 02138, USA
[c] School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA
[d] Center for Excellence in Tibetan Plateau Earth Sciences and National Tibetan Plateau Data Center, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China
[e] NASA-GSFC Hydrological Sciences Laboratory, Greenbelt, MD, USA

ABSTRACT

Global observations of near-surface air temperature and specific humidity over land are needed for a variety of applications, including to constrain global estimates of evapotranspiration (ET). Spaceborne hyperspectral observations, such as those from NASA's Atmospheric Infrared Sounder (AIRS) mission, show promise for meeting this need, yet there are surprisingly few validation studies of AIRS near-surface atmospheric state retrievals. In this study, we use triple collocation to validate AIRS Level 3 retrievals of near-surface atmospheric state over land using twelve years of gridded station observations and two reanalyses. Deseasonalized AIRS retrievals correlate well with deseasonalized ground observations outside the tropics, but correlate less well in the tropics. Lower temporal sensitivity near the surface in the tropics contributes to the lower correlation for near-surface air temperature and is consistent with known physics of the tropical atmosphere, in which temperatures outside the boundary layer (which dominate the AIRS retrieval signal) are poorly correlated with those near the surface. Retrievals in the tropics may also be more susceptible to errors in cloud-clearing algorithms, and to uncertainty in surface emissivity. Since ET is greatest in the tropics, and tropical measurement networks are particularly sparse, this work motivates new approaches for measuring ET in the tropics.

1. Introduction

The near-surface atmospheric state – in particular, near-surface air temperature and specific humidity – plays a critical role in human health, agriculture, and ecosystem function. More generally, the near-surface state both partially constrains, and is partially controlled by, surface fluxes of heat and moisture. For example, evapotranspiration (ET) is the second largest flux in the terrestrial water budget after precipitation, and links the water, energy and carbon cycles (Friedlingstein et al., 2013; Green et al., 2019). ET is controlled, in part, by the near-surface atmospheric state. All else being equal, a higher atmospheric temperature implies a higher vapor pressure deficit (VPD), and thus a higher atmospheric demand driving ET; similarly, lower specific humidity also increases VPD and atmospheric demand. In contrast, increased ET moistens and cools the near-surface atmosphere, creating a negative feedback between ET and the near-surface atmospheric state (Seneviratne et al., 2010). Broadly speaking, ET is not accurately represented in models (Mueller and Seneviratne, 2014). Since models exhibit large biases in near-surface temperatures over many land regions (Ma et al., 2018; Wehrli et al., 2018), errors in near-surface atmospheric variables may be both a cause and effect of errors in modelled ET (McColl et al., 2019b; McColl and Rigden, 2020; Salvucci and Gentine, 2013).

For these and other reasons, global observations of the near-surface atmospheric state over land are urgently needed. Satellite observations of the near-surface atmospheric state show great potential for meeting this need. NASA's Atmospheric Infrared Sounder (AIRS; Chahine et al., 2006) retrievals have been used extensively to evaluate the accuracy of surface warming trends (Susskind et al., 2019), and climate and weather model predictions (Gettelman et al., 2006; Jiang et al., 2012; Tian et al., 2013). Several widely-used ET schemes have used observations of near-surface state variables from AIRS as inputs, including near-surface air temperature and specific humidity (Badgley et al., 2015; Mallick et al., 2015; Martens et al., 2017; Vinukollu et al., 2011). It has also been used for estimating related quantities, such as vapor pressure deficit (Giardina et al., 2018).

However, while promising, there is some reason for skepticism regarding the accuracy of near-surface state retrievals from spaceborne

* Corresponding author at: Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA 02138, USA.
E-mail address: kmccoll@seas.harvard.edu (K.A. McColl).

https://doi.org/10.1016/j.rse.2020.112146
Received 13 March 2020; Received in revised form 24 September 2020; Accepted 17 October 2020
Available online 3 November 2020

hyperspectral observations. Consistent with its primary mission objectives, the AIRS retrieval is mainly based on information from the free troposphere, with relatively little contribution from the atmospheric boundary layer. Any spectral signal from the near-surface environment must either be strong enough to overwhelm competing signals higher in the atmosphere, or be strongly correlated with them (Wulfmeyer et al., 2015).

Given their increasingly widespread use, it is somewhat surprising that relatively few validation studies have been conducted of AIRS near-surface state retrievals over land. Most previous validation studies of near-surface AIRS retrievals have focused on individual sites or focus regions (Dang et al., 2017; Ferguson and Wood, 2010; Gao et al., 2008; Hearty et al., 2018; Prakash et al., 2019; Tobin et al., 2006), or globally-averaged performance over land (Divakarla et al., 2006). The AIRS retrieval algorithm has been substantially updated since most of these original validation studies were conducted (Susskind et al., 2014). There is a clear need for validation of AIRS near-surface temperature and specific humidity over land that is both spatially resolved and global in coverage, and is up-to-date with changes in the AIRS retrieval algorithm.

This study meets that need, primarily by comparing AIRS observations to gridded station measurements of near-surface air temperature and specific humidity with approximately global coverage over land (the Hadley Centre's Integrated Surface Database, or HadISD (Dunn et al., 2012, 2014, 2016)). One approach to this comparison would be to estimate differences between AIRS observations and station observations, and attribute differences between the two to errors in the AIRS observations. However, an important confounding factor in making this comparison is the scale mismatch between point-scale station measurements, and spatially-distributed Level 3 AIRS retrievals, which, in this study, can be thought of as approximate averages over $1^\circ \times 1^\circ$ regions. The scale mismatch induces so-called 'representativeness errors' in the station data. That is, even if AIRS retrievals were free of all errors, we would still expect there to be differences between the station observations (which measure quantities at a point) and the AIRS retrievals (which are gridded $1^\circ \times 1^\circ$ spatial averages). Essentially, they are measuring different, but correlated, quantities. Therefore, attributing a difference between an AIRS retrieval and a station measurement solely to errors in the AIRS retrieval would overestimate the AIRS retrieval error: some of the difference is due to the scale mismatch between the station observation and AIRS retrieval (Prakash et al., 2019).

In this study, we use an established technique for handling the scale mismatch in satellite validation studies, called 'triple collocation' (Stoffelen, 1998), extended by McColl et al. (2014). The technique is robust to the presence of representativeness errors induced by the scale mismatch, resulting in an unbiased assessment of the performance of AIRS retrievals. Triple collocation (TC) requires the use of a third estimate of near-surface air temperature and specific humidity, with errors that are largely uncorrelated with those of AIRS and HadISD. We use reanalysis estimates for this purpose (and discuss the strengths and weaknesses of this choice later). TC has been used to validate satellite retrievals of soil moisture (e.g., Draper et al., 2013; Gruber et al., 2016), wind speed (e.g., Stoffelen, 1998; Vogelzang et al., 2011), precipitation (e.g., Alemohammad et al., 2015; Roebeling et al., 2012), landscape freeze/thaw state (Lyu et al., 2018; McColl et al., 2016) and other geophysical variables. To our knowledge, this study is the first application of TC to validating retrievals of near-surface air temperature and specific humidity. Further details on TC and the datasets used in this study are given in Section 2; the results are presented in Section 3, and interpreted through the lens of known physics of the atmosphere in Section 4.

2. Data and methodology

In this section, we describe the datasets used in this study, detail how they are compared and deseasonalized, and give an overview of TC.

2.1. Data

In this study, five global datasets (AIRS L3, TES L3, MERRA2, ERA-interim, and HadISD) are used, spanning the time period 30 August 2002 to 31 December 2014. In order to match the data in space and time, we selected time series that overlap across the three datasets and regridded the data onto a common grid (detailed below).

2.1.1. Satellite datasets

The primary focus of this study is on AIRS retrievals. However, to provide context for our results, we also examined retrievals from the Tropospheric Emission Spectrometer (TES; Beer, 2006).

2.1.1.1. AIRS. AIRS launched into orbit on May 4, 2002 aboard NASA's Aqua satellite (Aumann et al., 2003; Chahine et al., 2006; Tobin et al., 2006). It provides retrievals at 100 vertical levels with nominal accuracy of 1 K/km, although the true vertical resolution varies with height and location (Maddy and Barnet, 2008), as does the true accuracy. AIRS has 2378 spectral channels, and measures infrared brightness from radiation emitted from Earth's surface and the atmosphere (Susskind et al., 2011, 2014). Each infrared wavelength is sensitive to temperature and water vapor over a particular range of heights in the atmosphere (Menzel et al., 2018). Based on overlapping trapezoidal perturbation functions, air temperature and water vapor retrievals are obtained by optimizing the fit to 147 and 66 channels, respectively. Cloud-cleared radiances are used to retrieve the AIRS Standard Product (Susskind et al., 2011).

The product was separated into ascending (1:30 PM local time) and descending (1:30 AM local time) 'observations' per day. Only the ascending overpass was used in this study, as we are primarily interested in daytime conditions. Specifically, we used the variables SurfAirTemp and H2O_MMR_Surf of the AIRS Level 3 Version 6 Daily Standard Physical Retrieval product (AIRS3STD.006), with a horizontal resolution of $1^\circ \times 1^\circ$ (Susskind et al., 2011), as near-surface air temperature and specific humidity. Level 3 AIRS products only include retrieved quantities with Level 2 quality flags labelled "best" or "good". Quality flags are determined based on a weighted sum of several parameters found to correlate with retrieval accuracy, including internal indicators of scene contrast, retrieval convergence, and differences between results at different stages of the retrieval (Susskind et al., 2011).

2.1.1.2. TES. TES was launched in July 2004 aboard the EOS Aura mission (Beer, 2006). Like AIRS, TES measures infrared brightness from radiation emitted from Earth's surface and the atmosphere. TES has a higher spectral resolution ($\sim$0.12 cm^{-1}) compared with that of AIRS ($\sim$1 cm^{-1}), but AIRS has nearly 1000 times the sampling density of TES (Worden et al., 2019).

TES is in a sun-synchronous orbit with a local overpass time of 1:30 PM local time, available every other day. Specifically, for near-surface air temperature, we used the variable TATMAtSurface from the TES/Aura L3 Atmospheric Temperatures Daily Gridded V005 product; and H2OAtSurface from the TES/Aura L3 Water Vapor Daily Gridded V005 product. Both products have a spatial resolution of $2^\circ \times 4^\circ$. In this study, for TES, we use data spanning the period August 22, 2004 – December 31, 2014.

2.1.2. In-situ dataset

The U.K. Met Office Hadley Centre's Integrated Surface Database (HadISD) is a global sub-daily, quality-controlled and station-based dataset which includes observations of near-surface air temperature and specific humidity (Dunn et al., 2012, 2014, 2016). The major climate variables, including temperature and dewpoint temperature, have passed quality control tests, which aimed to remove erroneous observations but not extreme values (see Dunn et al. (2012, 2014, 2016) for further details). In this study, we used version 2.0.2.2017f, consisting of 8103 stations, which were selected based on their record length and reporting frequency.

2.1.3. Reanalysis datasets

In addition to the satellite and in-situ datasets, two reanalysis datasets of near-surface air temperature and specific humidity are used in this study. Triple collocation requires three different datasets with largely uncorrelated errors (discussed further in Section 2.2.1).

2.1.3.1. MERRA-2. The second Modern-Era Retrospective Analysis for Research and Applications (MERRA-2), produced by NASA's Global Modeling and Assimilation Office (GMAO), is the latest satellite reanalysis product of the modern era (Gelaro et al., 2017). Based on the first MERRA, MERRA-2 assimilates a range of satellite and other observations into the GEOS model (Jiang et al., 2015; Molod et al., 2015). It has spatial and temporal resolutions of $0.5^\circ \times 0.625^\circ$ and 1 h, respectively.

2.1.3.2. ERA-Interim. ERA-Interim, produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), is a global atmospheric reanalysis and covers the period from 1979 to the present. In this study, we used $0.75^\circ \times 0.75^\circ$ gridded surface data with a temporal resolution of 6 h (Berrisford et al., 2011; Dee et al., 2011). Because it only has four analyses per day at 00, 06, 12 and 18 UTC, we used cubic spline interpolation to obtain hourly data to temporally match the reanalyses to the AIRS overpass time. The results of this analysis are qualitatively insensitive to the choice of interpolation method. We used 2-m temperature as the near-surface air temperature and calculated the near-surface specific humidity from the 2-m dewpoint temperature and surface pressure.

2.1.3.3. Other data processing. The Level 3 AIRS and TES observations used in this study are provided at a much coarser resolution than the MERRA-2, ERA-Interim and HadISD observations. In order to match AIRS (TES) data in space, MERRA-2 and ERA-Interim data were resampled onto a $1^\circ \times 1^\circ$ ($2^\circ \times 4^\circ$) grid prior to analysis using nearest neighbor resampling. HadISD station data were resampled to the AIRS (TES) observation scale by simple averaging of all station observations within a given AIRS (TES) grid cell (Fig. 1 shows the number of HadISD used in the average for each AIRS grid cell). Similarly, MERRA-2, ERA-Interim and HadISD data were temporally matched to the ascending (~1:30 PM local time) AIRS (TES) observations by nearest neighbor resampling.

Prior to performing triple collocation, the seasonal cycle was removed from each dataset: that is, the monthly mean of each dataset was subtracted from each observation in the dataset. We remove the seasonal cycle to allow a fairer comparison between the tropics (where the seasonal cycle is typically minimal) and higher latitudes (where the seasonal cycle is often larger). The correlation coefficient can be thought of as a normalized signal-to-noise ratio (McColl et al., 2014). Since the seasonal cycle often contributes substantially to the observed temperature and humidity signals, this implies that the AIRS retrievals would exhibit lower correlation coefficients in the tropics compared to higher latitudes, even if there were no differences in the measurement noise of the AIRS retrievals between the tropics and higher latitudes. Removing the seasonal cycle eliminates this confounding effect on the estimated correlation coefficient.

2.2. Methodology

2.2.1. Triple collocation

Given three different types of observations of a given target variable, triple collocation (TC) estimates the error standard deviations (Stoffelen, 1998) and correlation coefficients (McColl et al., 2014) of each observation type with respect to the target variable, without assuming any of the three types of observations are free of errors. This is critical since, as discussed earlier, station observations contain substantial representativeness errors: the number of HadISD stations included in each AIRS pixel in the analysis is typically one or two (Fig. 1). TC treats all three measurements of the target variable as linearly but noisily related to the target variable:

$$X_i = \alpha_i + \beta_i T + \varepsilon_i$$

where the X_i (i=1, 2, 3) are observations from the three collocated measurement systems; T is the unknown target variable; α_i and β_i are the ordinary least squares intercepts and slopes, respectively; and ε_i are mean-zero additive random errors. This is a common assumption that is often made implicitly in many validation studies (Gruber et al., 2016). In this study, the unknown target variables are near-surface air temperature and specific humidity. The three types of observations used are HadISD ($i = 1$), a satellite product ($i = 2$; either AIRS or TES) and a reanalysis product ($i = 3$; either MERRA2 or ERA-Interim).

TC assumes that the three observation types have errors which are uncorrelated with one another ($\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$), and with the target variable ($\mathrm{Cov}(\varepsilon_i, T) = 0$). These assumptions are likely to be at least partially violated (Yilmaz and Crow, 2014), although there is little information available to refine this assertion. We note that these assumptions are not unique to TC, and are implicitly made (and likely violated) in most validation studies. For example, Gruber et al. (2016) showed that adopting a traditional validation strategy – estimating the correlation coefficient and root-mean-squared difference (RMSD)

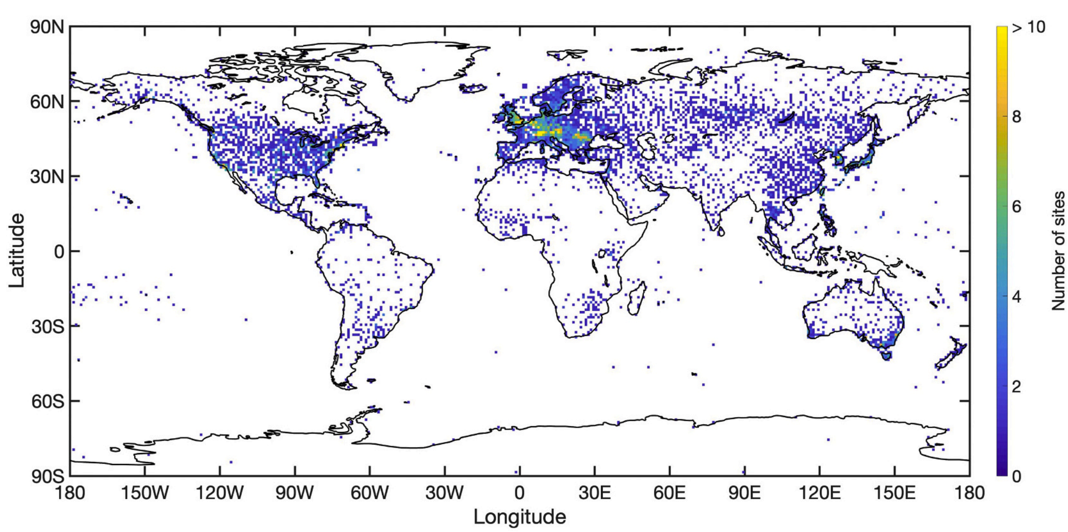

Fig. 1. Number of HadISD stations included in each AIRS pixel in the analysis.

between satellite observations and ground observations, and interpreting higher correlation coefficients and lower RMSDs as indicators of better satellite performance – requires exactly the same assumptions.

Given these assumptions, the TC estimation equations for the standard deviation of the random error σ_{TC} and the coefficient of determination R^2_{TC} are:

$$\boldsymbol{\sigma_{TC}} = \begin{bmatrix} \sqrt{Q_{11} - \frac{Q_{12}Q_{13}}{Q_{23}}} \\ \sqrt{Q_{22} - \frac{Q_{12}Q_{23}}{Q_{13}}} \\ \sqrt{Q_{33} - \frac{Q_{13}Q_{23}}{Q_{12}}} \end{bmatrix}, \boldsymbol{R^2_{TC}} = \begin{bmatrix} \frac{Q_{12}Q_{13}}{Q_{11}Q_{23}} \\ \frac{Q_{12}Q_{23}}{Q_{22}Q_{13}} \\ \frac{Q_{13}Q_{23}}{Q_{33}Q_{12.}} \end{bmatrix}$$

where Q_{ij} represents the covariance between sample time series from observations X_i and X_j; and σ_{TCi} and R^2_{TCi} are the noise error standard deviation and correlation coefficient of observation X_i with respect to the target variable, respectively.

TC is not able to estimate absolute values of the additive and multiplicative bias terms (α_i and β_i, respectively). However, it can estimate relative values: that is, if the HadISD station observations are treated as unbiased ($\alpha_1 = 0$ and $\beta_1 = 1$), the relative additive and multiplicative biases for the AIRS and reanalysis observations are given by (McColl et al., 2014):

$$\widehat{\beta}_2 = \frac{Q_{23}}{Q_{13}}, \widehat{\beta}_3 = \frac{Q_{23}}{Q_{12}}$$

$$\widehat{\alpha}_2 = \overline{X_2} - \widehat{\beta}_2\overline{X_1}, \widehat{\alpha}_3 = \overline{X_3} - \widehat{\beta}_3\overline{X_1}$$

where $\overline{X_i}$ is the sample mean; and $\widehat{\alpha}_i$ and $\widehat{\beta}_i$ are the relative additive and multiplicative biases, respectively. For simplicity of notation, we drop the $\widehat{\ }$-symbol and denote the relative bias terms as α_i and β_i for the remainder of the manuscript.

The multiplicative bias β can be interpreted as the temporal 'sensitivity' of the measurement to the underlying target variable T: small values of β result in small temporal fluctuations in the measurement X even for large temporal fluctuations in T. The term 'sensitivity' has different meanings in different contexts. In this work, the AIRS temporal 'sensitivity' refers to β estimated for the Level 3 AIRS product in its current form. It does not refer to the sensitivity of the AIRS instrument. For example, the Level 3 AIRS product may exhibit lower temporal sensitivity to the observed temperature than the AIRS-observed radiances due to artifacts of the retrieval algorithm or other processing. We also distinguish 'temporal sensitivity' (estimated in this study) from 'vertical sensitivity', which is a measure of the spatial (vertical) resolution of the AIRS profile (Maddy and Barnet, 2008). This study focuses solely on AIRS near-surface products, and therefore does not evaluate vertical sensitivity.

Like all validation metrics, quantities estimated by TC are subject to sampling error. We used bootstrapping (Efron and Tibshirani, 1994; chapter 6) with 5000 replicates to quantify the uncertainty in estimates of α_i and β_i. When plotting α_i, estimates of α_i with a 95% confidence interval that overlapped zero were manually set equal to zero. When plotting β_i, estimates of β_i with a 95% confidence interval that overlapped one were manually set equal to one. This ensures that reported non-zero estimates of α_i and non-unity estimates of β_i are unlikely to be artifacts of sampling error. In addition, if any TC-estimated σ^2_{TCi} was negative, or any TC-estimated R^2_{TCi} was negative or greater than one, it was discarded. Similarly, in rare cases in which estimates of β_i were negative or greater than two, they were discarded, along with the corresponding α_i. These values can arise if sampling error is significant or if one of the assumptions of TC is violated.

Since the primary focus of this study is on the error statistics of the AIRS products, rather than the HadISD or reanalysis products, we simplify our notation for the remainder of the study. Specifically, instead of writing σ_{TC2} and R^2_{TC2} for the standard deviation of the random error and the coefficient of determination for the AIRS products, respectively, we write σ_{TC}(AIRS) and R^2_{TC}(AIRS) instead. Similarly, instead of writing α_2 and β_2, we write α(AIRS) and β(AIRS).

3. Results

In this section, we present the major results of the triple collocation validation analysis of AIRS retrievals of near-surface air temperature and specific humidity. The estimated coefficient of determination R^2_{TC}(AIRS) is relatively high at mid- and high-latitudes for both air temperature and specific humidity (Fig. 2). Averaging reported R^2_{TC}(AIRS) values over latitudes outside the region [10°S, 10°N] gives 0.71 and 0.58 for air temperature and specific humidity over land, respectively. However, within the tropics, performance of AIRS retrievals over land degrades substantially. Averaging reported R^2_{TC}(AIRS) over latitudes within the region [10°S, 10°N] gives 0.38 and 0.19 for air temperature and specific humidity, respectively. This result is qualitatively consistent if the analysis is performed separately for different seasons, for both air temperature (Fig. 3) and specific humidity (Fig. 4).

The standard deviation of the noise error in the AIRS retrievals, σ_{TC}(AIRS), is shown in Fig. 5. For AIRS retrievals of near-surface air temperature over land, σ_{TC}(AIRS) is lowest in the tropics. However, for retrievals of near-surface specific humidity over land, σ_{TC}(AIRS) is highest in the tropics. These results are also qualitatively consistent if the analysis is performed separately for different seasons, for both air temperature (Fig. 6) and specific humidity (Fig. 7).

Maps of relative additive and multiplicative biases (Figs. 8 and 9, respectively) are presented, again, for retrievals of both near-surface air temperature and specific humidity. In most parts of the world, relative additive biases are indistinguishable from zero for AIRS retrievals of specific humidity (Fig. 8b). For air temperature, they are negative in most parts of the world, and are most negative in the eastern United States and Europe (Fig. 8a). In the tropics, the relative additive bias is closer to zero. The relative multiplicative bias is less than one for AIRS retrievals of both air temperature (Fig. 9a) and specific humidity (Fig. 9b). It is particularly low in the tropics for AIRS retrievals of air temperature (lack of observations in the tropics makes it difficult to evaluate the equivalent claim for specific humidity).

To evaluate the impact of choice of reanalysis on the TC analysis, the presented results were repeated using ERA-Interim instead of MERRA2 (not shown). The results are qualitatively similar to those using MERRA2, indicating differences in reanalysis choice do not have a substantial effect on the results of this study. Furthermore, the results are qualitatively similar if TES retrievals of near-surface air temperature and specific humidity are used instead of AIRS (not shown). TES retrievals exhibit systematically lower correlations with ground measurements (not shown), which are likely a result of its much coarser spatial resolution.

There is some concern in the use of TC in this study that its assumptions are violated by including reanalyses, which ingest AIRS observations, perhaps inducing error correlations between measurements that are assumed to be zero by TC. In addition, the AIRS retrievals include a component based on a neural net trained on ECMWF reanalysis (Blackwell and Milstein, 2014; Milstein and Blackwell, 2016). Near the surface, the AIRS retrieval may be substantially influenced by the reanalysis training set, again potentially creating error correlations between measurements that violate the assumptions of TC. In Appendix A, we demonstrate that error cross-correlation between the AIRS retrieval and the reanalyses is unlikely to explain the estimates of lower AIRS multiplicative bias in the tropics (Fig. 9). We show that, if anything, the presence of error cross-correlation would overestimate the multiplicative bias of AIRS in the tropics. Therefore, our results are unlikely to be

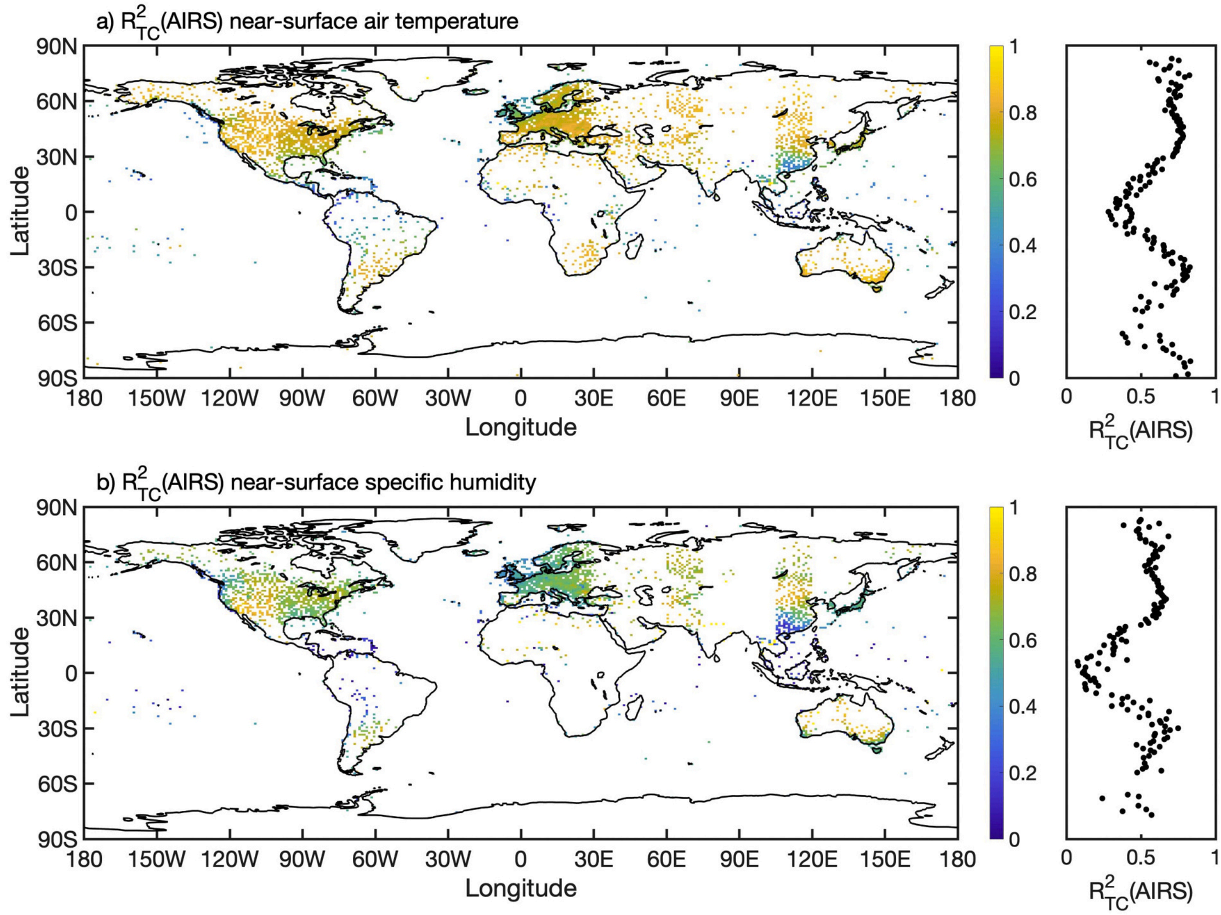

Fig. 2. Global maps and latitudinal averages of triple collocation (TC)-estimated coefficient of determination R^2_{TC}(AIRS) for deasonalized near-surface a) air temperature and b) specific humidity over land, using HadISD, AIRS and MERRA-2 at the ascending time.

an artifact of violations of the assumptions of TC.

4. Discussion

4.1. Reconciling latitudinal patterns of R^2_{TC}(AIRS) and σ^2_{TC}(AIRS)

A striking feature of Fig. 2 is the relatively low R^2_{TC}(AIRS) in the tropics, both for near-surface air temperature and specific humidity over land. This suggests AIRS retrievals of near-surface air temperature and specific humidity have poorest performance over land in the tropics. However, for near-surface air temperature, the noise error standard deviation σ_{TC}(AIRS) is also lowest in the tropics (Fig. 5): on this measure of performance, AIRS retrievals of near-surface air temperature exhibit strongest performance over land in the tropics. The same results hold if the analyses are conducted separately for each season (Figs. 3 and 6), suggesting it is not an artifact of differences in seasonality between the tropics and higher latitudes.

How should these results be reconciled? The correlation coefficient is an increasing function of the signal-to-noise ratio, meaning, for a given noise error standard deviation, it can take on any value between zero and one (McColl et al., 2014). Low correlations have three major causes: 1) large random error ('noise') in the observation (i.e., high σ); 2) small observation temporal sensitivity to the true signal (i.e., low β) and/or 3) small variability in the true signal (i.e,., low standard deviation of T). The differing results in the tropics tell us that, while the noise error in retrievals of near-surface air temperature over land is lowest in the tropics, the measured signal must be proportionally lower, either due to lower β, lower variability in T, or both. TC is not able to estimate the variance of T, but it is likely that lower variability in temperature and humidity in the tropics contributes to the lower correlations observed in the tropics (although differences in the seasonal cycle between the tropics and higher latitudes do not contribute, since all time series were deseasonalized prior to analysis, and qualitatively similar results are obtained if the analysis is conducted separately for each season). However, in addition to this effect, a substantial contributor to the reduction in measured signal is the relatively low multiplicative bias β(AIRS) which dampens the observed signal relative to station observations (i.e., β(AIRS) < 1), particularly in the tropics (Fig. 9). Therefore, the low AIRS correlation coefficients in the tropics for near-surface air temperature over land are due, at least in part, to relatively low temporal sensitivity (i.e., low β) rather than relatively high noise (high σ), above and beyond likely differences in variability of near-surface air temperature and specific humidity between the tropics and mid-latitudes.

4.2. Possible causes of lower temporal sensitivity in the tropics over land

A major result of this study is that AIRS retrievals of terrestrial near-surface state show significant potential in the extra-tropics but less potential in the tropics, where correlation with ground observations is relatively low. Why does AIRS perform well outside the tropics, but not in the tropics? In particular, why is the temporal sensitivity β(AIRS) – which we have identified as one cause of the low correlation in near-surface air temperature over land – systematically lower in the tropics?

Here, we review several possible causes for the poorer performance of AIRS in the tropics compared with higher latitudes. The list of possible causes reviewed in this section is clearly not exhaustive, but is provided to contextualize the results presented in the previous section.

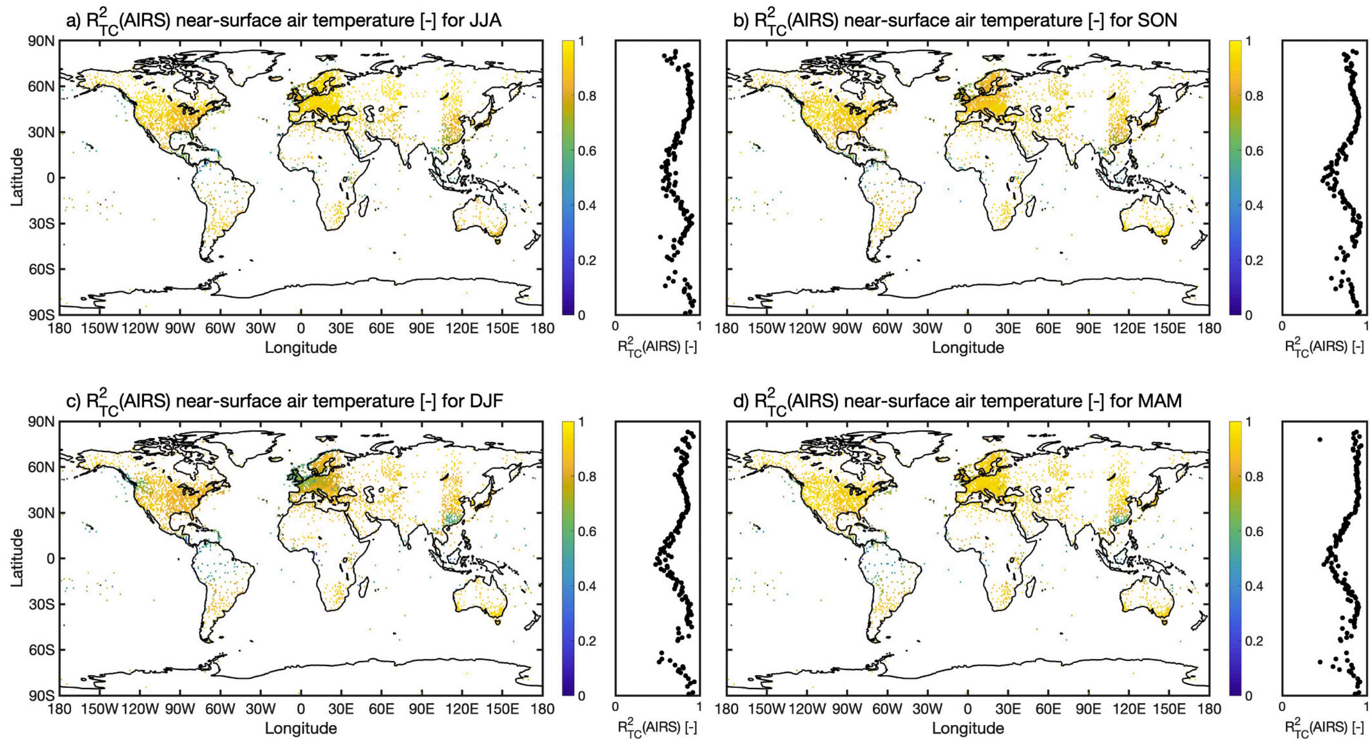

Fig. 3. Global maps and latitudinal averages of triple collocation (TC)-estimated coefficient of determination R^2_{TC}(AIRS) for deseasonalized near-surface air temperature over land for a) June–August (JJA) b) September–November (SON) c) December–February (DJF) d) March–May (MAM). HadISD, AIRS and MERRA2 data at the ascending time were used in the triple collocation analysis for this figure.

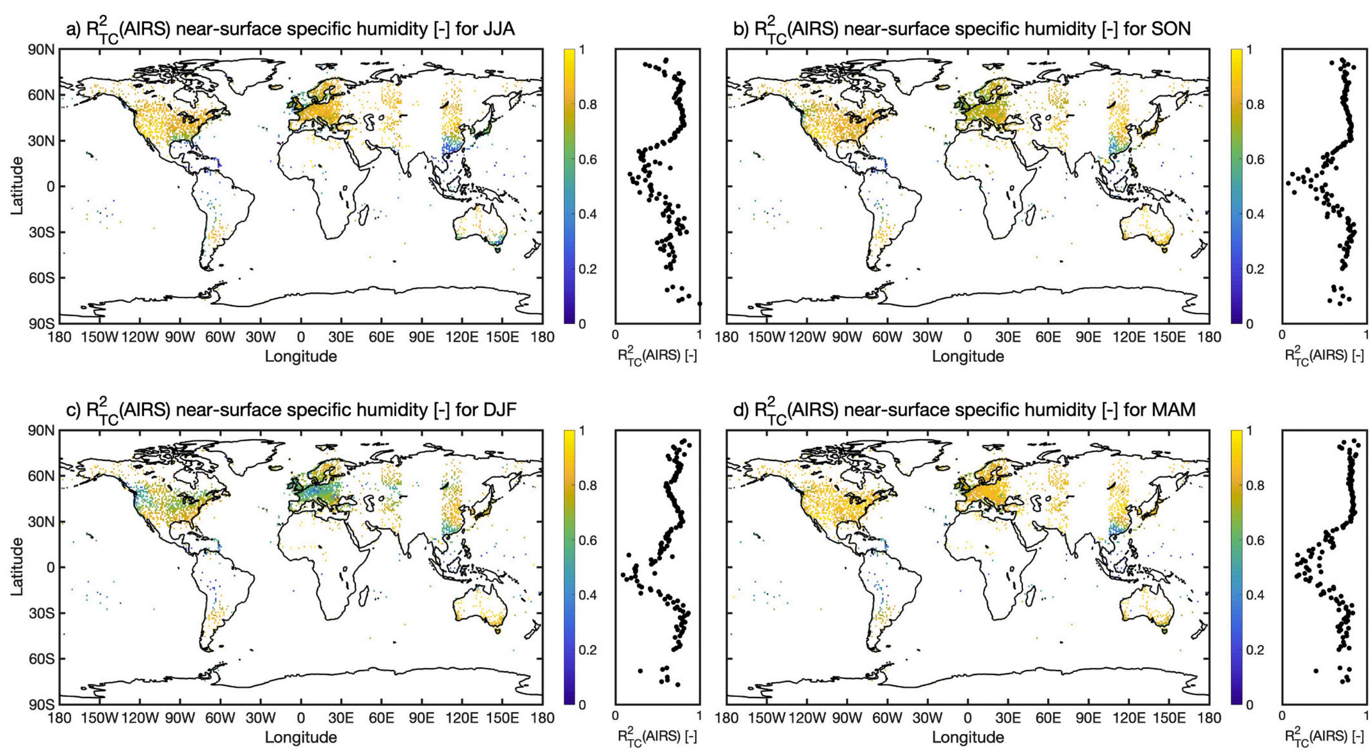

Fig. 4. Global maps and latitudinal averages of triple collocation (TC)-estimated coefficient of determination R^2_{TC}(AIRS) for deseasonalized near-surface specific humidity over land for a) June–August (JJA) b) September–November (SON) c) December–February (DJF) d) March–May (MAM). HadISD, AIRS and MERRA2 data at the ascending time were used in the triple collocation analysis for this figure.

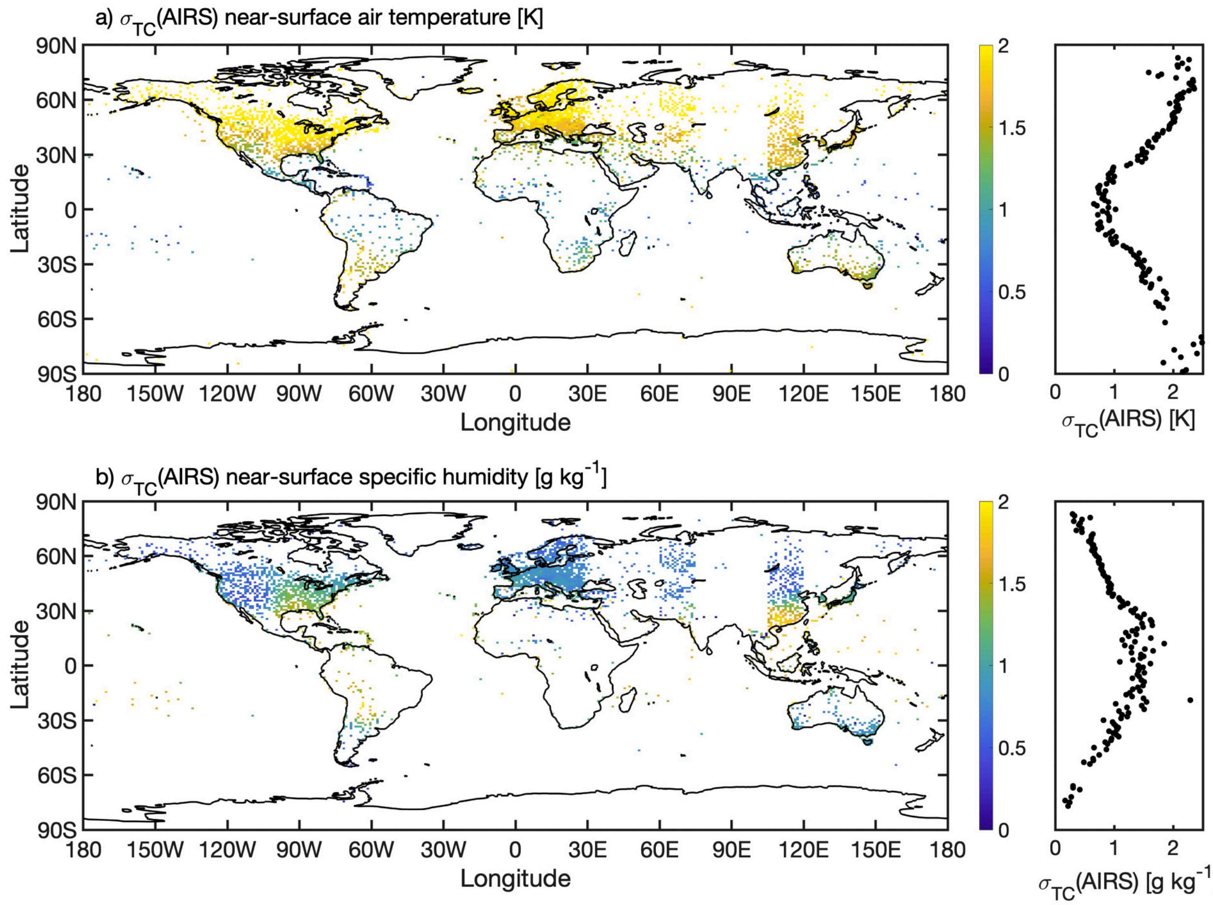

Fig. 5. Global maps and latitudinal averages of triple collocation (TC)-estimated noise error standard deviations in AIRS retrievals of deseasonalized a) near-surface air temperature and b) near-surface specific humidity over land, using HadISD, AIRS and MERRA2 data at the ascending time.

First, clouds are more prevalent in the tropics, and likely confound retrievals to a greater extent than at higher latitudes. AIRS includes a cloud-clearing algorithm to mitigate this problem (Susskind et al., 2003), but errors remain (Chahine et al., 2006), and will likely be greater in the tropics.

Second, it is possible that systematic uncertainties in surface emissivity also contribute (Chahine et al., 2006). While surface emissivity directly impacts retrievals of surface temperature, it also contributes indirectly to retrievals of near-surface air temperature. If uncertainties in surface emissivity are greater for tropical forests than other land cover types, or greater for coastal regions, then this may partially explain the poorer performance in the tropics.

Third, differences between the first-order structure of the atmosphere in the tropics and extra-tropics may also contribute. Waves in the tropical free troposphere spread temperature signals horizontally, resulting in a relatively constant temperature in the free troposphere (Sobel et al. (2001) term this the "weak temperature gradient" (WTG) approximation). This implies that anomalies in near-surface atmospheric temperature that propagate into the free troposphere will be rapidly smoothed out by tropical waves; further, it implies that free tropospheric temperatures will be relatively insensitive to, and poorly correlated with, near-surface temperatures. Since the AIRS retrieval signal is dominated by contributions from the free troposphere (Susskind et al., 2003; Wulfmeyer et al., 2015), it suggests that Level 3 AIRS retrievals – in their current form – will be only weakly sensitive to, and therefore only poorly correlated with, near-surface air temperatures in the tropics (AIRS also provides a more direct retrieval of surface temperature, but this differs significantly from the near-surface air temperature of interest in this study). Previous studies have found that AIRS air temperature retrievals from the boundary layer and free troposphere have relatively low correlation in the tropics (Holloway and Neelin, 2007; Wu et al., 2006). This result is also present in radiosonde datasets, so is not an artifact of AIRS (Holloway and Neelin, 2007; Wu et al., 2006). Outside the tropics, the WTG approximation does not apply, and free tropospheric temperatures are more sensitive to variations in near-surface temperatures. The increased temporal sensitivity leads to higher correlations between AIRS observations and ground observations, despite the fact that AIRS is primarily measuring the free troposphere. In contrast, the WTG approximation does not directly apply to specific humidity, which is generally sensitive to cloud microphysics, entrainment and other spatially variable processes (Emanuel, 2018). This may partially explain why noise error contributes more to lower correlations for specific humidity in the tropics (Fig. 5b), compared with air temperature (Fig. 5a).

4.3. Implications for satellite retrievals

These results have implications for other satellite retrievals of near-surface atmospheric temperature and specific humidity, such as those from MODIS. The same error sources listed in the previous section that impact retrievals in the tropics will likely impact retrievals from MODIS and other satellites. While global validation studies are useful (Famiglietti et al., 2018), surface observations are relatively sparse in the tropics, meaning global validation exercises may overstate the global accuracy of satellite retrievals. Accurate estimates of the terrestrial near-surface state are particularly important in the tropics since that is where terrestrial ET is largest (Budyko et al., 1980; Fisher et al., 2008). Our work suggests that separate validation studies focused on the tropics are

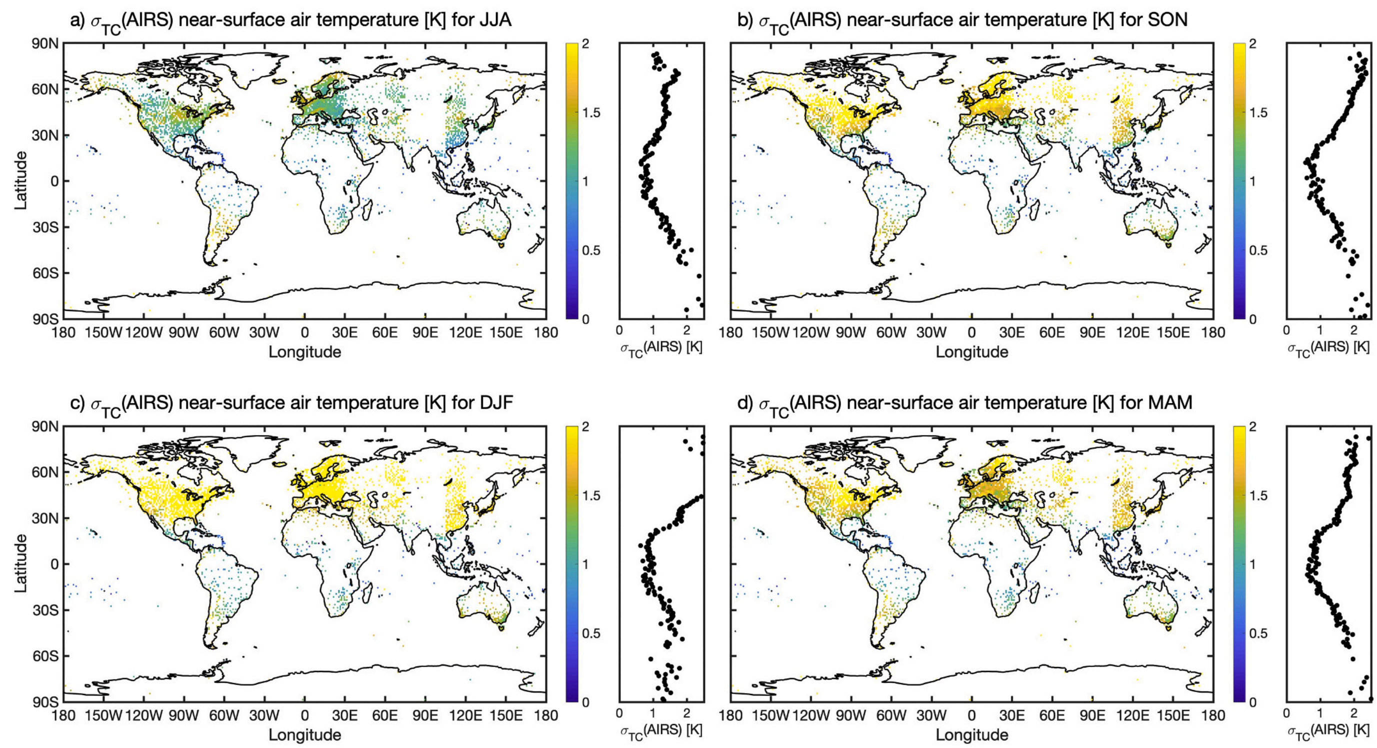

Fig. 6. Global maps and latitudinal averages of triple collocation (TC)-estimated noise error standard deviations in AIRS retrievals of deseasonalized near-surface air temperature over land for a) June–August (JJA) b) September–November (SON) c) December–February (DJF) d) March–May (MAM). HadISD, AIRS and MERRA2 data at the ascending time were used in the triple collocation analysis for this figure.

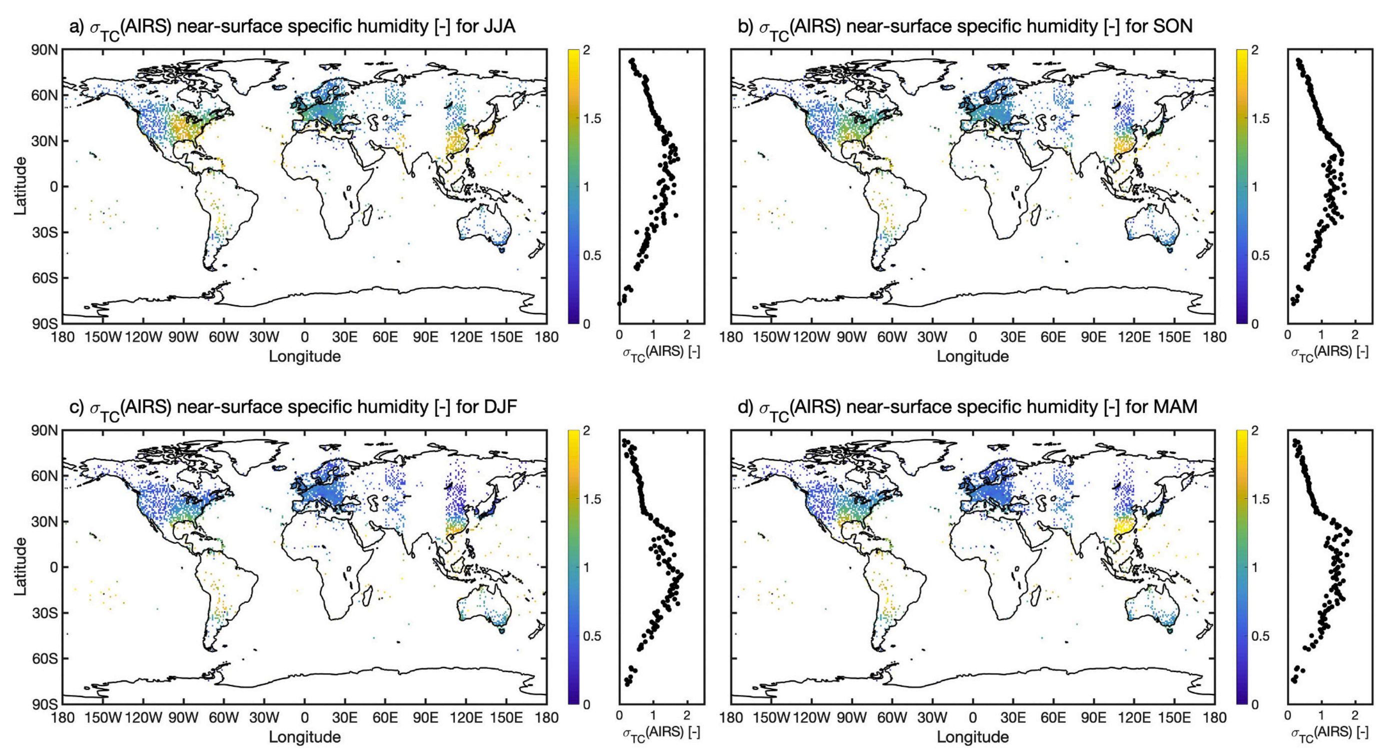

Fig. 7. Global maps and latitudinal averages of triple collocation (TC)-estimated noise error standard deviations in AIRS retrievals of deseasonalized near-surface specific humidity over land for a) June–August (JJA) b) September–November (SON) c) December–February (DJF) d) March–May (MAM). HadISD, AIRS and MERRA2 data at the ascending time were used in the triple collocation analysis for this figure.

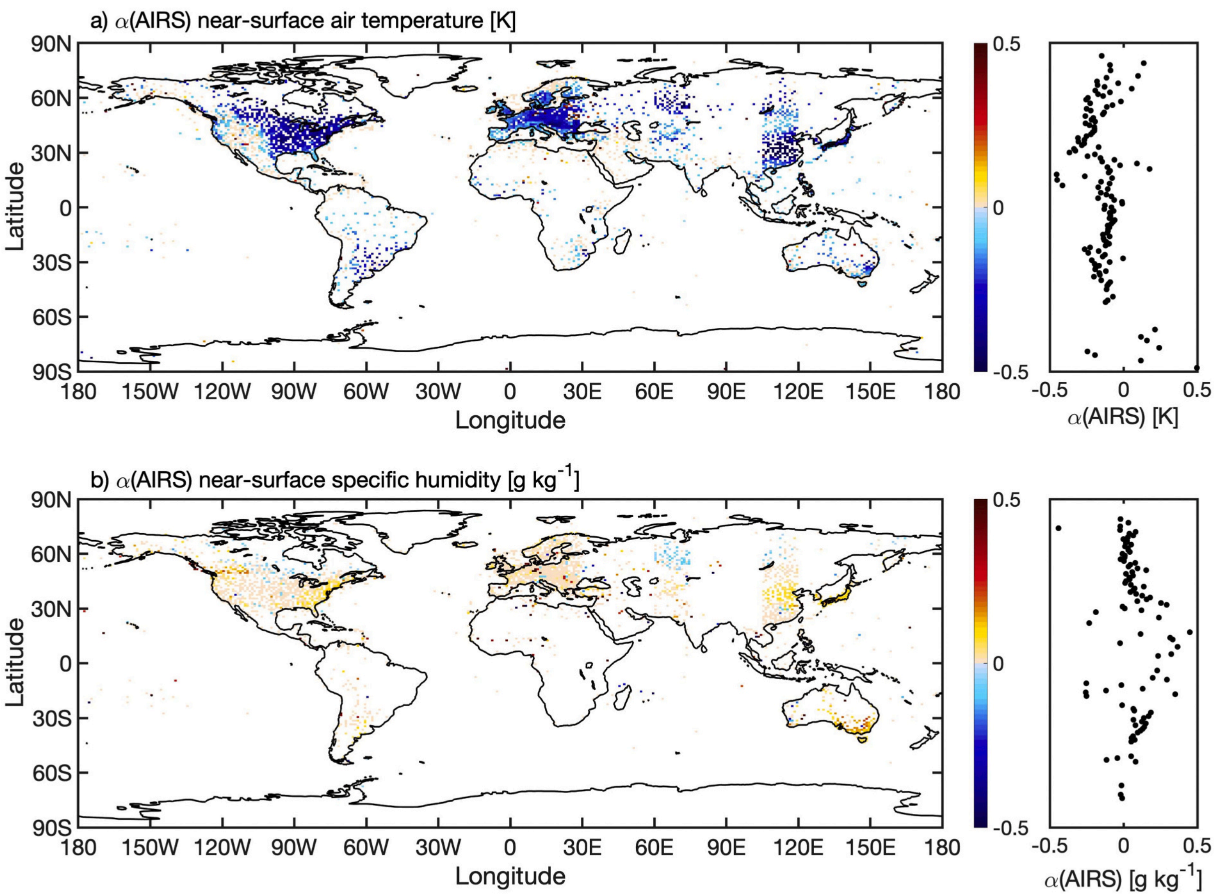

Fig. 8. Global maps and latitudinal averages of triple collocation (TC)-estimated additive biases in AIRS retrievals of deseasonalized a) near-surface air temperature and b) near-surface specific humidity over land, using HadISD, AIRS and MERRA2 data at the ascending time. Estimated values that were not statistically significantly different from zero were manually set to zero.

warranted. It also motivates the development of new techniques for estimating ET in the tropics.

Errors in retrievals of near-surface air temperature and specific humidity in the tropics can significantly impact satellite-derived estimates of tropical ET. Although a full error propagation analysis is beyond the scope of this study, we provide a first-order estimate of the induced errors in ET for typical conditions in the tropics. A typical set of conditions in the tropics (Fisher et al., 2009) are used as a reference state: available energy (the difference between net radiation and ground heat flux) $R_n - G = 200$ W/m^2, aerodynamic conductance $g_a = 1/50$ m/s, surface conductance $g_s = 1/100$ m/s, surface pressure $P = 101{,}325$ Pa, near-surface air temperature $T_a = 300$ K, and relative humidity RH $= 0.7$. An ensemble ($N = 10{,}000$) of conditions are generated by adding independent Gaussian zero-mean errors to the reference near-surface air temperature and specific humidity, with standard deviations of 1 K and 1 g/kg, respectively, consistent with typical values estimated in this study. No errors are added to the other reference variables. The ensemble of reference conditions is then used to generate an ensemble of ET estimates using standard bulk flux gradient equations and the surface energy budget. The mean of the ensemble of ET estimates is equal to the synthetic true value: 167 W/m^2. The standard deviation of the resulting ensemble of ET estimates is 16 W/m^2, which represents uncertainty in the estimate due to random errors in near-surface air temperature and specific humidity. This estimate of ET error does not include the effects of biases in air temperature and specific humidity, or errors of any kind in other input forcings (net radiation, ground heat flux or pressure) or parameters (surface conductance and aerodynamic conductance). A recent global intercomparison of different ET estimates found typical root-mean-squared-errors of 21–56 W/m^2 (Michel et al., 2016). While these numbers are not directly comparable, the comparison suggests that errors in estimates of near-surface air temperature and specific humidity will contribute substantially to total ET errors in the tropics. However, additional analyses are required to fully characterize the impact of errors on ET over the full range of conditions, which is left to future work.

Outside the tropics, there are many regions in which the AIRS retrievals perform well. There is significant potential in these regions for estimating ET using satellite retrievals of near-surface air temperature and specific humidity (e.g., Martens et al., 2017). While ET is instantaneously a function of more than just these two variables, recent work suggests that at daily and longer time scales, near-surface air temperature and specific humidity explain most of the observed variability in evaporative fraction (McColl et al., 2019b; McColl and Rigden, 2020; Salvucci and Gentine, 2013). In addition, outside the tropics, AIRS retrievals have the potential to better constrain land-atmosphere coupling at scales relevant to models (Roundy and Santanello, 2017). For example, satellite air temperature retrievals could be combined with satellite soil moisture observations to estimate soil moisture-air temperature correlations in regions at mid-latitudes with significant soil moisture memory (Koster and Suarez, 2001; McColl et al., 2017a, 2017b, 2019a; Seneviratne and Koster, 2011) and potential for land-atmosphere feedbacks (Koster et al., 2004; Tuttle and Salvucci, 2016).

5. Summary and Conclusions

This study has evaluated the performance of AIRS retrievals of near-surface air temperature and specific humidity over land. Our evaluation is novel in at least two respects. First, to our knowledge, this is the first

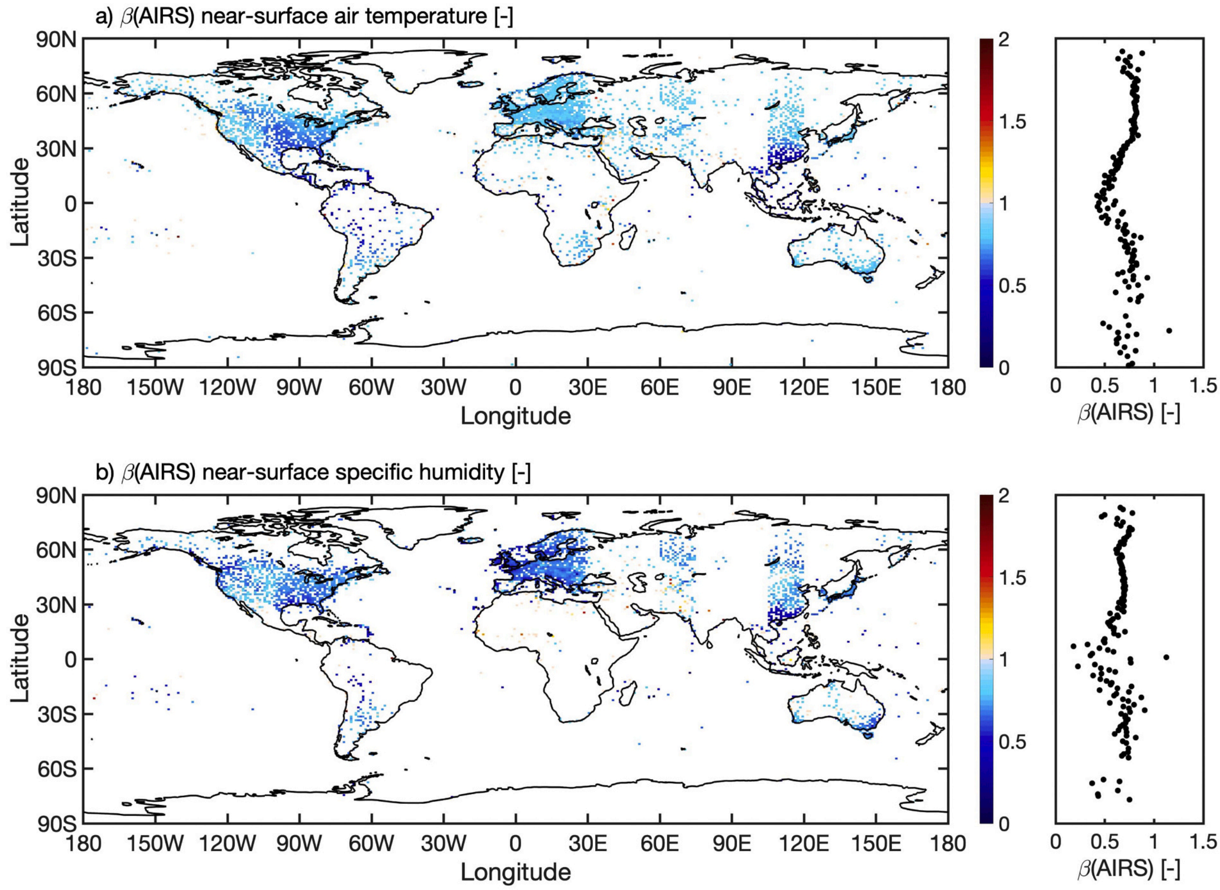

Fig. 9. Global maps and latitudinal averages of triple collocation (TC)-estimated multiplicative biases in AIRS retrievals of deseasonalized a) near-surface air temperature and b) near-surface specific humidity over land, using HadISD, AIRS and MERRA2 data at the ascending time. Estimated values that were not statistically significantly different from one were manually set to one.

study to apply triple collocation to evaluating retrievals of near-surface air temperature and specific humidity. Second, it is the first evaluation study of any kind of AIRS near-surface atmospheric measurements that is both global (rather than specific to a particular site or region) and spatially resolved (rather than averaging results, for example, over all land surfaces). The novel aspects of the study's methodology allow us to reach the main new finding of this study: AIRS retrievals of the near-surface atmospheric state are less accurate in the tropics compared to higher latitudes, at least with respect to the correlation coefficient and temporal sensitivity, even after removing the seasonal cycle. We provide several plausible reasons for why this might be expected, including higher uncertainties due to clouds, surface emissivity and the weak correlation between the near-surface atmosphere and the free troposphere in the tropics. Finally, implications are discussed for ET products that use AIRS as inputs to estimate ET in the tropics. While further studies are required to comprehensively quantify the impact of errors in AIRS retrievals on relevant ET products, a first-order estimate suggests that errors of around 10% should be expected in the tropics, solely due to random noise error. Including additive and multiplicative biases, and errors in other inputs, will increase the expected error. Since ET is greatest in the tropics, and tropical measurement networks are particularly sparse in that region, this work motivates new approaches for measuring ET in the tropics.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Thanks to Daniel Short Gianotti, Antonia Gambacorta, and the anonymous reviewers for providing helpful feedback. This work was supported by the National Basic Research Program of China (Grant No. 2017YFA0603703). K.A.M. acknowledges funding from a Winokur Seed Grant in the Environmental Sciences from Harvard University's Center for the Environment. All data used in this study are publicly-available. Code for performing TC is available at https://github.com/kaighin/ETC.

Appendix A. Impact of error cross-correlation on AIRS temporal sensitivity estimates

In this section, we examine how positive error cross-correlation between AIRS and the reanalysis data ($\mathrm{Cov}(\varepsilon_2, \varepsilon_3) > 0$) could impact estimates of AIRS temporal sensitivity ($\widehat{\beta}_2$). Subscripts 1, 2 and 3 refer to the HadISD station data, AIRS observations and reanalysis, respectively.

Relaxing the assumption that $\mathrm{Cov}(\varepsilon_2, \varepsilon_3) = 0$, the standard triple-collocation estimate for $\widehat{\beta}_2$ (McColl et al., 2014) becomes

$$\widehat{\beta}_2 = \frac{Q_{23}}{Q_{13}} = \frac{\beta_2\beta_3\sigma_T^2 + \text{Cov}(\varepsilon_2, \varepsilon_3)}{\beta_3\sigma_T^2}$$

It is clear from this equation that the estimate is unbiased when $\text{Cov}(\varepsilon_2, \varepsilon_3) = 0$. However, if the error covariance is positive, the temporal sensitivity estimate $\widehat{\beta}_2$ is positively-biased:

$$\widehat{\beta}_2 = \frac{Q_{23}}{Q_{13}} = \beta_2 + \frac{\text{Cov}(\varepsilon_2, \varepsilon_3)}{\beta_3\sigma_T^2} > \beta_2$$

In the tropics, natural variability (σ_T^2) is more likely to be systematically lower than at higher latitudes, rather than higher (even after removing the seasonal cycle). From the above equation, this implies that, if anything, positive error cross-correlation would cause TC-estimated $\widehat{\beta}_2$ to be *higher* in the tropics compared to outside the tropics. Since we observe $\widehat{\beta}_2$ to be *lower* in the tropics in Fig. 9, this pattern is unlikely to be an artifact caused by violations of the assumptions of TC.

References

Alemohammad, S.H., McColl, K.A., Konings, A.G., Entekhabi, D., Stoffelen, A., 2015. Characterization of precipitation product errors across the United States using multiplicative triple collocation. Hydrol. Earth Syst. Sci. 19, 3489–3503. https://doi.org/10.5194/hess-19-3489-2015.

Aumann, H.H., Chahine, M.T., Gautier, C., Goldberg, M.D., Kalnay, E., McMillin, L.M., Revercomb, H., Rosenkranz, P.W., Smith, W.L., Staelin, D.H., Strow, L.L., Susskind, J., 2003. AIRS/AMSU/HSB on the aqua mission: design, science objectives, data products, and processing systems. IEEE Trans. Geosci. Remote Sens. 41, 253–264. https://doi.org/10.1109/TGRS.2002.808356.

Badgley, G., Fisher, J.B., Jiménez, C., Tu, K.P., Vinukollu, R., 2015. On uncertainty in global terrestrial evapotranspiration estimates from choice of input forcing datasets. J. Hydrometeorol. 16, 1449–1455. https://doi.org/10.1175/JHM-D-14-0040.1.

Beer, R., 2006. TES on the Aura mission: scientific objectives, measurements, and analysis overview. IEEE Trans. Geosci. Remote Sens. 44, 1102–1105. https://doi.org/10.1109/TGRS.2005.863716.

Berrisford, P., Kållberg, P., Kobayashi, S., Dee, D., Uppala, S., Simmons, A.J., Poli, P., Sato, H., 2011. Atmospheric conservation properties in ERA-interim. Q. J. R. Meteorol. Soc. 137, 1381–1399. https://doi.org/10.1002/qj.864.

Blackwell, W.J., Milstein, A.B., 2014. A neural network retrieval technique for high-resolution profiling of cloudy atmospheres. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7, 1260–1270. https://doi.org/10.1109/JSTARS.2014.2304701.

Budyko, M.I., Berlyand, T.G., Yefimova, N.A., Zubenok, L.I., Strokina, L.A., 1980. Heat Balance of the Earth. NASA.

Chahine, M.T., Pagano, T.S., Aumann, H.H., Atlas, R., Barnet, C., Blaisdell, J., Chen, L., Divakarla, M., Fetzer, E.J., Goldberg, M., Gautier, C., Granger, S., Hannon, S., Irion, F.W., Kakar, R., Kalnay, E., Lambrigtsen, B.H., Lee, S.-Y., Le Marshall, J., Mcmillan, W.W., Mcmillin, L., Olsen, E.T., Revercomb, H., Rosenkranz, P., Smith, W. L., Staelin, D., Strow, L.L., Susskind, J., Tobin, D., Wolf, W., Zhou, L., 2006. AIRS: improving weather forecasting and providing new data on greenhouse house gases. Bull. Am. Meteorol. Soc. 87, 911–926. https://doi.org/10.1175/BAMS-87-7-911.

Dang, H.V.T., Lambrigtsen, B., Manning, E., 2017. AIRS/AMSU/HSB Version 6 Level 2 Performance and Test Report. Jet Propulsion Laboratory, California Institute of Technology.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., Vitart, F., 2011. The ERA-interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137, 553–597. https://doi.org/10.1002/qj.828.

Divakarla, M.G., Barnet, C.D., Goldberg, M.D., McMillin, L.M., Maddy, E., Wolf, W., Zhou, L., Liu, X., 2006. Validation of atmospheric infrared sounder temperature and water vapor retrievals with matched radiosonde measurements and forecasts. J. Geophys. Res. Atmos. 111. https://doi.org/10.1029/2005JD006116.

Draper, C., Reichle, R., de Jeu, R., Naeimi, V., Parinussa, R., Wagner, W., 2013. Estimating root mean square errors in remotely sensed soil moisture over continental scale domains. Remote Sens. Environ. 137, 288–298. https://doi.org/10.1016/j.rse.2013.06.013.

Dunn, R.J.H., Willett, K.M., Thorne, P.W., Woolley, E.V., Durre, I., Dai, A., Parker, D.E., Vose, R.S., 2012. HadISD: a quality-controlled global synoptic report database for selected variables at long-term stations from 1973–2011. Clim. Past 8, 1649–1679. https://doi.org/10.5194/cp-8-1649-2012.

Dunn, R.J.H., Willett, K.M., Morice, C.P., Parker, D.E., 2014. Pairwise homogeneity assessment of HadISD. Clim. Past 10, 1501–1522. https://doi.org/10.5194/cp-10-1501-2014.

Dunn, R.J.H., Willett, K.M., Parker, D.E., Mitchell, L., 2016. Expanding HadISD: quality-controlled, sub-daily station data from 1931. Geosci. Instrum. Methods Data Syst. 5, 473–491. https://doi.org/10.5194/gi-5-473-2016.

Efron, B., Tibshirani, R.J., 1994. An Introduction to the Bootstrap. CRC Press.

Emanuel, K., 2018. Inferences from simple models of slow, convectively coupled processes. J. Atmos. Sci. https://doi.org/10.1175/JAS-D-18-0090.1.

Famiglietti, C.A., Fisher, J.B., Halverson, G., Borbas, E.E., 2018. Global validation of MODIS near-surface air and dew point temperatures. Geophys. Res. Lett. https://doi.org/10.1029/2018GL077813.

Ferguson, C.R., Wood, E.F., 2010. An evaluation of satellite remote sensing data products for land surface hydrology: atmospheric infrared sounder. J. Hydrometeorol. 11, 1234–1262. https://doi.org/10.1175/2010JHM1217.1.

Fisher, J.B., Tu, K.P., Baldocchi, D.D., 2008. Global estimates of the land–atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. Remote Sens. Environ. 112, 901–919. https://doi.org/10.1016/j.rse.2007.06.025.

Fisher, J.B., Malhi, Y., Bonal, D., Rocha, H.R.D., Araújo, A.C.D., Gamo, M., Goulden, M. L., Hirano, T., Huete, A.R., Kondo, H., Kumagai, T., Loescher, H.W., Miller, S., Nobre, A.D., Nouvellon, Y., Oberbauer, S.F., Panuthai, S., Roupsard, O., Saleska, S., Tanaka, K., Tanaka, N., Tu, K.P., Randow, C.V., 2009. The land–atmosphere water flux in the tropics. Glob. Chang. Biol. 15, 2694–2714. https://doi.org/10.1111/j.1365-2486.2008.01813.x.

Friedlingstein, P., Meinshausen, M., Arora, V.K., Jones, C.D., Anav, A., Liddicoat, S.K., Knutti, R., 2013. Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. J. Clim. 27, 511–526. https://doi.org/10.1175/JCLI-D-12-00579.1.

Gao, W., Zhao, F., Xu, Y., Feng, X., 2008. Validation of the surface air temperature products retrieved from the atmospheric infrared sounder over China. IEEE Trans. Geosci. Remote Sens. 46, 1783–1789. https://doi.org/10.1109/TGRS.2008.916640.

Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A.M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J.E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S.D., Sienkiewicz, M., Zhao, B., 2017. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). J. Clim. 30, 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1.

Gettelman, A., Collins, W.D., Fetzer, E.J., Eldering, A., Irion, F.W., Duffy, P.B., Bala, G., 2006. Climatology of upper-tropospheric relative humidity from the atmospheric infrared sounder and implications for climate. J. Clim. 19, 6104–6121. https://doi.org/10.1175/JCLI3956.1.

Giardina, F., Konings, A.G., Kennedy, D., Alemohammad, S.H., Oliveira, R.S., Uriarte, M., Gentine, P., 2018. Tall Amazonian forests are less sensitive to precipitation variability. Nat. Geosci. 11, 405. https://doi.org/10.1038/s41561-018-0133-5.

Green, J.K., Seneviratne, S.I., Berg, A.M., Findell, K.L., Hagemann, S., Lawrence, D.M., Gentine, P., 2019. Large influence of soil moisture on long-term terrestrial carbon uptake. Nature 565, 476. https://doi.org/10.1038/s41586-018-0848-x.

Gruber, A., Su, C.-H., Zwieback, S., Crow, W., Dorigo, W., Wagner, W., 2016. Recent advances in (soil moisture) triple collocation analysis. In: Int. J. Appl. Earth Obs. Geoinformation, Advances in the Validation and Application of Remotely Sensed Soil Moisture - Part 1 45, Part B, pp. 200–211. https://doi.org/10.1016/j.jag.2015.09.002.

Hearty, T.J., Lee, J.N., Wu, D.L., Cullather, R., Blaisdell, J.M., Susskind, J., Nowicki, S.M. J., 2018. Intercomparison of surface temperatures from AIRS, MERRA, and MERRA-2 with NOAA and GC-net weather stations at summit, greenland. J. Appl. Meteorol. Climatol. 57, 1231–1245. https://doi.org/10.1175/JAMC-D-17-0216.1.

Holloway, C.E., Neelin, J.D., 2007. The convective cold top and quasi equilibrium. J. Atmos. Sci. 64, 1467–1487. https://doi.org/10.1175/JAS3907.1.

Jiang, J.H., Su, H., Zhai, C., Perun, V.S., Genio, A.D., Nazarenko, L.S., Donner, L.J., Horowitz, L., Seman, C., Cole, J., Gettelman, A., Ringer, M.A., Rotstayn, L., Jeffrey, S., Wu, T., Brient, F., Dufresne, J.-L., Kawai, H., Koshiro, T., Watanabe, M., LÉcuyer, T.S., Volodin, E.M., Iversen, T., Drange, H., Mesquita, M.D.S., Read, W.G., Waters, J.W., Tian, B., Teixeira, J., Stephens, G.L., 2012. Evaluation of cloud and water vapor simulations in CMIP5 climate models using NASA "A-train" satellite observations. J. Geophys. Res. Atmos. 117 https://doi.org/10.1029/2011JD017237.

Jiang, J.H., Su, H., Zhai, C., Wu, L., Minschwaner, K., Molod, A.M., Tompkins, A.M., 2015. An assessment of upper troposphere and lower stratosphere water vapor in MERRA, MERRA2, and ECMWF reanalyses using Aura MLS observations. J. Geophys. Res. Atmos. 120 (11), 468–11,485. https://doi.org/10.1002/2015JD023752.

Koster, R.D., Suarez, M.J., 2001. Soil moisture memory in climate models. J. Hydrometeorol. 2, 558–570. https://doi.org/10.1175/1525-7541(2001)002<0558:SMMICM>2.0.CO;2.

Koster, R.D., Dirmeyer, P.A., Guo, Z., Bonan, G., Chan, E., Cox, P., Gordon, C.T., Kanae, S., Kowalczyk, E., Lawrence, D., Liu, P., Lu, C.-H., Malyshev, S., McAvaney, B., Mitchell, K., Mocko, D., Oki, T., Oleson, K., Pitman, A., Sud, Y.C., Taylor, C.M., Verseghy, D., Vasic, R., Xue, Y., Yamada, T., 2004. Regions of strong coupling between soil moisture and precipitation. Science 305, 1138–1140. https://doi.org/10.1126/science.1100217.

Lyu, H., McColl, K.A., Li, X., Derksen, C., Berg, A., Black, T.A., Euskirchen, E., Loranty, M., Pulliainen, J., Rautiainen, K., Rowlandson, T., Roy, A., Royer, A., Langlois, A., Stephens, J., Lu, H., Entekhabi, D., 2018. Validation of the SMAP freeze/thaw product using categorical triple collocation. Remote Sens. Environ. 205, 329–337. https://doi.org/10.1016/j.rse.2017.12.007.

Ma, Klein, Xie, S., Zhang, Tang, Tang, Q., Morcrette, C.J., Van Weverberg, K., Petch, J., Ahlgrimm, M., Berg, L.K., Cheruy, F., Cole, J., Forbes, R., Gustafson, W.I., Huang, M., Liu, Y., Merryfield, W., Qian, Y., Roehrig, R., Wang, Y.-C., 2018. CAUSES: on the role of surface energy budget errors to the warm surface air temperature error over the central United States. J. Geophys. Res. Atmos. https://doi.org/10.1002/2017JD027194.

Maddy, E.S., Barnet, C.D., 2008. Vertical resolution estimates in version 5 of AIRS operational retrievals. IEEE Trans. Geosci. Remote Sens. 46, 2375–2384. https://doi.org/10.1109/TGRS.2008.917498.

Mallick, K., Jarvis, A., Wohlfahrt, G., Kiely, G., Hirano, T., Miyata, A., Yamamoto, S., Hoffmann, L., 2015. Components of near-surface energy balance derived from satellite soundings – part 1: noontime net available energy. Biogeosciences 12, 433–451. https://doi.org/10.5194/bg-12-433-2015.

Martens, B., Miralles, D.G., Lievens, H., van der Schalie, R., de Jeu, R.A.M., Fernández-Prieto, D., Beck, H.E., Dorigo, W.A., Verhoest, N.E.C., 2017. GLEAM v3: satellite-based land evaporation and root-zone soil moisture. Geosci. Model Dev. 10, 1903–1925. https://doi.org/10.5194/gmd-10-1903-2017.

McColl, K.A., Rigden, A.J., 2020. Emergent Simplicity of Continental Evapotranspiration. Geophys. Res. Lett. 47 https://doi.org/10.1029/2020GL087101 e2020GL087101.

McColl, K.A., Vogelzang, J., Konings, A.G., Entekhabi, D., Piles, M., Stoffelen, A., 2014. Extended triple collocation: estimating errors and correlation coefficients with respect to an unknown target. Geophys. Res. Lett. 41 https://doi.org/10.1002/2014GL061322, 2014GL061322.

McColl, K.A., Roy, A., Derksen, C., Konings, A.G., Alemohammed, S.H., Entekhabi, D., 2016. Triple collocation for binary and categorical variables: application to validating landscape freeze/thaw retrievals. Remote Sens. Environ. 176, 31–42. https://doi.org/10.1016/j.rse.2016.01.010.

McColl, K.A., Alemohammad, S.H., Akbar, R., Konings, A.G., Yueh, S., Entekhabi, D., 2017a. The global distribution and dynamics of surface soil moisture. Nat. Geosci. 10, 100. https://doi.org/10.1038/ngeo2868.

McColl, K.A., Wang, W., Peng, B., Akbar, R., Short Gianotti, D.J., Lu, H., Pan, M., Entekhabi, D., 2017b. Global characterization of surface soil moisture drydowns. Geophys. Res. Lett. https://doi.org/10.1002/2017GL072819, 2017GL072819.

McColl, K.A., He, Q., Lu, H., Entekhabi, D., 2019a. Short-term and long-term surface soil moisture memory time scales are spatially anticorrelated at global scales. J. Hydrometeorol. 20, 1165–1182. https://doi.org/10.1175/JHM-D-18-0141.1.

McColl, K.A., Salvucci, G.D., Gentine, P., 2019b. Surface flux equilibrium theory explains an empirical estimate of water-limited daily evapotranspiration. J. Adv. Model. Earth Syst. 11, 2036–2049. https://doi.org/10.1029/2019MS001685.

Menzel, W.P., Schmit, T.J., Zhang, P., Li, J., 2018. Satellite-based atmospheric infrared sounder development and applications. Bull. Am. Meteorol. Soc. 99, 583–603. https://doi.org/10.1175/BAMS-D-16-0293.1.

Michel, D., Jiménez, C., Miralles, D.G., Jung, M., Hirschi, M., Ershadi, A., Martens, B., McCabe, M.F., Fisher, J.B., Mu, Q., Seneviratne, S.I., Wood, E.F., Fernández-Prieto, D., 2016. The WACMOS-ET project – Part 1: tower-scale evaluation of four remote-sensing-based evapotranspiration algorithms. Hydrol. Earth Syst. Sci. 20, 803–822. https://doi.org/10.5194/hess-20-803-2016.

Milstein, A.B., 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. https://doi.org/10.1002/2015JD024008.

Molod, A., Takacs, L., Suarez, M., Bacmeister, J., 2015. Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2. Geosci. Model Dev. 8, 1339–1356. https://doi.org/10.5194/gmd-8-1339-2015.

Mueller, B., Seneviratne, S.I., 2014. Systematic land climate and evapotranspiration biases in CMIP5 simulations. Geophys. Res. Lett. 41, 128–134. https://doi.org/10.1002/2013GL058055.

Prakash, S., Shati, F., Norouzi, H., Blake, R., 2019. Observed differences between near-surface air and skin temperatures using satellite and ground-based data. Theor. Appl. Climatol. 137, 587–600. https://doi.org/10.1007/s00704-018-2623-1.

Roebeling, R.A., Wolters, E.L.A., Meirink, J.F., Leijnse, H., 2012. Triple collocation of summer precipitation retrievals from SEVIRI over Europe with gridded rain gauge and weather radar data. J. Hydrometeorol. 13, 1552–1566. https://doi.org/10.1175/JHM-D-11-089.1.

Roundy, J.K., Santanello, J.A., 2017. Utility of satellite remote sensing for land–atmosphere coupling and drought metrics. J. Hydrometeorol. 18, 863–877. https://doi.org/10.1175/JHM-D-16-0171.1.

Salvucci, G.D., Gentine, P., 2013. Emergent relation between surface vapor conductance and relative humidity profiles yields evaporation rates from weather data. Proc. Natl. Acad. Sci. 110, 6287–6291. https://doi.org/10.1073/pnas.1215844110.

Seneviratne, S.I., Koster, R.D., 2011. A revised framework for analyzing soil moisture memory in climate data: derivation and interpretation. J. Hydrometeorol. 13, 404–412. https://doi.org/10.1175/JHM-D-11-044.1.

Seneviratne, S.I., Corti, T., Davin, E.L., Hirschi, M., Jaeger, E.B., Lehner, I., Orlowsky, B., Teuling, A.J., 2010. Investigating soil moisture–climate interactions in a changing climate: a review. Earth-Sci. Rev. 99, 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004.

Sobel, A.H., Nilsson, J., Polvani, L.M., 2001. The weak temperature gradient approximation and balanced tropical moisture waves. J. Atmos. Sci. 58, 3650–3665. https://doi.org/10.1175/1520-0469(2001)058<3650:TWTGAA>2.0.CO;2.

Stoffelen, A., 1998. Toward the true near-surface wind speed: error modeling and calibration using triple collocation. J. Geophys. Res. 103, 7755–7766.

Susskind, J., Barnet, C.D., Blaisdell, J.M., 2003. Retrieval of atmospheric and surface parameters from AIRS/AMSU/HSB data in the presence of clouds. IEEE Trans. Geosci. Remote Sens. 41, 390–409. https://doi.org/10.1109/TGRS.2002.808236.

Susskind, J., Blaisdell, J.M., Iredell, L., Keita, F., 2011. Improved temperature sounding and quality control methodology using AIRS/AMSU data: the AIRS science team version 5 retrieval algorithm. IEEE Trans. Geosci. Remote Sens. 49, 883–907. https://doi.org/10.1109/TGRS.2010.2070508.

Susskind, J., Blaisdell, J.M., Iredell, L., 2014. Improved methodology for surface and atmospheric soundings, error estimates, and quality control procedures: the atmospheric infrared sounder science team version-6 retrieval algorithm. J. Appl. Remote. Sens. 8, 084994 https://doi.org/10.1117/1.JRS.8.084994.

Susskind, J., Schmidt, G.A., Lee, J.N., Iredell, L., 2019. Recent global warming as confirmed by AIRS. Environ. Res. Lett. 14, 044030 https://doi.org/10.1088/1748-9326/aafd4e.

Tian, B., Fetzer, E.J., Kahn, B.H., Teixeira, J., Manning, E., Hearty, T., 2013. Evaluating CMIP5 models using AIRS tropospheric air temperature and specific humidity climatology. J. Geophys. Res. Atmos. 118, 114–134. https://doi.org/10.1029/2012JD018607.

Tobin, D.C., Revercomb, H.E., Knuteson, R.O., Lesht, B.M., Strow, L.L., Hannon, S.E., Feltz, W.F., Moy, L.A., Fetzer, E.J., Cress, T.S., 2006. Atmospheric radiation measurement site atmospheric state best estimates for atmospheric infrared sounder temperature and water vapor retrieval validation. J. Geophys. Res. Atmos. 111. https://doi.org/10.1029/2005JD006103.

Tuttle, S., Salvucci, G., 2016. Empirical evidence of contrasting soil moisture–precipitation feedbacks across the United States. Science 352, 825–828. https://doi.org/10.1126/science.aaa7185.

Vinukollu, R.K., Wood, E.F., Ferguson, C.R., Fisher, J.B., 2011. Global estimates of evapotranspiration for climate studies using multi-sensor remote sensing data: evaluation of three process-based approaches. Remote Sens. Environ. 115, 801–823. https://doi.org/10.1016/j.rse.2010.11.006.

Vogelzang, J., Stoffelen, A., Verhoef, A., Figa-Saldaña, J., 2011. On the quality of high-resolution scatterometer winds. J. Geophys. Res. 116 https://doi.org/10.1029/2010JC006640.

Wehrli, K., Guillod, B.P., Hauser, M., Leclair, M., Seneviratne, S.I., 2018. Assessing the dynamic versus thermodynamic origin of climate model biases. Geophys. Res. Lett. 45, 8471–8479. https://doi.org/10.1029/2018GL079220.

Worden, J.R., Kulawik, S.S., Fu, D., Payne, V.H., Lipton, A.E., Polonsky, I., He, Y., Cady-Pereira, K., Moncet, J.-L., Herman, R.L., Irion, F.W., Bowman, K.W., 2019. Characterization and evaluation of AIRS-based estimates of the deuterium content of water vapor. Atmos. Meas. Tech. 12, 2331–2339. https://doi.org/10.5194/amt-12-2331-2019.

Wu, W., Dessler, A.E., North, G.R., 2006. Analysis of the correlations between atmospheric boundary-layer and free-tropospheric temperatures in the tropics. Geophys. Res. Lett. 33 https://doi.org/10.1029/2006GL026708.

Wulfmeyer, V., Hardesty, R.M., Turner, D.D., Behrendt, A., Cadeddu, M.P., Girolamo, P. D., Schlüssel, P., Baelen, J.V., Zus, F., 2015. A review of the remote sensing of lower tropospheric thermodynamic profiles and its indispensable role for the understanding and the simulation of water and energy cycles. Rev. Geophys. 53, 819–895. https://doi.org/10.1002/2014RG000476.

Yilmaz, M.T., Crow, W.T., 2014. Evaluation of assumptions in soil moisture triple collocation analysis. J. Hydrometeorol. https://doi.org/10.1175/JHM-D-13-0158.1.