JGR Atmospheres

RESEARCH ARTICLE
10.1029/2023JD040536

Special Collection:
Twenty Years of Observations from the Atmospheric Infrared Sounder

Assessing the Tropospheric Temperature and Humidity Simulations in CMIP3/5/6 Models Using the AIRS Obs4MIPs V2.1 Data

Baijun Tian[1], **Eric J. Fetzer**[1], **and Joao Teixeira**[1]

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Key Points:

- CMIP Phase 3, 5 and 6 climate models share similar tropospheric air temperature, specific and relative humidity biases relative to Atmospheric Infrared Sounder
- Cold biases and positive relative humidity biases are prevalent in free troposphere almost globally with maxima over the mid-latitude oceans
- Double-intertropical convergence zone bias-related specific and relative humidity biases are evident in tropical free troposphere: positive (negative) off (near) the equator

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

B. Tian,
baijun.tian@jpl.nasa.gov

Citation:

Tian, B., Fetzer, E. J., & Teixeira, J. (2024). Assessing the tropospheric temperature and humidity simulations in CMIP3/5/6 models using the AIRS Obs4MIPs V2.1 data. *Journal of Geophysical Research: Atmospheres*, *129*, e2023JD040536. https://doi.org/10.1029/2023JD040536

Received 5 DEC 2023
Accepted 25 JUL 2024

Author Contributions:

Conceptualization: Baijun Tian
Data curation: Baijun Tian
Formal analysis: Baijun Tian
Funding acquisition: Baijun Tian
Investigation: Baijun Tian
Methodology: Baijun Tian
Project administration: Baijun Tian
Resources: Baijun Tian
Software: Baijun Tian
Validation: Baijun Tian
Visualization: Baijun Tian
Writing – original draft: Baijun Tian

Abstract In this study, the Atmospheric Infrared Sounder (AIRS) Observations for Model Intercomparison Projects (Obs4MIPs) V2.1 tropospheric air temperature, specific humidity, and relative humidity data are utilized to evaluate the global tropospheric temperature and humidity simulations in the fully coupled global climate models from the Coupled Model Intercomparison Project phases 3, 5, and 6 (CMIP3, CMIP5, and CMIP6), and possible simulation improvement in CMIP6 models in comparison to CMIP3 and CMIP5 models. Our analyses indicate that all three phases of CMIP models share similar tropospheric air temperature, specific humidity, and relative humidity biases in their multi-model ensemble means relative to AIRS. Cold biases up to 4 K and positive relative humidity biases up to 20% are found in the free troposphere almost globally with maxima over the mid-latitude storm tracks. Warm biases up to 2 K are seen over the Southern Ocean in the lower troposphere. Positive specific and relative humidity biases exist over the off-equatorial oceans while negative specific and relative humidity biases are seen near the equator in the tropical free troposphere, which are related to the double-intertropical convergence zone bias in the models. Both the air temperature and specific humidity biases are important to the relative humidity biases except in the tropical free troposphere where the specific humidity biases dominate. The tropospheric air temperature, specific humidity, and relative humidity biases are reduced from CMIP3 to CMIP5 and to CMIP6 at almost all pressure levels except at 300 hPa for specific humidity and in the boundary layer for relative humidity.

Plain Language Summary In this study, we use the Atmospheric Infrared Sounder (AIRS) data to assess how well the fully coupled global climate models simulate temperature and humidity in the troposphere. We focus on the climate models from the recent three phases of Coupled Model Intercomparison Project (CMIP3, CMIP5, and CMIP6) and investigate if there is any bias reduction in the recent CMIP6 models compared to the earlier CMIP3 and CMIP5 models. To conduct our evaluation, we compare the CMIP model ensemble averages with the AIRS data and find that all three phases of CMIP models exhibit similar biases in tropospheric temperature and humidity. For example, cold biases up to 4 K and positive relative humidity biases up to 20% are seen in the free troposphere over most regions, with the maximum biases over the midlatitudes. Positive specific and relative humidity biases are found over the off-equatorial oceans while negative specific and relative humidity biases are seen near the equator in the tropical free troposphere. We also note a possible temperature and humidity bias reduction from the CMIP3 models to the CMIP5 models and then to the CMIP6 models at most pressure levels.

1. Introduction

The Coupled Model Intercomparison Project (CMIP) (Meehl et al., 1997, 2005) collects the standardized global climate model outputs from various global climate model groups and makes them publicly available through the Earth System Grid Federation (ESGF) data centers (Cinquini et al., 2014). The main objective of CMIP is to better understand the past, present, and future climate variations and changes arising from natural and unforced variability, and in response to changes in radiative forcing due to human activities in a multi-model context. Started in early 1990s, CMIP has undergone five phases and become a central component of national and international climate change assessments over the past three decades. The global climate model outputs from the CMIP Phase 3 (CMIP3, Meehl et al., 2007), Phase 5 (CMIP5, Taylor et al., 2012), and Phase 6 (CMIP6, Eyring et al., 2016) have formed the foundation of the Intergovernmental Panel on Climate Change (IPCC) Fourth, Fifth, and Sixth Assessment Reports (AR4, AR5, and AR6) (IPCC, 2007, 2013, 2021), respectively.

Writing – review & editing: Baijun Tian, Eric J. Fetzer, Joao Teixeira

The Observations for Model Intercomparison Projects (Obs4MIPs) (Ferraro et al., 2015; Teixeira et al., 2014; Waliser et al., 2020) was established in 2010 and aims to facilitate the effective utilization of satellite observations for CMIP climate model evaluation and research. To achieve this goal, Obs4MIPs follows the following four specific strategies: (a) It focuses on observed variables that can be directly compared to CMIP model variables. (b) It adheres to data set formatting specifications and metadata requirements that closely align with CMIP model outputs. (c) It provides concise technical documentation for each data set that is tailored to non-experts and includes information on data set uncertainty, strengths, limitations, and relevance for model evaluation. (d) It ensures the satellite data to be easily searched and accessed alongside the climate model outputs by disseminating the satellite data through the ESGF platforms.

The Atmospheric Infrared Sounder (AIRS)/Advanced Microwave Sounding Unit (AMSU) is a key suite of instruments for measuring atmospheric temperature and humidity on NASA's Aqua satellite (Aumann et al., 2003; Chahine et al., 2006). Since the launch of the Aqua satellite on 4 May 2002, AIRS/AMSU has provided a two-decade long high-quality atmospheric temperature and humidity sounding data record that has been widely used by the weather and climate communities. As part of the Obs4MIPs project, the AIRS Obs4MIPs data are the AIRS/AMSU atmospheric temperature and humidity sounding data designed specifically for CMIP climate model evaluation and research (Tian & Hearty, 2020; Tian et al., 2013a, 2019). Three versions of the AIRS Obs4MIPs data sets have been published on the ESGF data centers for public use: V1.0 (Tian et al., 2013a), V2.0 (Tian et al., 2019), and V2.1 (Tian & Hearty, 2020). The AIRS Obs4MIPs V2.1 data are the latest version and include the sampling-bias-corrected monthly mean tropospheric air temperature (ta), specific humidity (hus), and relative humidity (hur) profiles from September 2002 to September 2016 (Tian & Hearty, 2020). The AIRS Obs4MIPs V2.1 data have accounted for the sampling difference among the AIRS data, reanalysis data, and climate model outputs to make the AIRS Obs4MIPs V2.1 data, reanalysis data, and climate model outputs more directly comparable, and should be used for CMIP climate model evaluation and research (Tian & Hearty, 2020).

Previous studies have used the AIRS tropospheric air temperature and humidity data to evaluate the tropospheric temperature and humidity biases in the CMIP3 and CMIP5 models and have well demonstrated the scientific value of the AIRS tropospheric air temperature and humidity data in climate model evaluation (e.g., Fasullo & Trenberth, 2012; Gettelman et al., 2006; Jiang et al., 2012; John & Soden, 2007; Pierce et al., 2006; Su et al., 2014; Takahashi et al., 2013; Tian, 2015; Tian et al., 2013a). For example, Gettelman et al. (2006) compared the AIRS Version 3 (V3) Level-2 (L2) relative humidity data to a simulation from the Community Atmosphere Model version 3 (CAM3) and found that the CAM3 model reproduces realistic mean relative humidity distributions in comparison to AIRS but is slightly more moist than AIRS in the middle and upper troposphere. The CAM3 model also has difficulties reproducing many scales of AIRS observed relative humidity variability, particularly in the tropics. Pierce et al. (2006) and John and Soden (2007) compared the AIRS Version 4 (V4) Level 3 (L3) and reanalysis specific humidity profiles to those simulated from the CMIP3 models. They found that most CMIP3 models have a large moist bias in the free troposphere (more than 100%) especially over the extratropics, but a dry bias in the boundary layer (up to 25%) over the tropics. John and Soden (2007) also compared the AIRS V4 L3 and reanalysis tropospheric air temperature profiles to those simulated from the CMIP3 models. They found that the CMIP3 simulated temperatures are systematically colder than the observations by 1–2 K throughout the troposphere. This cold bias generally increases with altitude in the free troposphere, with maxima located near 200 hPa in the extratropics. The inter-model spread of the cold and moist biases in the CMIP3 models is also large (John & Soden, 2007). Tian et al. (2013a) evaluated the long-term mean tropospheric air temperature and specific humidity simulations in the CMIP5 models using the AIRS Obs4MIPs V1.0 data. A tropospheric cold bias of ∼2 K and the double-intertropical convergence zone (ITCZ) bias in the troposphere from 1000 hPa to 300 hPa were found in the CMIP5 models. The inter-model spread of the cold bias and the double-ITCZ bias in the CMIP5 models is also large (Tian et al., 2013a). Tian (2015) further quantified the double-ITCZ bias in the CMIP3 and CMIP5 models using the AIRS specific humidity data and the Tropical Rainfall Measurement Mission (TRMM)/Global Precipitation Climatology Project (GPCP) precipitation (pr) data and examined the connection of the double-ITCZ bias and the equilibrium climate sensitivity (ECS) in the CMIP3 and CMIP5 models. Tian (2015) found that the spatial patterns of the mid-tropospheric specific humidity and surface precipitation are similar over the tropical Pacific, and both are related to the double-ITCZ biases in the CMIP3 and CMIP5 models. Tian (2015) proposed a tropical mid-tropospheric humidity asymmetry index like the tropical precipitation asymmetry index to quantify the double-ITCZ bias and demonstrated that both the tropical mid-tropospheric humidity asymmetry index and the southern ITCZ index are linearly correlated to the ECS and are emergent constraints for ECS in the

CMIP3 and CMIP5 models. This indicates that the AIRS specific humidity data can help constrain model ECS and improve future climate projections. Su et al. (2014) evaluated the relative humidity simulations in the CMIP5 models using the AIRS Version 6 (V6) L3 data and found that the close interaction between the large-scale circulation and clouds is important for cloud feedback and climate sensitivity.

Tropospheric temperature and humidity biases are also expected in the CMIP6 models and thus need to be identified, quantified, and understood. However, very few studies have so far worked on this issue (Bock et al., 2020; Jiang et al., 2021). For example, Bock et al. (2020) quantified the tropospheric air temperature and specific humidity biases in the CMIP6 models using the Earth System Model Evaluation Tool (ESMValTool) v2.0 (Eyring et al., 2020; Lauer et al., 2020; Righi et al., 2020; Weigel et al., 2021) and the ERA-Interim reanalysis data (Dee et al., 2011). They found some improvements in zonally averaged air temperature and specific humidity simulations moving from the CMIP3 or CMIP5 models to the CMIP6 models. However, they did not examine the bias spatial patterns at different pressure levels, did not examine the relative humidity, and did not use the AIRS Obs4MIPs V2.1 data. Jiang et al. (2021) assessed the representation of clouds and water vapor structures in 28 CMIP6 models using NASA satellite data and found measurable improvements in the CMIP6 models relative to the CMIP5 models for both clouds and water vapor. However, their results focus on the tropical oceans only and use the AIRS V6 L3 data instead of the AIRS Obs4MIPs V2.1 data.

To the best of our knowledge, global tropospheric air temperature, specific humidity, and relative humidity biases in the CMIP6 models and the possible reductions of these biases from the CMIP3 or CMIP5 models to the CMIP6 models have not been systematically characterized, especially based on the AIRS Obs4MIPs V2.1 data. The objective of this study is to utilize the AIRS Obs4MIPs V2.1 data to evaluate the representation of global tropospheric air temperature, specific humidity, and relative humidity in the fully coupled global climate models that participated in the most recent three CMIP phases, namely CMIP3, CMIP5, and CMIP6, and to identify potential simulation improvement or bias reduction in the CMIP6 models compared to the CMIP3 and CMIP5 models. By conducting this comprehensive analysis across the three physical variables (air temperature, specific humidity, and relative humidity) and the three model generations (CMIP3, CMIP5, and CMIP6), we aim to provide insights into the strengths and weaknesses of the CMIP models in representing global tropospheric thermodynamics. The remainder of this paper is structured as follows. In Section 2, we provide a comprehensive overview of the data and methodology employed. Section 3 presents the main results, and a summary is provided in Section 4.

2. Data and Methodology

2.1. Satellite and Reanalysis Data

AIRS/AMSU is the NASA's atmospheric temperature and humidity sounding instrument suite on the Aqua satellite launched on 4 May 2002 (Aumann et al., 2003; Chahine et al., 2006). The AIRS and AMSU instruments are co-aligned cross-track scanning nadir sounders. Their combined operations began on 31 August 2002 and ended when the AMSU-A2 instrument stopped working on 24 September 2016. The AIRS instrument is a 2378-channel infrared grating spectrometer at wavelengths in the range 3.7–15.4 μm with a horizontal resolution of about 13.5 km at nadir (Aumann et al., 2003). The AMSU instrument is a 15-channel microwave radiometer with a horizontal resolution of about 45 km at nadir (Lambrigtsen, 2003). These infrared and microwave wavelengths are sensitive to atmospheric air temperature in the troposphere and stratosphere and specific humidity in the troposphere as well as clouds, precipitation, minor gases, and surface properties. The AIRS/AMSU retrieval algorithm uses an iterative least squares physical inversion method and a set of one AMSU microwave spectrum and nine associated AIRS infrared spectra, referred to as Level-1 (L1) products, for geophysical quantity retrievals (Susskind et al., 2014). The resulting AIRS/AMSU atmospheric air temperature and specific humidity are referred to as L2 products and have a temporal resolution of twice daily, a horizontal resolution of 45 km at nadir, and a vertical resolution of about 1 km for air temperature and 2 km for specific humidity (Susskind et al., 2003). The AIRS/AMSU atmospheric air temperature and specific humidity profiles have a near global coverage every day for cloud cover up to about 70% (Susskind et al., 2014). Their uncertainties are 1 K in 1-km layers for air temperature and 15% of mean specific humidity in 2-km layers for specific humidity and have been confirmed empirically through validation studies (Divakarla et al., 2006; Gettelman et al., 2004; Tobin et al., 2006). The AIRS/AMSU atmospheric relative humidity is a derived quantity based on the AIRS/AMSU retrieved atmospheric air temperature and specific humidity (Olsen et al., 2013). The relative humidity is calculated as the ratio

of atmospheric specific humidity and saturation specific humidity, equivalent to the ratio of water vapor pressure and saturation pressure, while the water vapor saturation pressure is calculated based on the retrieved atmospheric air temperature over equilibrium phase (liquid above 273.15 K and ice below 273.16 K) using the formula from Murphy and Koop (2005). The AIRS L3 data products are regularly averaged and gridded L2 swath products on global 1° × 1° latitude-longitude grids, at specified vertical pressure levels, and at specified daily or monthly temporal resolutions (Tian et al., 2013b).

The AIRS Obs4MIPs data are AIRS/AMSU L3 atmospheric temperature and humidity sounding data designed specifically for CMIP climate model evaluation and research (Tian & Hearty, 2020; Tian et al., 2013a, 2019). Three versions of the AIRS Obs4MIPs data sets have been generated and published on the ESGF data centers for public use: V1.0 (Tian et al., 2013a), V2.0 (Tian et al., 2019), and V2.1 (Tian & Hearty, 2020). These three versions of AIRS Obs4MIPs data sets share the common characteristics of a monthly mean temporal resolution, a global spatial resolution of 1°-longitude by 1°-latitude, and available on the eight CMIP mandatory vertical pressure levels ranging from 1000 hPa to 300 hPa (e.g., 1000, 925, 850, 700, 600, 500, 400, and 300 hPa). However, these three versions differ in terms of the variables provided, the time periods covered, and whether the sampling biases are corrected or not. The AIRS Obs4MIPs V1.0 data are based on the AIRS Version 5 (V5) L3 standard monthly AIRS/AMSU combined retrieval products and published in September 2011 (Tian et al., 2013a). They consist of monthly mean tropospheric air temperature (ta) and specific humidity (hus) profiles from September 2002 to May 2011. The AIRS Obs4MIPs V2.0 data were published in April 2018 and encompass monthly mean tropospheric air temperature (ta), specific humidity (hus), and relative humidity (hur) profiles spanning September 2002 to September 2016 (Tian et al., 2019). They are based on the AIRS V6 L3 standard monthly products in the "TqJoint" grids from the AIRS/AMSU combined retrievals. The V2.0 data added relative humidity (hur) and extended the time coverage of the AIRS Obs4MIPs data. The AIRS Obs4MIPs V2.1 data are the latest version and were published in August 2020. They include the sampling-bias-corrected monthly mean tropospheric air temperature (ta), specific humidity (hus), and relative humidity (hur) profiles from September 2002 to September 2016 (Tian & Hearty, 2020). A significant improvement in V2.1 relative to V2.0 is the removal of sampling biases in the AIRS Obs4MIPs V2.0 data. Due to the limitations of the Aqua satellite orbit (limited swath width and limited sampling of the diurnal cycle and synoptic events) and the AIRS retrieval algorithm (e.g., reduced sample in cloudy regions), large sampling biases exist in the AIRS Obs4MIPs V2.0 data (Tian et al., 2019). These sampling biases were estimated using the method of Hearty et al. (2014) and the fifth generation of the European Centre for Medium-Range Weather Forecasts (ECMWF) (ERA5) reanalysis (Hersbach et al., 2020), and removed from the AIRS Obs4MIPs V2.0 data to produce the sampling-bias-corrected AIRS Obs4MIPs V2.1 data (Tian & Hearty, 2020). As a result, the AIRS Obs4MIPs V2.1 data have accounted for the sampling difference among the AIRS data, reanalysis data, and climate model outputs to make the AIRS Obs4MIPs V2.1 data, reanalysis data, and climate model outputs directly comparable and should be used in CMIP climate model evaluation and research (Tian & Hearty, 2020).

Our previous studies (Hearty et al., 2014; Tian & Hearty, 2020; Tian et al., 2013a) have indicated that differences still exist between the AIRS Obs4MIPs V2.1 data and the reanalysis data, even after accounting for the sampling difference between the AIRS and reanalysis data. These differences are due to either errors in the AIRS data from their retrieval algorithm, or errors in the reanalysis data from their data assimilation systems. Thus, as a cross-check or validation of the AIRS Obs4MIPs V2.1 data, we also consider two state-of-the-art reanalyses: the fifth generation European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis (ERA5) and the Modern-Era Retrospective Analysis for Research and Application, Version 2 (MERRA-2). ERA5 is produced by ECMWF using the Integrated Forecasting System (IFS) Cycle 41r2 and available from 1979 to present on 37 pressure levels from 1000 hPa to 1 hPa and on a regular 0.25° longitude and latitude grid (Hersbach et al., 2020). MERRA-2 is the latest modern satellite era atmospheric reanalysis produced by NASA's Global Modeling and Assimilation Office (GMAO) and available on 42 pressure levels from 1000 to 0.1 hPa and on a regular 0.5° latitude by 0.625° longitude grid (Gelaro et al., 2017).

Please note that the AIRS Obs4MIPs data are derived from the AIRS L1 radiances and both ERA5 and MERRA-2 have assimilated the AIRS L1 radiances. Thus, ERA5, MERRA-2, and AIRS Obs4MIPs data are not "totally independent" with each other. However, there are so many differences in these three data sets after the AIRS radiances, especially the ERA5 and MERRA-2 depend on numerical models to provide information and use many other observations, these three data sets are unique and different.

2.2. CMIP Model Data

Monthly mean air temperature (ta), specific humidity (hus), and relative humidity (hur) outputs from the "twentieth century" (CMIP3) or "historical" (CMIP5 and CMIP6) experiments of 91 fully coupled global climate models that participated in CMIP3, CMIP5, and CMIP6 are used and analyzed here. These include 22 CMIP3 models, 35 CMIP5 models, and 34 CMIP6 models (Table S1 in the Supporting Information S1). Due to the unavailability of the relative humidity outputs from two CMIP3 models (CGCM3.1(T47) and CGCM3.1(T63)), only 20 CMIP3 models are used for relative humidity analysis. Following Tian and Dong (2020), the first ensemble member run of variant label "r1i1p1" for the CMIP3 and CMIP5 models and "r1i1p1f1" for the CMIP6 models are used if available. Here, we analyze only the fully coupled global climate models because they are the foundation for future climate prediction and IPCC AR. Evaluating the Atmospheric Model Intercomparison Project (AMIP) simulations forced with observed sea surface temperatures (SSTs) (Gates et al., 1999) is also important, but beyond the scope of the current study.

2.3. Analysis Methodology

Following Tian and Dong (2020) and Bock et al. (2020), we focus on assessing the possible progress made over the years of the most recent three generations or phases of CMIP models by evaluating Multi-Model Ensemble Means (MMEMs) of the CMIP3, CMIP5, and CMIP6 models rather than trying to trace the possible progress or improvement of specific individual CMIP3, CMIP5, and CMIP6 models. As the MMEM could be a result of compensated biases of individual models and there is a big spread among individual models (Tian, 2015; Tian et al., 2013a), the conclusions for the MMEMs may not necessarily be valid for individual models.

To highlight the spatial pattern and vertical structure of the tropospheric air temperature, specific humidity, and relative humidity, we also focus on the model bias (CMIP–AIRS) maps and global statistics of the long-term or climatological annual means of tropospheric air temperature, specific humidity, and relative humidity at the eight standard CMIP tropospheric pressure levels from 1000 hPa to 300 hPa. The long-term annual means of tropospheric air temperature, specific humidity, and relative humidity from AIRS, ERA5, and MERRA-2 are calculated from September 2002 to September 2016 and those from the CMIP models are calculated based on the last 20 years of CMIP model outputs, that is, January 1980−December 1999 for CMIP3, January 1986−December 2005 for CMIP5, and January 1995−December 2014 for CMIP6. The long-term annual means of tropospheric air temperature, specific humidity, and relative humidity for the CMIP3, CMIP5, and CMIP6 MMEMs are calculated by averaging the long-term annual means of tropospheric air temperature, specific humidity, and relative humidity from each CMIP3, CMIP5, and CMIP6 models. AIRS, reanalysis, and CMIP model data have different atmospheric spatial grid resolutions (Table S1 in the Supporting Information S1). Thus, all AIRS, reanalysis, and CMIP model data are re-gridded onto a common spatial grid of 2.5° × 2.5° for comparisons.

3. Results

3.1. Air Temperature (ta)

The climatological annual mean tropospheric air temperature (ta, K) from AIRS over the whole globe at the eight tropospheric pressure levels of 1000, 925, 850, 700, 600, 500, 400, and 300 hPa is shown in the first column of Figure 1. As expected, the AIRS tropospheric air temperature data show the well-known vertical structure and spatial patterns of tropospheric air temperature. It decreases with both altitude (warm near the surface and cold in the upper troposphere) and latitude (warm near the equator and cold near the poles). Local maxima are found over the Indo-Pacific warm pool and Amazonia at each pressure levels. The air temperature is about 300 K at the equator and 1000 hPa and decreases to around 280 K in the polar regions at 1000 hPa and to around 240 K at the equator and 300 hPa.

The ERA5 and MERRA-2 tropospheric air temperature data are broadly consistent with AIRS in their vertical structures and spatial patterns of tropospheric air temperature (not shown), with high spatial pattern (Pearson) correlations between ERA5 (or MERRA-2) and AIRS ($r > 0.985$) and small Root Mean Square Differences (RMSDs) (<20% of the AIRS standard deviation) at all pressure levels, especially in the free troposphere above 700 hPa (Figures 2c and 2d). The differences between ERA5 and AIRS tropospheric air temperatures (also referred to as ERA5 biases) are generally smaller than 1 K except for a few limited regions in the boundary layer (second column of Figure 1), such as the Southern Ocean, the eastern Pacific, and the eastern Atlantic at 925 and

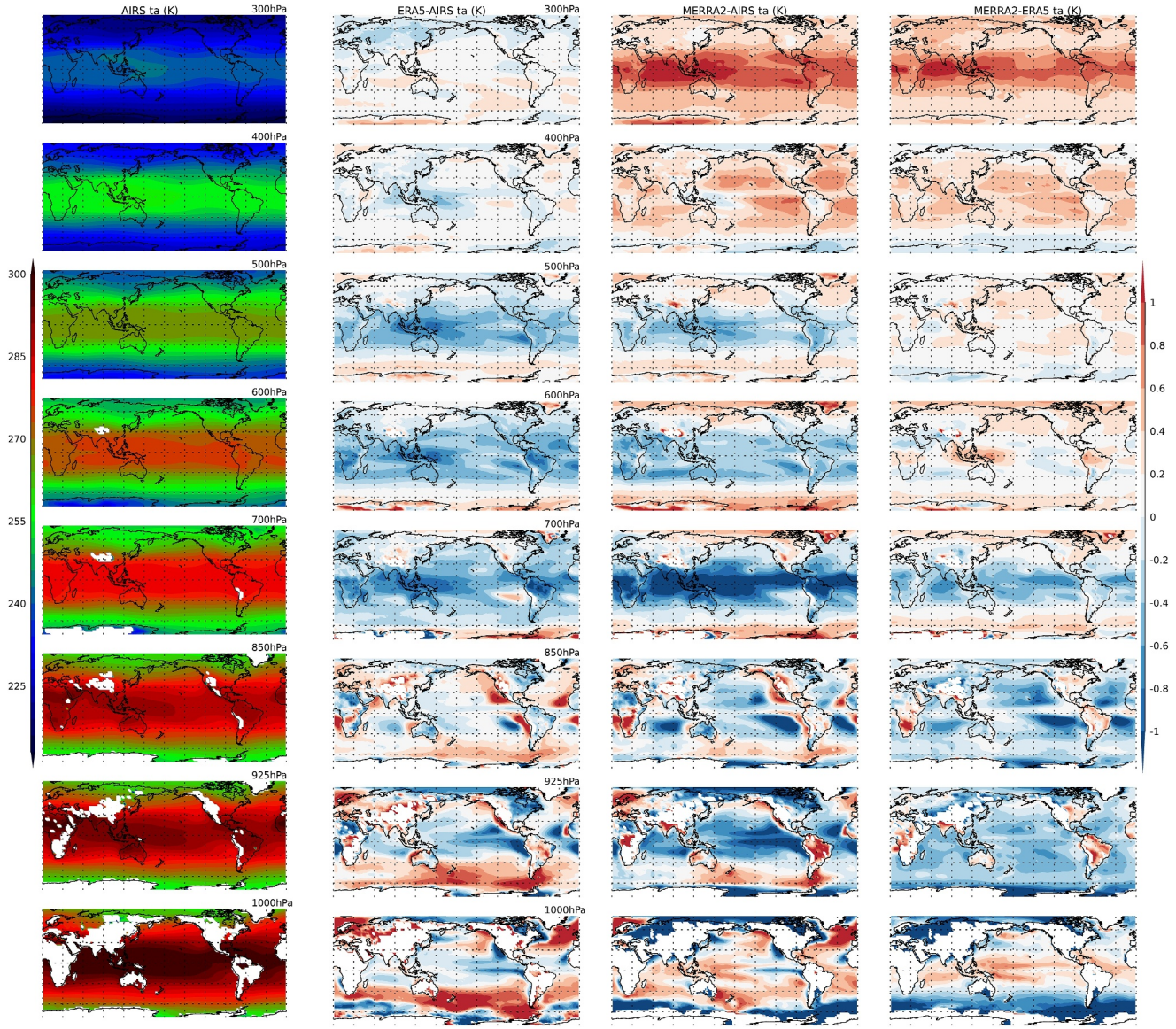

Figure 1. Long-term annual mean tropospheric air temperature (ta, K) from AIRS (first column) and differences among ERA5, MERRA-2 and AIRS tropospheric air temperature (ERA5–AIRS, second column; MERRA-2–AIRS, third column; MERRA-2–ERA5, fourth column) over the globe at the eight tropospheric pressure levels from 1000 to 300 hPa.

1000 hPa. The global mean ERA5 tropospheric air temperature biases are smaller than 0.3 K at all pressure levels with cold biases in the free troposphere from 500 to 700 hPa and warm biases in the upper free troposphere from 300 to 400 hPa and in the boundary layer (Figure 2a). The differences between MERRA-2 and AIRS tropospheric air temperatures (also referred to as MERRA-2 biases) are larger than the ERA5 biases but still smaller than 1 K over most regions and levels, except for the tropics at 300 hPa and 700 hPa, and the Southern Ocean, the Artic Ocean, and the eastern parts of the tropical oceans in the boundary layer (third column of Figure 1). The global mean MERRA-2 tropospheric air temperature biases are greater in magnitude than the global mean ERA5 tropospheric air temperature biases but still mostly smaller than 0.5 K, except at the 1000 hPa, with warm biases above the 600-hPa level (300−500 hPa) and cold biases below the 600-hPa level (700−1000 hPa) (Figure 2a). The patterns of the differences between MERRA-2 and ERA5 tropospheric air temperatures are similar to those of the differences between MERRA-2 and AIRS tropospheric air temperatures. The high consistency between the ERA5 and AIRS tropospheric air temperature data and the good quality of the ERA5 data (Hersbach et al., 2020)

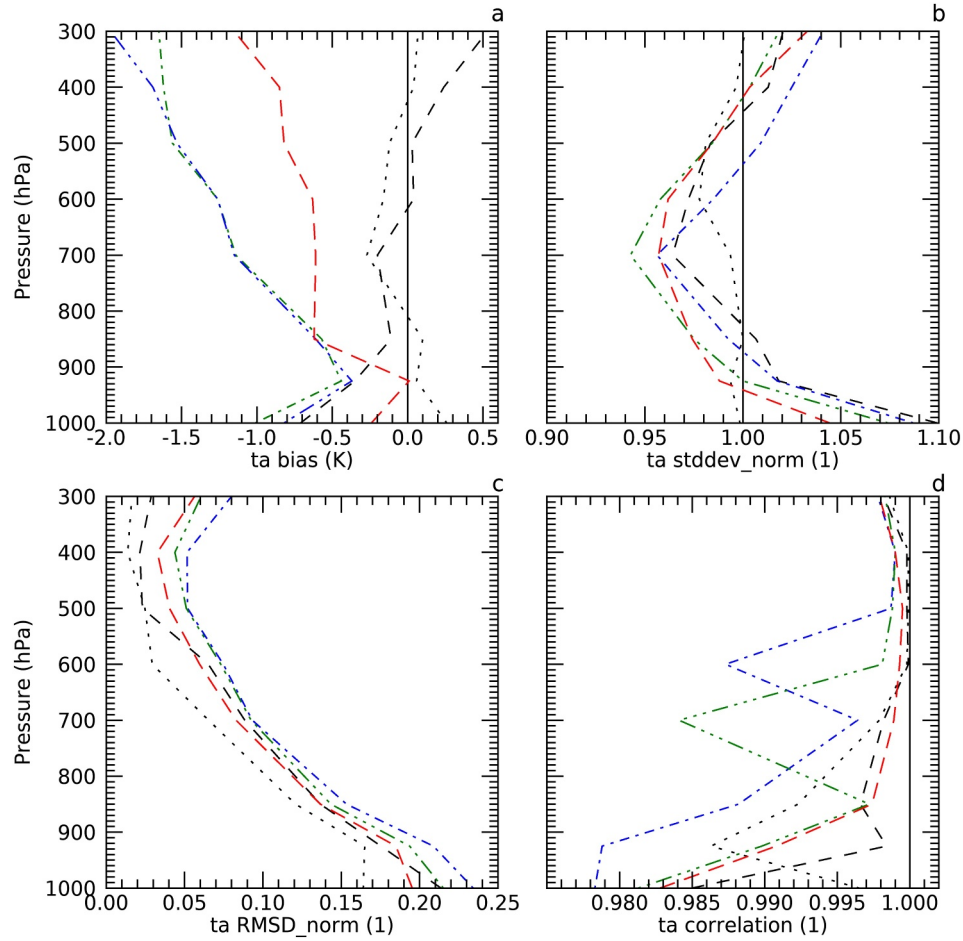

Figure 2. Long-term annual mean tropospheric air temperature (ta) global statistics including (a) bias (K), (b) standard deviation (stddev, 1), (c) root mean square difference (RMSD, 1), and (d) spatial correlation against AIRS at the eight tropospheric pressure levels (1000–300 hPa) from ERA5 (black dotted lines), MERRA-2 (black dashed lines), and the multi-model ensemble means (MMEMs) of CMIP3 (blue dash-dotted lines), CMIP5 (green dash-dot-dot-dotted lines), and CMIP6 (red long-dash lines) models. The stddev and RMSD at each pressure level are normalized by the AIRS stddev at that pressure level.

indicate that the AIRS Obs4MIPs V2.1 tropospheric air temperature data are of sufficient quality and can be used as a reference data set for climate model evaluation.

The horizontal spatial patterns and vertical structures of the long-term annual mean tropospheric air temperature from the MMEMs of the CMIP3, CMIP5, and CMIP6 models are also similar to those from AIRS (not shown) with high spatial pattern correlations ($r > 0.980$) and small RMSDs (<25% of the AIRS standard deviation) between the CMIP models and AIRS at all pressure levels (Figures 2c and 2d). Among the three CMIP phases, the CMIP6 models have the highest spatial pattern correlations and the smallest RMSDs relative to AIRS. This indicates a possible improvement from the CMIP3 and CMIP5 models to the CMIP6 models in the tropospheric air temperature simulations at all pressure levels. However, the CMIP3 and CMIP5 models have similar spatial pattern correlations and RMSDs relative to AIRS indicating the improvement from CMIP3 to CMIP5 is minor.

The global maps of the long-term annual mean tropospheric air temperature biases from the MMEMs of the CMIP3, CMIP5, and CMIP6 models relative to AIRS (CMIP–AIRS) are shown in Figure 3. Please note that only the CMIP tropospheric air temperature biases with the same signs relative to both AIRS and ERA5 are shown in Figure 3. This means that the tropospheric air temperature biases from the CMIP3, CMIP5, and CMIP6 models shown in Figure 3 are robust whether AIRS or ERA5 is used as a reference. The magnitudes of these CMIP

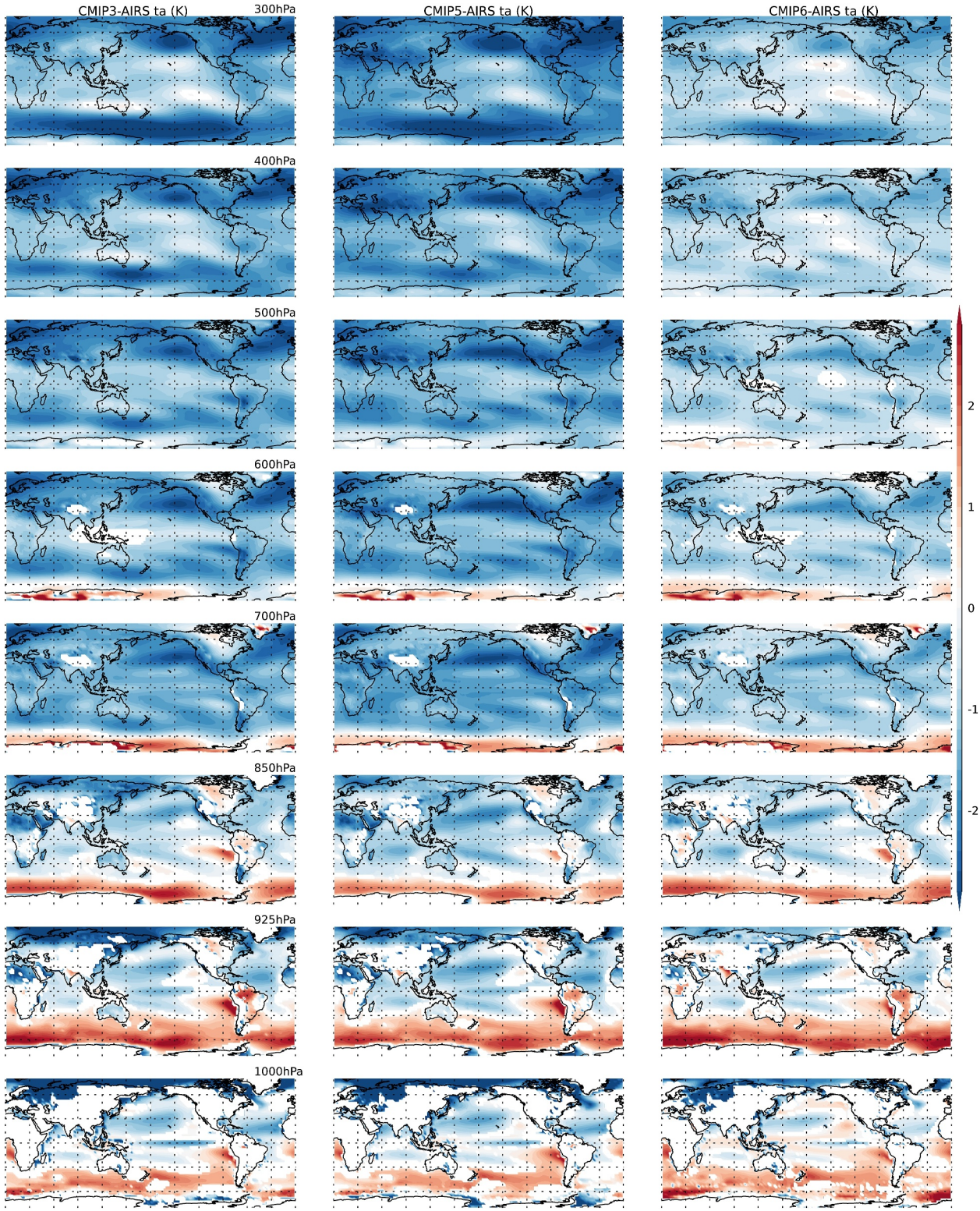

Figure 3. Long-term annual mean tropospheric air temperature (ta) biases (CMIP–AIRS) (K) over the globe at the eight tropospheric pressure levels from the MMEMs of CMIP3 (first column), CMIP5 (second column), and CMIP6 (third column) models. Only the CMIP tropospheric air temperature biases with the same signs relative to both AIRS and ERA5 are shown.

tropospheric air temperature biases can be different relative to either AIRS or ERA5, but their differences, the same as the ERA5 tropospheric air temperature biases relative to AIRS, are mostly small as discussed above and shown in the second column of Figure 1. The tropospheric air temperature bias spatial patterns and vertical structures are very similar among the CMIP3, CMIP5, and CMIP6 models (r = 0.940 between CMIP3 and CMIP5, r = 0.897 between CMIP3 and CMIP6, and r = 0.958 between CMIP5 and CMIP6). This indicates that all three phases of CMIP models share similar tropospheric air temperature biases as discussed below. The most prominent feature of the CMIP tropospheric air temperature biases is near-global cold biases in the free troposphere (700−300 hPa). Cold biases are particularly large (∼4 K) over the middle and high latitudes (south of 30°S and north of 30°N), especially over the mid-latitude storm tracks in the oceans. These mid-latitude free tropospheric cold biases in the CMIP models may be due to the biases in the extratropical storm tracks in the CMIP models (e.g., Priestley et al., 2020). They are much smaller (∼1 K) over the tropics (30°S−30°N) and have minima (close to zero) over the subtropical Pacific, especially in the upper troposphere (400−300 hPa). Warm biases of ∼1 K are seen over a narrow zonal belt near the Antarctic coast at 600 and 700 hPa. In the boundary layer (1000−850 hPa), cold biases are still prevalent over most regions, but their magnitudes are smaller than those in the free troposphere. However, warm biases at 600 and 700 hPa become larger and their area grows wider and moves northward as the altitude decreases from 700 to 925 hPa. As a result, warm biases of ∼2 K are found over the Southern Ocean near the Antarctic coast and the southeastern Pacific along the coast of Peru as well as Amazonia and subtropical South America (especially at 925 hPa). Our finding of the free tropospheric cold biases over almost the whole globe in the CMIP3, CMIP5, and CMIP6 models is consistent with previous studies examining the tropospheric air temperature biases in the AMIP-I models (Gates et al., 1999), the CMIP3 models (John & Soden, 2007), the CMIP5 models (Tian et al., 2013a), and the CMIP6 models (Bock et al., 2020). The boundary-layer warm biases over the Southern Ocean near Antarctica are also seen previously in the CMIP3 models (John & Soden, 2007) and in the CMIP5 models (Tian et al., 2013a). Given its consistency across studies, this is likely a real bias for the CMIP models although the AIRS air temperature is subjected to large uncertainties in the Southern Ocean due to cloud and ice influence.

For the three CMIP phases, the global mean tropospheric air temperature biases are all negative at all pressure levels (Figure 2a) due to the prominent near-global cold biases in the free troposphere discussed above (Figure 3). They maximize (∼2 K) at 300 hPa and decrease with altitude, with a minimum (<0.5 K) at 925 hPa in the boundary layer (Figure 2a). The decrease of the global mean tropospheric cold biases with altitude is due to the canceling effect of the increasing spatial area of the tropospheric warm biases over the Southern Ocean as the altitude decreases (Figure 3). In addition, the global mean tropospheric cold biases in the CMIP3 and CMIP5 models are similar, while the global mean tropospheric cold biases in the CMIP6 models are much smaller than the those in the CMIP3 and CMIP5 models (Figure 2a). This again indicate a simulation improvement or a bias reduction from the CMIP5 models to the CMIP6 models, but less so from the CMIP3 models to the CMIP5 models.

3.2. Specific Humidity (hus)

The long-term annual mean tropospheric specific humidity (hus, g kg^{-1}) from AIRS over the whole globe at the eight tropospheric pressure levels is shown in the first column of Figure 4. Also as expected, the AIRS tropospheric specific humidity data show the well-known horizontal and vertical patterns of tropospheric specific humidity. It decreases with both altitude (moist near the surface and dry in the upper troposphere) and latitude (moist near the equator and dry near the poles). The AIRS tropospheric specific humidity data also shows the well-known tropical moist deep convective features and the equatorial and subtropical dry descending regions as shown by the precipitation (Tian & Dong, 2020). For example, the ITCZ, the South Pacific convergence zone (SPCZ), the South Atlantic convergence zone (SACZ), and the equatorial Pacific cold tongue are evident in the first column of Figure 4.

The ERA5 and MERRA-2 tropospheric specific humidity data are also broadly consistent with the AIRS tropospheric specific humidity data in the vertical structures and spatial patterns, with a high spatial pattern (Pearson) correlation among AIRS, ERA5 and MERRA-2 (r = 0.999 between AIRS and ERA5, r = 0.997 between AIRS and MERRA-2, and r = 0.995 between ERA5 and MERRA-2; not shown). In particular, the AIRS and ERA5 tropospheric specific humidity data are highly consistent with each other, with very high spatial pattern correlations (>0.997) between them at all pressure levels and very small RMSDs (<10% of the AIRS standard deviation) at all pressure levels except for 300 hPa (Figures 5d and 5e). At 300 hPa, the relatively large RMSD

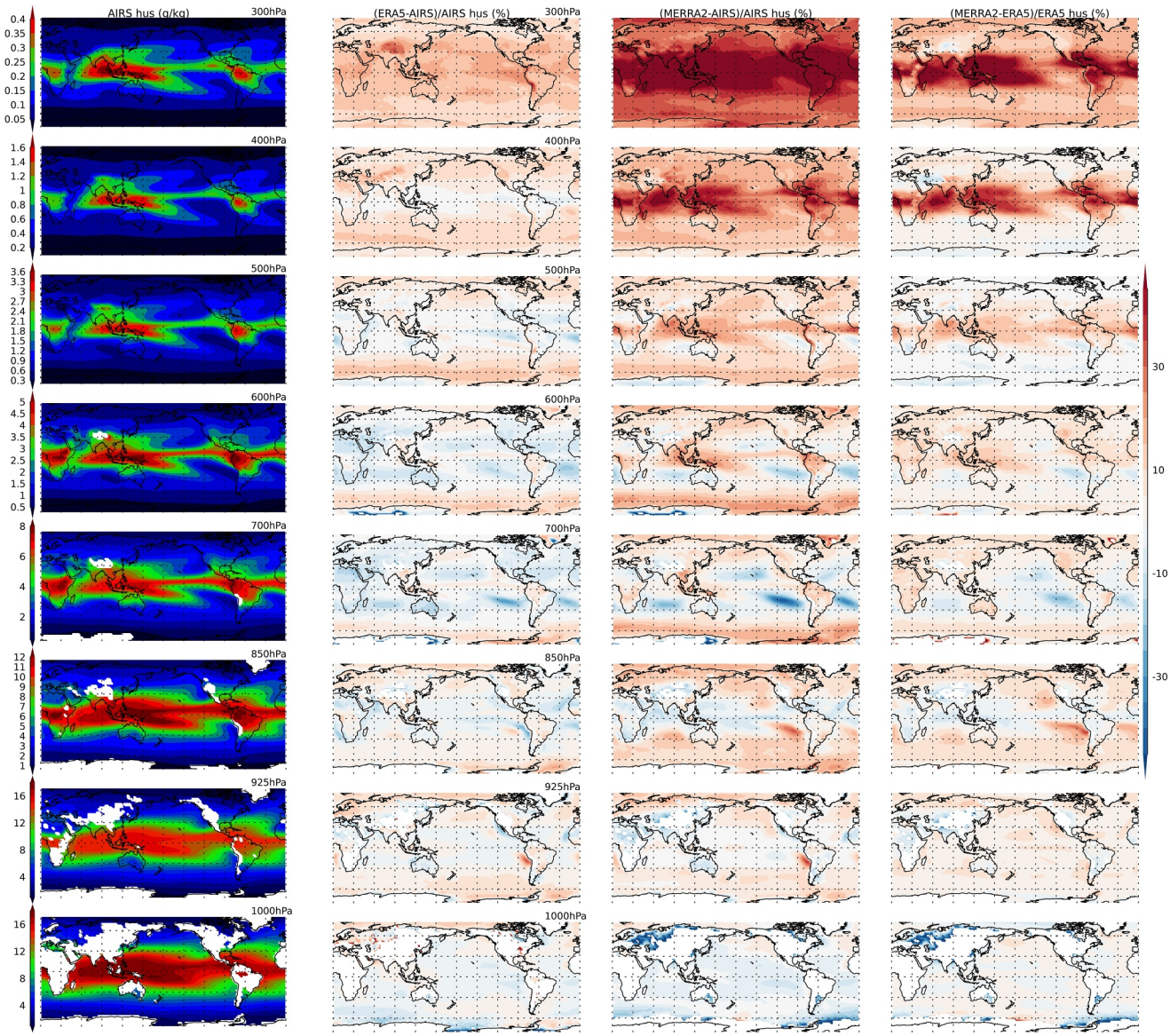

Figure 4. Same as Figure 1 but for tropospheric specific humidity (hus, g kg^{-1}) and the differences among ERA5, MERRA-2 and AIRS normalized by AIRS or ERA5 mean values ((ERA5–AIRS)/AIRS, second column; (MERRA-2–AIRS)/AIRS, third column; (MERRA-2–ERA5)/ERA5, fourth column).

between ERA5 and AIRS tropospheric specific humidity is only ~20% of the AIRS standard deviation. At this level, the ERA5 specific humidity is ~20% larger than the AIRS specific humidity almost globally (Figure 4) and the maximum absolute differences are located over the tropical moist deep convective regions (Figure S1 in Supporting Information S1). As discussed later, the MERRA-2 specific humidity is also mostly much larger than the AIRS specific humidity at 300 hPa. This indicates that the AIRS specific humidity may have dry biases at 300 hPa. In the middle and lower troposphere from 500 hPa and down, the ERA5 tropospheric specific humidity relative biases are smaller than 10% almost globally (Figure 4). The global mean ERA5 tropospheric specific humidity absolute biases are smaller than 0.1 g kg^{-1} at all pressure levels with positive specific humidity biases from 300 to 600 hPa and at 925 hPa and negative specific humidity biases at 1000 hPa (Figure 5a). The global mean ERA5 tropospheric specific humidity relative biases are smaller than 5% in the middle and lower troposphere from 500 hPa and down (Figure 5a).

The differences between MERRA-2 and AIRS tropospheric specific humidity data are much bigger than the differences between ERA5 and AIRS tropospheric specific humidity data (Figure 4 and Figure S1 in Supporting

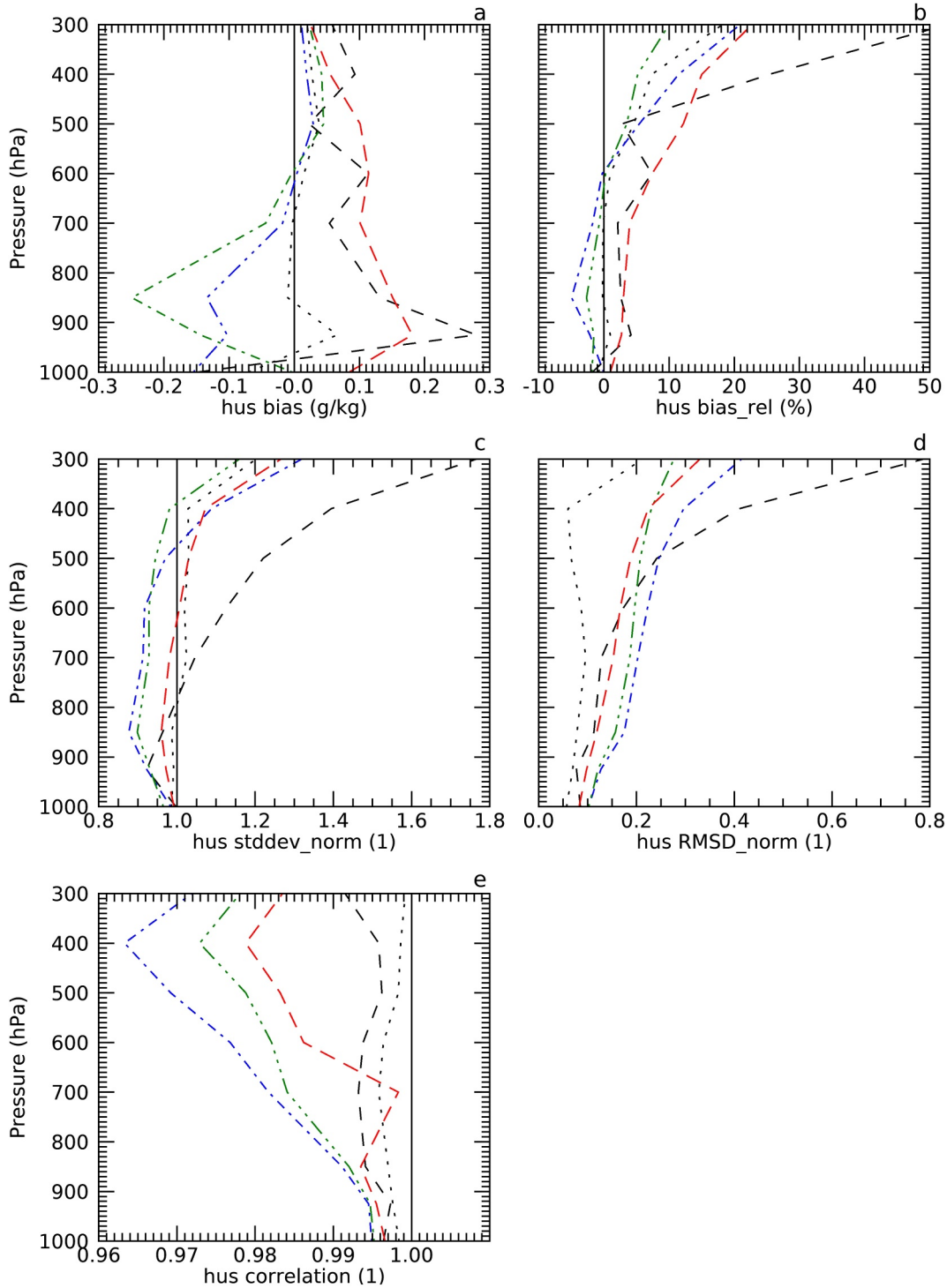

Figure 5. Long-term annual mean tropospheric specific humidity (hus) global statistics including (a) absolute bias (g kg^{-1}), (b) relative bias (%), (c) standard deviation (stddev, 1), (d) root mean square difference (RMSD, 1), and (e) spatial correlation against AIRS at the eight tropospheric pressure levels (1000–300 hPa) from ERA5 (black dotted lines), MERRA-2 (black dashed lines), and the multi-model ensemble means (MMEMs) of CMIP3 (blue dash-dotted lines), CMIP5 (green dash-dot-dot-dotted lines), and CMIP6 (red long-dash lines) models. The stddev and RMSD at each pressure level are normalized by the AIRS stddev at that pressure level.

Information S1). For example, at 300 hPa, MERRA-2 is much moister (~30%) than AIRS almost globally, particularly over the tropical moist deep convective regions (over 50%; Figure 4 and Figure S1 in Supporting Information S1). The RMSDs between the MERRA-2 and AIRS tropospheric specific humidity are around 80% of the AIRS standard deviation at 300 hPa (Figure 5d). As mentioned earlier, both ERA5 and MERRA-2 are moister than AIRS at 300 hPa, indicating that the AIRS specific humidity may have dry biases at 300 hPa. However, MERRA-2 is much moister than both AIRS and ERA5 at this level. The large difference between the MERRA-2 and AIRS (or ERA5) tropospheric specific humidity is consistent with previous studies showing MERRA-2 tropospheric specific humidity is biased (too moist) in the upper troposphere (Gelaro et al., 2017) and indicates a problem of the MERRA-2 tropospheric specific humidity data instead of the AIRS or ERA5 tropospheric specific humidity data. This MERRA-2 moist bias over the tropical moist deep convective regions is most obvious at 300 hPa and 400 hPa but can also be found in all pressure levels in the free troposphere, though their magnitudes and area decrease as the altitude decreases. In addition, MERRA-2 is drier than AIRS over the subtropical dry descending regions in the middle and lower free troposphere (500–700 hPa) and the difference magnitudes increase as the altitude decreases. At 850 hPa, MERRA-2 is drier than AIRS over tropical moist deep convective regions but moister than AIRS over subtropical dry descending regions. The global mean MERRA-2 specific humidity biases are positive in the troposphere except at 1000 hPa, and much bigger than the ERA5 specific humidity biases (Figure 5a). The spatial pattern (Pearson) correlations between the MERRA-2 and AIRS tropospheric specific humidity are around 0.993 and consistently lower than those between ERA5 and AIRS tropospheric specific humidity, and the RMSDs between MERRA-2 and AIRS tropospheric specific humidity are greater than 10% of the AIRS standard deviation at almost all pressure levels and consistently larger than RMSDs between the ERA5 and AIRS tropospheric specific humidity (Figures 5d and 5e).

It is well known that ERA5 has smaller absolute biases than MERRA-2 partly due to the higher resolution and the newer data assimilation system in ERA5 (Hersbach et al., 2020). The greater consistency between the ERA5 and AIRS tropospheric specific humidity data and the good quality of the ERA5 data (Hersbach et al., 2020) indicate that the AIRS Obs4MIPs V2.1 tropospheric specific humidity data are of sufficient quality to be used as a reference data set for climate model evaluation. Please note that both ERA5 and MERRA-2 have assimilated AIRS L1 radiances. Thus, whether or not AIRS data are assimilated cannot explain the differences between ERA5 and MERRA-2 tropospheric specific humidity.

The horizontal and vertical structures of the long-term annual mean tropospheric specific humidity from the MMEMs of the CMIP3, CMIP5, and CMIP6 models are similar to AIRS (not shown) with high spatial correlations between the CMIP models and AIRS ($r > 0.96$) and small RMSDs (<40% of the AIRS standard deviation) at all pressure levels (Figures 5d and 5e). Among the three CMIP phases, the CMIP6 models have the highest spatial pattern correlations and the smallest RMSDs relative to AIRS while the CMIP3 models have the lowest spatial pattern correlations and the largest RMSDs relative to AIRS at all tropospheric pressure levels except for 300 hPa (Figures 5d and 5e). This indicates a possible improvement from CMIP3 to CMIP5 and from CMIP5 to CMIP6 in tropospheric specific humidity simulation at all tropospheric pressure levels except 300 hPa.

Global maps of the long-term annual mean tropospheric specific humidity relative biases from the MMEMs of the CMIP3, CMIP5, and CMIP6 models (CMIP–AIRS) are shown in Figure 6 with the absolute biases shown in Figure S2 in Supporting Information S1. As with the CMIP model tropospheric air temperature biases shown in Figure 3, only the robust CMIP tropospheric specific humidity biases relative to both AIRS and ERA5 are shown in Figure 6 and Figure S2 in Supporting Information S1. Although the exact magnitudes of these CMIP tropospheric specific humidity biases depends on whether AIRS or ERA5 data are used as a reference, the differences between them are small, as discussed above and shown in the second column of Figure 4 or the first column of Figure S1 in Supporting Information S1. Strong similarities are seen in the tropospheric specific humidity bias horizontal patterns and vertical structures among the CMIP3, CMIP5, and CMIP6 models ($r = 0.941$ between CMIP3 and CMIP5, $r = 0.785$ between CMIP3 and CMIP6, and $r = 0.873$ between CMIP5 and CMIP6). This indicates that all three phases of CMIP models share similar tropospheric specific humidity biases that have different signs and magnitudes depending on different regions and altitudes considered. We first discuss the tropospheric specific humidity biases over the middle and high latitudes (poleward of 30°S and 30°N) in Figure 6. Negative specific humidity biases predominate in the free troposphere above the 600-hPa level (300−500 hPa), while mostly positive specific humidity biases are found in the lower troposphere below the 600-hPa level (700−1000 hPa). The magnitudes of the negative specific humidity biases in the free troposphere above the 600-

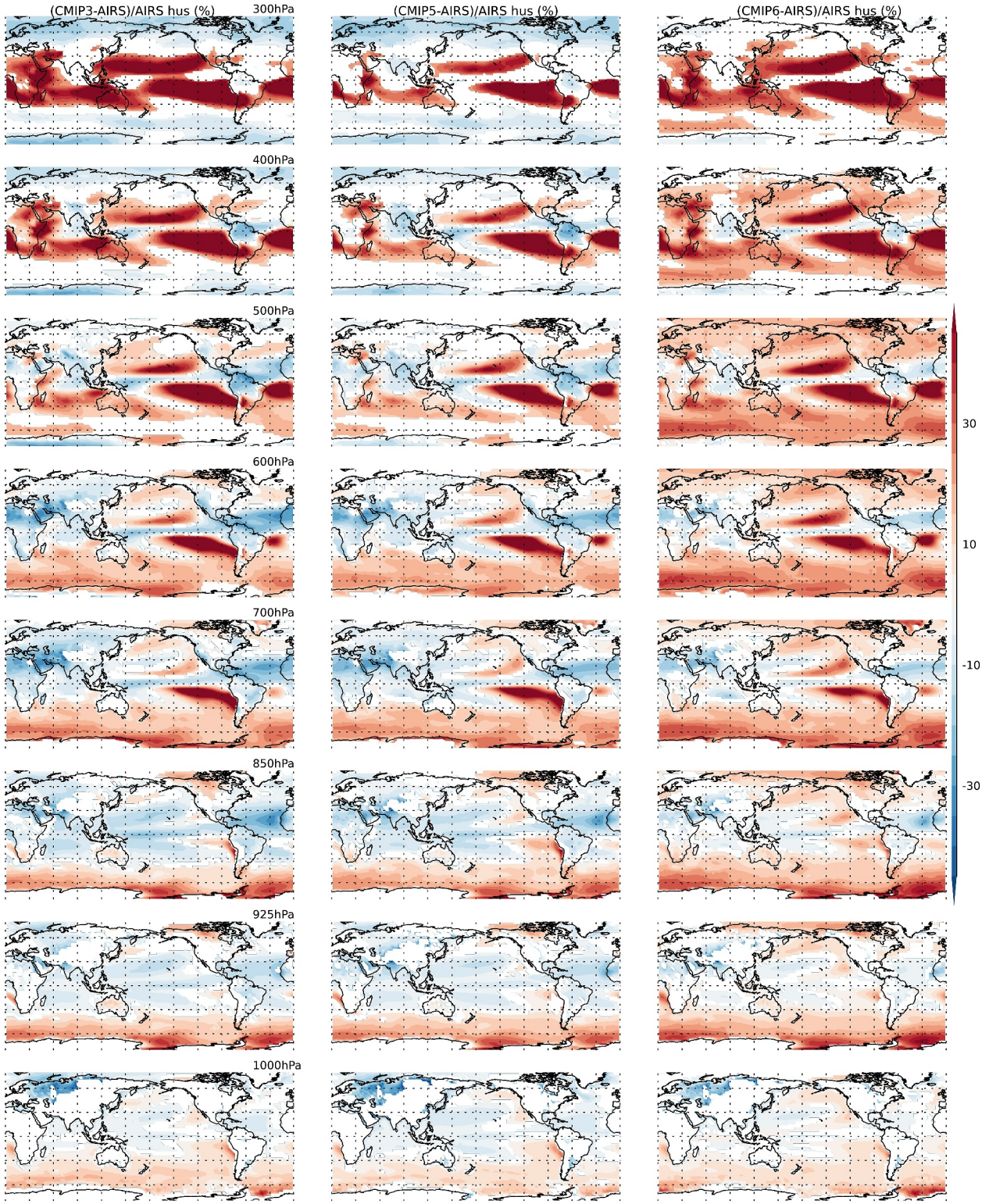

Figure 6. Same as Figure 3 but for the CMIP model tropospheric specific humidity (hus) relative biases ((CMIP–AIRS)/AIRS, %).

hPa level (300−500 hPa) decrease with the decreasing altitude from 300 hPa to 500 hPa and become almost zero at 600 hPa. In contrast, the magnitudes of the positive specific humidity biases in the lower troposphere the 600-hPa level (700−1000 hPa) increase as the altitude decreases from 700 hPa to 1000 hPa (Figure 6).

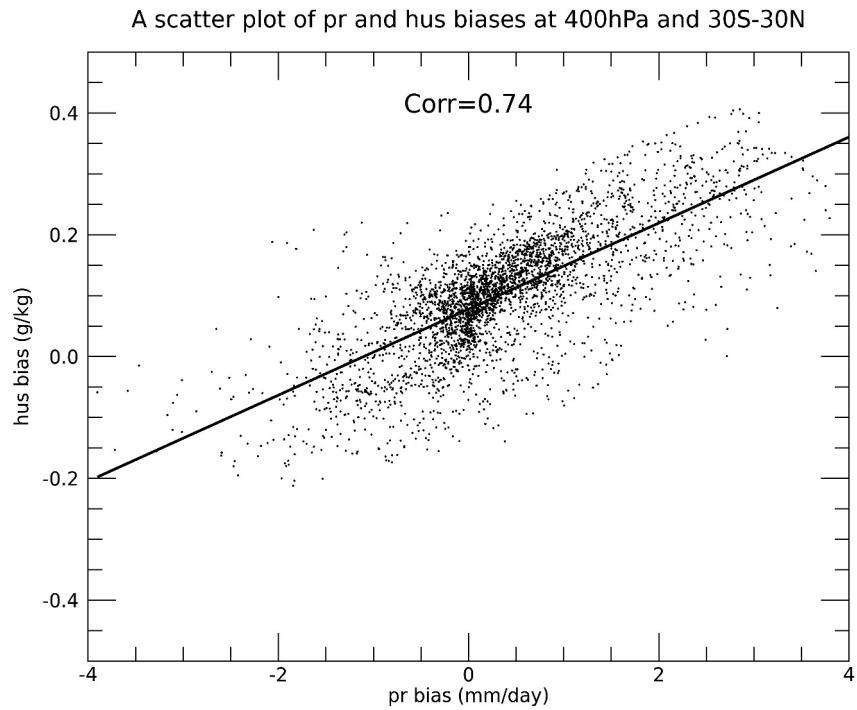

Figure 7. A scatter plot of precipitation biases and 400-hPa specific humidity biases in the tropics (30°S–30°N) from CMIP6 models. The tropical precipitation biases are adopted from Tian and Dong (2020).

We next discuss the CMIP tropospheric specific humidity biases over the tropics (30°S−30°N). At 300 hPa, positive specific humidity biases are found over the tropical North Pacific Ocean (5°−30°N) and over the tropical South Indian, Southeastern Pacific and South Atlantic Oceans (equator−30°S). Near zero or small negative CMIP specific humidity biases are seen over the equatorial North Atlantic (equator−10°N), central America, Amazonia, south Asia including the Bay of Bengal, the equatorial Indian Ocean, the Maritime Continent, the equatorial Pacific, and middle and high latitudes. However, these near zero or small negative CMIP specific humidity biases at 300 hPa depend on the reference data used: near zero biases based on AIRS and small negative biases based on ERA5. As we discussed earlier, the AIRS specific humidity data may have their own dry biases at 300 hPa, so the small negative CMIP specific humidity biases based on ERA5 may be more trustworthy at this level. As the altitude decreases, the area of the CMIP positive specific humidity biases becomes smaller and narrower, and moves further eastward over the tropical oceans, especially over the tropical Pacific. On the other hand, the area of the negative specific humidity biases gets larger and wider, and the magnitude of the negative specific humidity biases gets larger with the decreasing altitude. As a result, in the boundary layer (1000−850 hPa), negative specific humidity biases in CMIP can be seen almost everywhere except for small parts of the eastern Pacific and eastern Atlantic Oceans with positive specific humidity biases. For the three CMIP phases, the global mean tropospheric specific humidity biases are all positive above the 600 hPa level (300−500 hPa) but they have different signs for different CMIP ensembles below the 600-hPa level (700−1000 hPa): Negative for the CMIP3 and CMIP5 models but positive for the CMIP6 models (Figure 5a). From this information alone, it is hard to see any significant simulation improvement or bias reduction from CMIP3 to CMIP5 to CMIP6. Our finding of the tropospheric specific humidity biases in the CMIP3, CMIP5, and CMIP6 models against AIRS is consistent with previous studies who have examined the tropospheric specific humidity biases in the CMIP3 models (John & Soden, 2007; Pierce et al., 2006; Tian, 2015) and the CMIP5 models (Tian, 2015; Tian et al., 2013a).

The tropical free tropospheric (300−700 hPa) specific humidity bias patterns (Figure 6) in the CMIP3, CMIP5, and CMIP6 models against AIRS (and ERA5) are similar to the tropical precipitation bias pattern in the models as demonstrated by a high correlation between the tropical precipitation biases and the tropical 400-hPa specific humidity biases for CMIP6 shown in Figure 7. Similar high correlations are also found at other free tropospheric levels (300−700 hPa) and in CMIP3 and CMIP5 models but not shown. Thus, the tropical free tropospheric (300−700 hPa) specific humidity biases are related to the double-ITCZ bias in the models (Tian & Dong, 2020) as

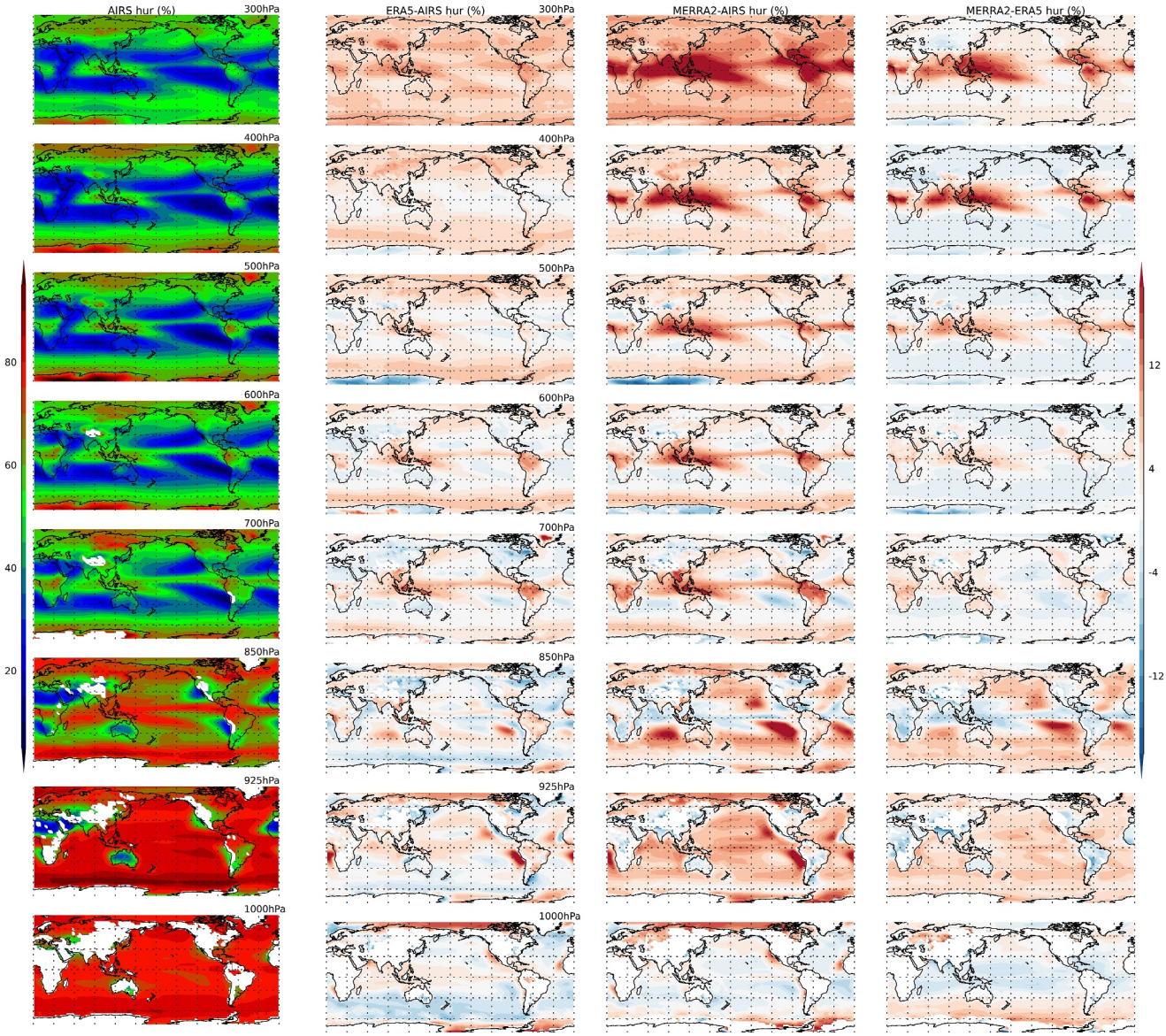

Figure 8. Same as Figure 1 but for relative humidity (hur, %).

has been previously shown by Tian et al. (2013a) and Tian (2015). There is a clear advantage to use the three-dimensional AIRS tropospheric specific humidity data instead of the two-dimensional TRMM/GPCP precipitation data to identify and understand the double-ITCZ bias in the models: the AIRS tropospheric specific humidity bias can reveal vertical structure, such as the westward shift with the increasing altitude over the tropical oceans, that cannot be seen in the TRMM/GPCP precipitation data.

3.3. Relative Humidity (hur)

The long-term annual mean tropospheric relative humidity (hur, %) from AIRS over the whole globe at the eight tropospheric pressure levels is shown in the first column of Figure 8. Again, as expected, the AIRS tropospheric relative humidity data show the well-known horizontal and vertical patterns of tropospheric relative humidity. In the boundary layer (i.e., 925 and 1000 hPa), higher relative humidity (>70%) is found almost everywhere over the oceans and over most land regions. In contrast, lower relative humidity (<50%) is seen over several well-known large land deserts, such as the Sahara and Kalahari deserts in Africa, the Arabian, Syrian, Karakum, Dasht-e Kavir, Taklamakan, and Gobi deserts in the Middle East and Asia, the Great Australian desert in Australia,

the Great Basin desert in North America, and the Patagonian desert in South America, as well as the coastal oceans near these deserts. At the top of the boundary layer (850 hPa), relatively high relative humidity (>70%) is found in the middle and high latitudes (poleward of 40°N and 40°S) and in the tropical moist deep convective regions, such as the equatorial Africa, the Indo-Pacific warm pool, Amazonia, the ITCZ, the SPCZ, and the SACZ. Lower relative humidity (<50%) is also seen over the well-known dry descending subtropical regions over the eastern parts of the oceans and the western parts of the continents, such as the subtropical southeastern and northeastern Pacific and Atlantic, the Sahara desert and the Middle East, southern Africa, and the southern Indian Ocean (Tian & Dong, 2020). In the free troposphere, as the altitude increases from 700 to 300 hPa, the dry subtropical regions become drier and wider and move further westward over the oceans, while the tropical moist deep convective regions become smaller. As a result, in the upper troposphere such as 300 hPa, higher relative humidity (>70%) is seen over the middle and high latitudes and the tropical moist deep convective regions, while lower relative humidity (<50%) is seen over the dry subtropical regions.

The ERA5 and MERRA-2 tropospheric relative humidity data show overall consistency with AIRS in their horizontal and vertical structures and spatial patterns of tropospheric relative humidity with a high spatial pattern (Pearson) correlation among them (r = 0.9779 between AIRS and ERA5, r = 0.9644 between AIRS and MERRA-2, and r = 0.9827 between ERA5 and MERRA-2; not shown). However, these correlations are smaller than those for the tropospheric air temperature and specific humidity. This is expected because both the tropospheric air temperature and specific humidity biases contribute to the tropospheric relative humidity bias.

The absolute differences between the ERA5 and AIRS tropospheric relative humidity are smaller than 10% almost everywhere (Figure 8). The global mean ERA5 tropospheric relative humidity biases are positive and smaller than 5% at all pressure levels except for 300 hPa where they are ∼6% (Figure 9a). The spatial correlations between the ERA5 and AIRS tropospheric relative humidity are high (>0.94) at all pressure levels except 1000 hPa (Figure 9d). This indicates that the AIRS and ERA5 tropospheric relative humidity data are highly consistent with each other. This is expected because both the tropospheric air temperature and specific humidity data are highly consistent between AIRS and ERA5 as we discussed in Sections 3.1 and 3.2.

The differences between the MERRA-2 and AIRS tropospheric relative humidity data are much bigger than the differences between the ERA5 and AIRS (Figure 8). For example, at 300 hPa, MERRA-2 is much moister than AIRS almost globally, particularly over the tropical moist deep convective regions. This MERRA-2 moist bias in relative humidity at 300 hPa is consistent with and caused by the warm bias and positive specific humidity bias in MERRA-2 at 300 hPa. The MERRA-2 positive relative humidity biases over the tropical moist deep convective regions can be found at all pressure levels in the free troposphere and their magnitudes and area decrease when moving downward. In addition, MERRA-2 is drier than AIRS over the subtropical dry descending regions in the lower free troposphere (500–700 hPa) with an increasing magnitude as the altitude decreases. At 850 hPa, MERRA-2 is drier than AIRS over the tropical moist deep convective regions but moister than AIRS over the subtropical dry descending regions. The global mean ERA5 and MERRA-2 tropospheric relative humidity biases are positive at all pressure levels with a minimum at 850 hPa and a maximum at 300 hPa (Figure 9a). In addition, the global mean MERRA-2 tropospheric relative humidity biases are consistently larger (∼5%) than those from ERA5 (∼3%) at all pressure levels. This indicates the better quality of ERA5 tropospheric relative humidity data over MERRA-2. This is also supported by the consistent higher spatial pattern correlations and smaller RMSDs between ERA5 and AIRS tropospheric relative humidity at all pressure levels than those between MERRA-2 and AIRS tropospheric relative humidity (Figures 9c and 9d). This is expected because we have found bigger biases from MERRA-2 than ERA5 in both the tropospheric air temperature and specific humidity data as we discussed in Sections 3.1 and 3.2. In addition, the MERRA-2 tropospheric relative humidity bias pattern is very similar to the MERRA-2 tropospheric specific humidity bias pattern. This indicates that the MERRA-2 tropospheric relative humidity biases are mainly from the MERRA-2 specific humidity biases instead of from the MERRA-2 air temperature biases.

Again, it is well known that ERA5 has smaller biases than MERRA-2 partly due to the higher resolution and the newer data assimilation system in ERA5 (Hersbach et al., 2020). Similar to the specific humidity discussed in Section 3.2, the relatively high consistency between the ERA5 and AIRS tropospheric relative humidity data and the good quality of the ERA5 data (Hersbach et al., 2020) indicate that the AIRS Obs4MIPs V2.1 tropospheric relative humidity data are of sufficient quality and can be used as a reference data set for climate model evaluation. Also like the specific humidity discussed in Section 3.2, the large difference between the MERRA-2 and

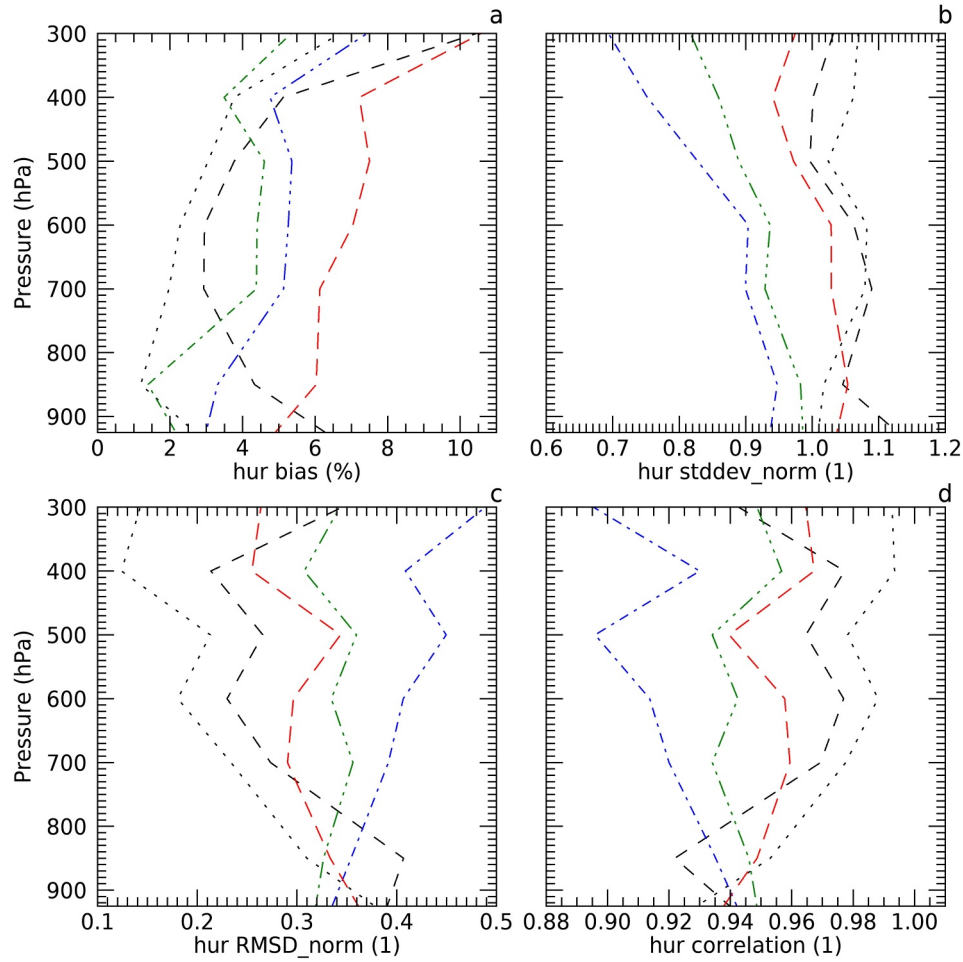

Figure 9. Same as Figure 2 but for relative humidity (hur, %) from 300 to 925 hPa.

AIRS or ERA5 tropospheric relative humidity data indicates a problem of the MERRA-2 tropospheric relative humidity data instead of the AIRS or ERA5 tropospheric relative humidity data.

The horizontal and vertical structures of the long-term annual mean tropospheric relative humidity from the MMEMs of the CMIP3, CMIP5, and CMIP6 models are also similar to AIRS (not shown) with high spatial pattern correlations between the CMIP models and AIRS ($r > 0.90$) at all pressure levels except for 1000 hPa (Figure 9d). In the free troposphere, the CMIP6 models have the highest pattern correlations and the smallest RMSDs while the CMIP3 models have the lowest pattern correlations and the largest RMSDs relative to AIRS among the three CMIP phases (Figures 9c and 9d). This indicates a possible improvement from CMIP3 to CMIP5 and to CMIP6 in the free tropospheric relative humidity simulation. In the boundary layer, the situation is different and complicated and there is no obvious improvement in the CMIP6 models in comparison to the CMIP3 or CMIP5 models.

The global maps of the long-term annual mean tropospheric relative humidity biases from the MMEMs of the CMIP3, CMIP5, and CMIP6 models (CMIP–AIRS) are shown in Figure 10. As the CMIP model tropospheric air temperature and specific humidity biases shown in Figures 3 and 6, only the robust CMIP tropospheric relative humidity biases relative to both AIRS and ERA5 are shown in Figure 10. Although the exact magnitudes of these CMIP tropospheric relative humidity biases depends on whether AIRS or ERA5 is used as a reference, the differences between them are small as discussed above and shown in the second column of Figure 8. Like the tropospheric air temperature and specific humidity biases discussed in Sections 3.1 and 3.2, strong similarities are found in the tropospheric relative humidity bias horizontal and vertical structures among the CMIP3, CMIP5, and CMIP6 models ($r = 0.927$ between CMIP3 and CMIP5, $r = 0.769$ between CMIP3 and CMIP6, and $r = 0.905$

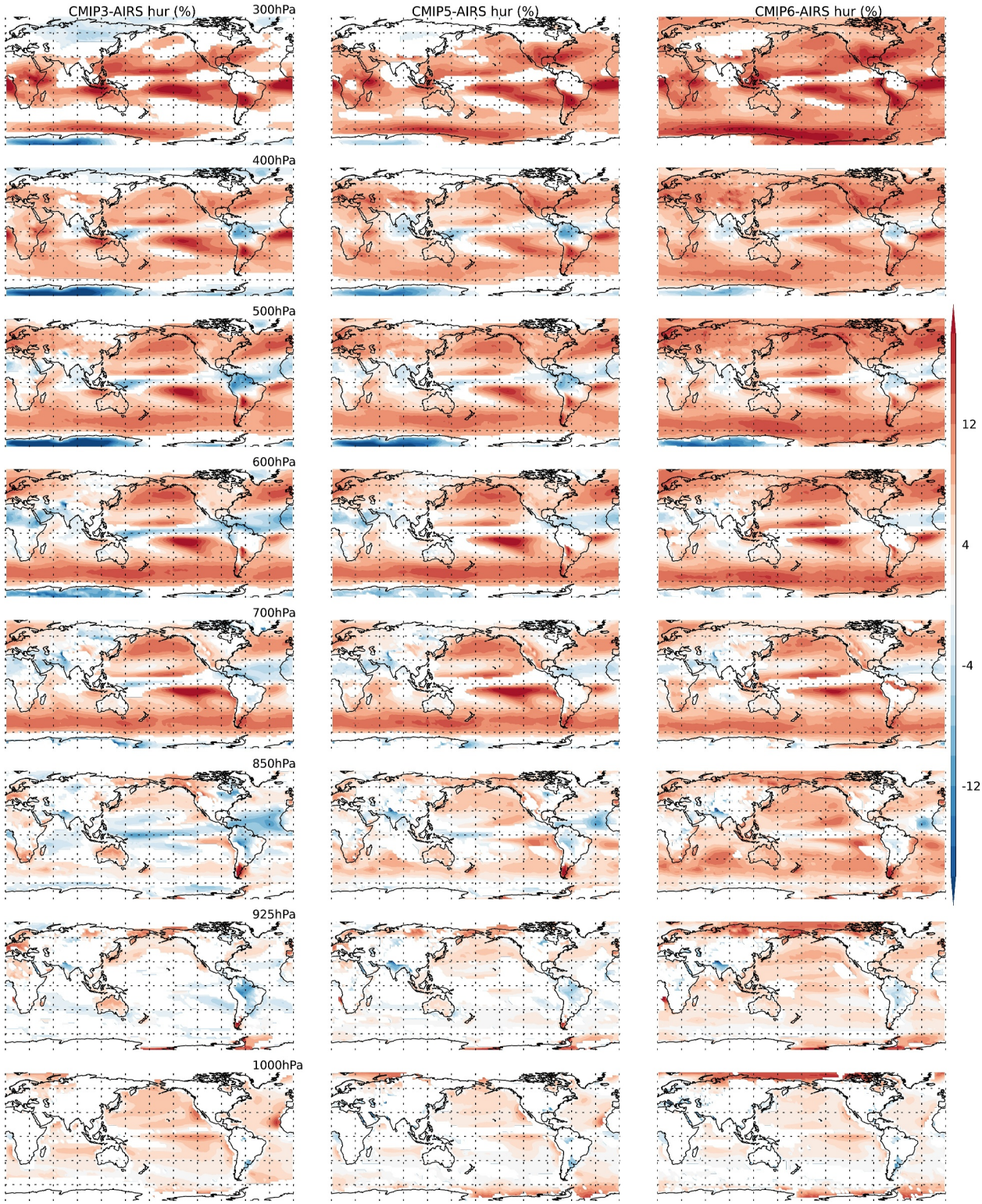

Figure 10. Same as Figure 3 but for the CMIP model tropospheric relative humidity (hur) biases (CMIP–AIRS, %).

between CMIP5 and CMIP6). This indicates that all three phases of CMIP models share similar tropospheric relative humidity biases as described below. This is also expected because all three phases of CMIP models share similar tropospheric air temperature and specific humidity biases as described in Sections 3.1 and 3.2.

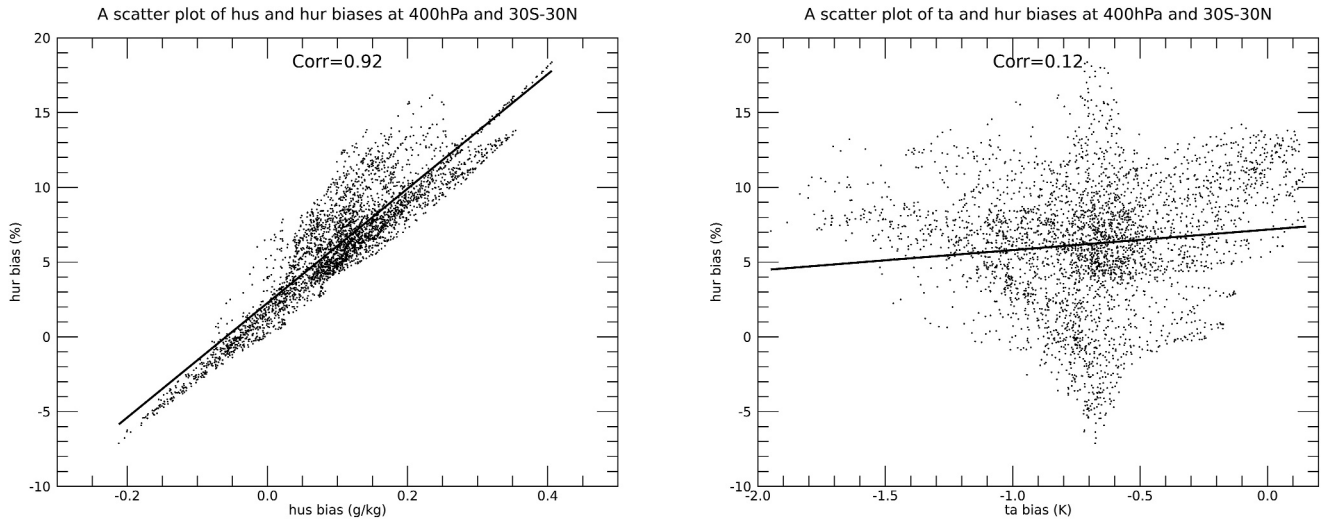

Figure 11. Scatter plots of the specific and relative humidity biases (left column) and the air temperature and relative humidity biases (right column) at 400 hPa in the tropics (30°S–30°N) from CMIP6 models.

In the boundary layer, the relative humidity biases are generally small (<10%) except for a few regions such as the Arctic Ocean and the coast of Antarctica. However, the relative humidity biases are generally large (up to 20%) in the free troposphere. The free tropospheric relative humidity biases have different signs and magnitudes in different regions like the free tropospheric specific humidity biases. Large positive free tropospheric relative humidity biases are found over the middle latitude oceans (poleward of 30°S and 30°N). Negative free tropospheric relative humidity biases are found over the high latitudes (poleward of 70°S and 70°N) especially for CMIP3 and CMIP5 models. In the tropics (30°S−30°N), large positive free tropospheric relative humidity biases are generally found over the North Pacific Ocean (5°−30°N) and over the South Indian, Southeastern Pacific and South Atlantic Oceans (equator−30°S). Near zero or small negative free tropospheric relative humidity biases can be seen over Amazonia, central America, the equatorial North Atlantic (equator−10°N), south Asia including the Bay of Bengal, the equatorial Indian Ocean, the Maritime Continent, the equatorial Pacific, and the middle and high latitudes. As with tropospheric specific humidity at 300 hPa, these near zero or small negative relative humidity biases at 300 hPa depend on the reference data used: near zero biases based on AIRS and small negative biases based on ERA5. As we discussed earlier, the AIRS relative humidity may have dry biases at 300 hPa and, so the small negative relative humidity CMIP biases based on ERA5 may be more trustworthy at this level. This tropical free tropospheric relative humidity bias pattern resembles the tropical free tropospheric specific humidity bias pattern as demonstrated by a high correlation (0.92) between the tropical specific and relative humidity biases at 400 hPa for CMIP6 shown in Figure 11. On the other hand, the resemblance of the tropical free tropospheric relative humidity and air temperature bias patterns is low as demonstrated by a low correlation (0.12) between the tropical air temperature and relative humidity biases at 400 hPa for CMIP6 (Figure 11). Similar high correlations between the tropical specific and relative humidity biases and similar low correlations between the tropical air temperature and relative humidity biases are also found at other free tropospheric levels (300−700 hPa) and in CMIP3 or CMIP5 models. Thus, the tropical free tropospheric relative humidity biases should arise mainly from the tropical free tropospheric specific humidity biases instead of the tropical free tropospheric air temperature biases and be related to the double-ITCZ bias in the models (Figure 6).

For all three CMIP phases, the global mean tropospheric relative humidity biases are positive at all pressure levels. The CMIP6 models have the largest global mean tropospheric relative humidity biases while the CMIP5 models have the smallest global mean tropospheric relative humidity biases except at 1000 hPa. Thus, from this perspective, it is hard to discern any possible improvement or bias reduction from the CMIP3 and CMIP5 models to the CMIP6 models. Our current finding of the tropospheric relative humidity biases in the CMIP3, CMIP5, and CMIP6 models is consistent with previous studies examining the tropospheric relative humidity biases in the CAM3 models (Gettelman et al., 2006) and the CMIP5 models (Su et al., 2014).

Given the tropospheric air temperature and specific humidity biases, the tropospheric relative humidity bias can be estimated using the following equation:

$$\Delta \mathrm{RH} = [\mathrm{RH}\Delta(\ln q)] + \left[RH\left(\frac{c_p}{R_d} - \frac{L}{R_v T}\right)\Delta(lnT)\right] \quad (1)$$

Here, RH is relative humidity and ΔRH is relative humidity bias. T is air temperature and ΔT is air temperature bias. q is specific humidity or water vapor mixing ratio and Δq is specific humidity or water vapor mixing ratio bias. $c_p = 1004\ \mathrm{J\ K^{-1}\ kg^{-1}}$ is the specific heat capacity of air. $L = 2.5 \times 10^6\ \mathrm{J\ kg^{-1}}$ is latent heat of vapourization. $R_d = 287\ \mathrm{J\ K^{-1}\ kg^{-1}}$ and $R_v = 461\ \mathrm{J\ K^{-1}\ kg^{-1}}$ are specific gas constants for dry air and water vapor, respectively. The detailed derivation of this equation can be found in Supporting Information S1 file.

To further understand the reasons of the tropospheric relative humidity biases in CMIP models, we show in Figure 12 the tropospheric relative humidity biases in CMIP6 models estimated using Equation 1 and the tropospheric air temperature and specific humidity biases in CMIP6 models showed in Figures 3 and 6. Similar results are found in CMIP3 and CMIP5 models too. As expected, the estimated tropospheric relative humidity biases in CMIP6 models (Figure 12) are very similar to the actual tropospheric relative humidity biases in CMIP6 models showed in Figure 10 (the spatial pattern Pearson correlation is around 0.7). This gives us some confidence on Equation 1. In the tropical free troposphere, the relative humidity biases are mainly from the specific humidity biases while the contribution of the air temperature biases is small (Figure 12). This is consistent with the high resemblance of the tropical free tropospheric relative humidity and specific humidity bias patterns and the low resemblance of the tropical free tropospheric relative humidity and air temperature bias patterns as discussed earlier (Figure 11). In the tropical boundary layer, the contributions from the air temperature biases and the specific humidity biases to the relative humidity biases are both important. In the middle and high latitude troposphere, both the air temperature biases and the specific humidity biases are also important to the relative humidity biases. First looking at 300 hPa in the middle and high latitude troposphere. At this level, both the cold biases and the positive specific humidity biases contribute to the positive relative humidity biases with a higher contribution from the cold biases than the positive specific humidity biases. As the altitude decreases, the contribution of the extratropical cold biases to the extratropical relative humidity biases decreases. In particular, the extratropical cold biases become extratropical warm biases over the Southern Ocean and these warm biases cause negative relative humidity biases that reduce the positive relative humidity biases generated by the positive specific humidity biases over the Southern Ocean. This is particularly true in the boundary layer. As a result, the boundary-layer relative humidity biases are small although the boundary-layer specific humidity biases are large over the Southern Ocean.

4. Summary and Conclusions

This study seeks to quantify the tropospheric air temperature, specific humidity, and relative humidity biases in the CMIP6 models using the AIRS Obs4MIPs V2.1 data (Tian & Hearty, 2020) and possible bias reduction from the CMIP3 (Meehl et al., 2007) and CMIP5 (Taylor et al., 2012) models to the CMIP6 (Eyring et al., 2016) models. Here, we have analyzed outputs of 91 CMIP models including 22 CMIP3 models, 35 CMIP5 models, and 34 CMIP6 models (Table S1 in the Supporting Information S1). We have focused on the MMEMs, the long-term (~20 years) annual means, the model annual mean bias (CMIP–AIRS) maps, and global statistics at the eight standard tropospheric pressure levels from 1000 hPa to 300 hPa.

As a quality check and validation of the AIRS Obs4MIPs V2.1 data, we first inter-compare the AIRS, ERA5, and MERRA-2 data. We find that the AIRS and ERA5 tropospheric air temperature, specific humidity, and relative humidity data are highly consistent with each other (Figures 1, 4, and 8), that is, biases are small, pattern correlations are high and RMSDs are small between AIRS and ERA5 (Figures 2, 5 and 9). The global mean tropospheric air temperature, specific humidity, and relative humidity between ERA5 and AIRS are smaller than 0.3 K, 0.1 g $\mathrm{kg^{-1}}$ or 20%, and 5% at all pressure levels (Figures 2, 5 and 9). This indicates that the AIRS Obs4MIPs V2.1 tropospheric air temperature, specific and relative humidity data are of sufficient quality to be used as a reference data set for climate model evaluation. On the other hand, the differences between the MERRA-2 and AIRS tropospheric air temperature, specific and relative humidity data are much larger than the differences between the ERA5 and AIRS tropospheric air temperature, specific and relative humidity data, especially for specific and relative humidity over the tropical moist deep convective regions in the free troposphere, where

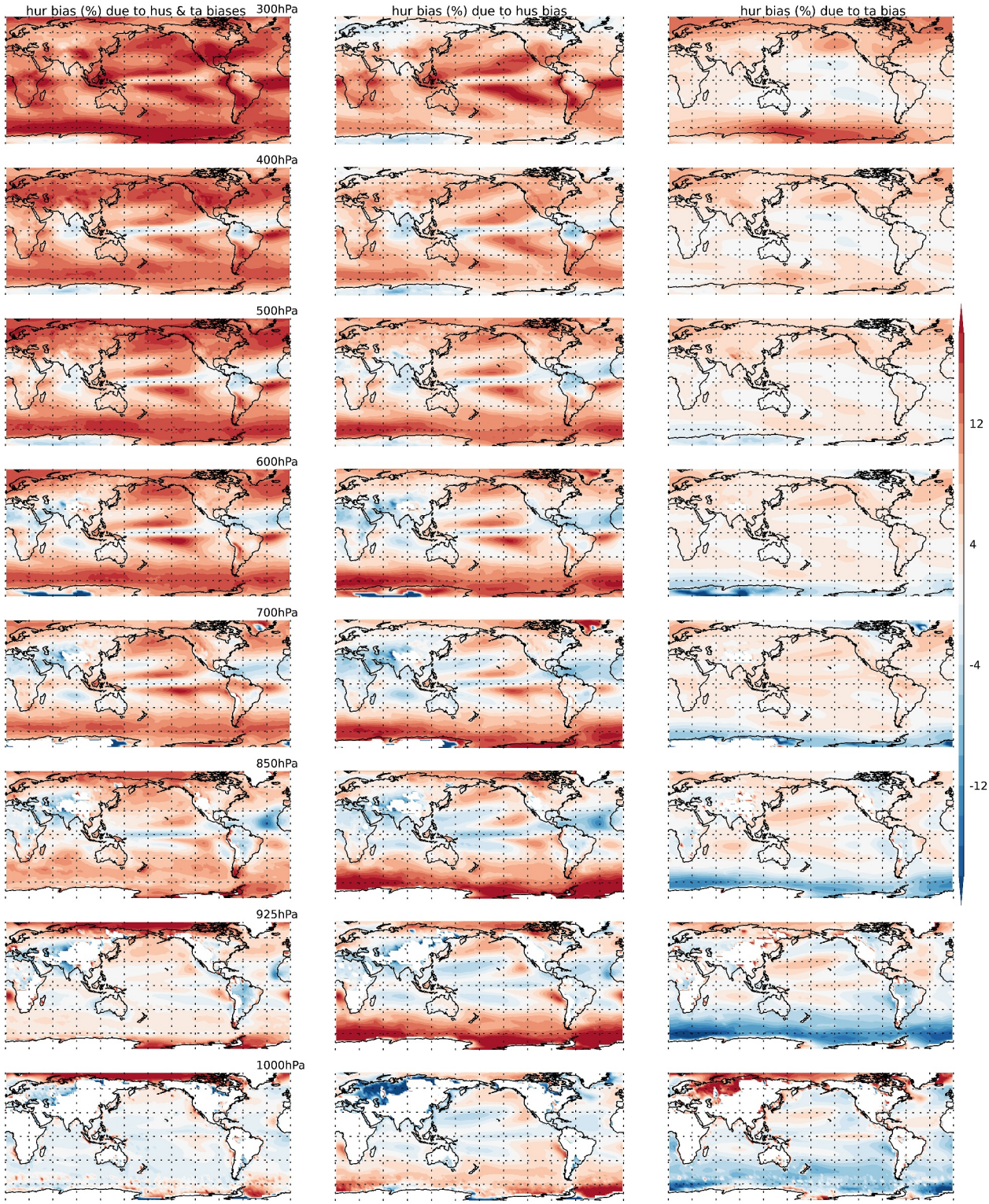

Figure 12. Tropospheric relative humidity (hur) biases (%) in CMIP6 models estimated using Equation 1 (first column) and its contributions from the tropospheric specific humidity (second column) and air temperature (third column) biases in CMIP6 models.

MERRA-2 is shown to be much moister than both AIRS and ERA5 (Figures 1, 2, 4, 5, 8, and 9). These are consistent with previous studies showing MERRA-2 tropospheric specific humidity is biased too moist in the upper troposphere (Gelaro et al., 2017).

We then compare the MMEMs of the CMIP3, CMIP5, and CMIP6 models against the AIRS Obs4MIPs V2.1 data. Our analyses indicate that all three phases of CMIP models share similar tropospheric air temperature, specific humidity, and relative humidity biases relative to AIRS. These tropospheric temperature and humidity biases in the CMIP models may impact the CMIP model ECS (Tian, 2015) and the future climate projections based on these models.

First, CMIP tropospheric air temperature bias: In the free troposphere (700−300 hPa), cold biases up to 4 K exist over almost the whole globe with maxima (~4 K) over the middle and high latitudes (poleward of 30°S and 30°N), especially over the mid-latitude storm tracks in the oceans, and minima (~1 K) over the tropics (30°S−30°N), especially over the subtropical Pacific (Figure 3). In the boundary layer (1000−850 hPa), cold biases are still prevalent over most regions. However, warm biases of up to 2 K are found over the Southern Ocean near the Antarctic coast and the southeastern Pacific along the coast of Peru as well as Amazonia and subtropical South America (especially at 925 hPa) (Figure 3). These warm biases start at the lower free troposphere and become larger and wider (northward) in the boundary layer. For all three CMIP phases, the global mean tropospheric air temperature biases are negative at all pressure levels. They maximize (~2 K) at 300 hPa and decrease with the decreasing altitude with a minimum (<0.5 K) at 925 hPa in the boundary layer (Figure 2a). The CMIP6 models have the smallest global mean tropospheric cold biases, the highest pattern correlations, and the smallest RMSDs among CMIP3/5/6 models (Figures 2c and 2d), indicating a possible cold bias reduction from CMIP3 and CMIP5 to CMIP6.

Second, CMIP tropospheric specific humidity bias: Over the middle and high latitudes, negative specific humidity biases are seen in the upper free troposphere at altitudes above 600 hPa (300−500 hPa) while positive specific humidity biases are seen in the lower troposphere below 600 hPa (700−1000 hPa) with a gradual transition as the altitude decreases from 300 to 1000 hPa (Figure 6). Over the tropics, positive free tropospheric specific humidity biases are found over the off-equatorial oceans, including the North Pacific, South Indian, Southeastern Pacific, and South Atlantic Oceans, while negative free tropospheric specific humidity biases can be seen near the equator (Figure 6). As the altitude decreases from 300 to 925 hPa, the area of positive free tropospheric specific humidity biases becomes smaller and narrower and moves further eastward over the tropical oceans, especially over the tropical Pacific Ocean, while the area of negative free tropospheric specific humidity biases becomes larger and wider, and their magnitudes increase. In the boundary layer negative specific humidity biases can be seen almost everywhere except for small parts of the eastern Pacific and eastern Atlantic Oceans (Figure 6). The tropical free tropospheric specific humidity bias pattern is similar to the tropical precipitation bias pattern and related to the double-ITCZ bias in the models (Tian, 2015; Tian & Dong, 2020). The global mean tropospheric specific humidity biases are positive for all three CMIP phases above 600 hPa (300−500 hPa) but have different signs for different CMIP ensembles below 600 hPa (700−1000 hPa): Negative for CMIP3 and CMIP5 but positive for CMIP6 (Figures 5a and 5b). Among all three CMIP phases, CMIP6 models have the highest pattern correlations and the smallest RMSDs relative to AIRS while CMIP3 models have the lowest pattern correlations and the largest RMSDs relative to AIRS at all tropospheric pressure levels except 300 hPa (Figures 5d and 5e). This indicates a possible improvement from CMIP3 to CMIP5 and to CMIP6 in tropospheric specific humidity simulation at all tropospheric pressure levels except 300 hPa.

Third, CMIP tropospheric relative humidity bias: The relative humidity biases are generally small (<10%) in the boundary layer but large (up to 20%) in the free troposphere (Figure 10). Over the middle and high latitudes, large positive free tropospheric relative humidity biases are found over the oceans (Figure 10). Here, both the air temperature biases (Figure 3) and the specific humidity biases (Figure 6) are important to the relative humidity biases (Figures 10 and 12). In the tropics, the free tropospheric relative humidity bias pattern (Figure 10) closely resembles the free tropospheric specific humidity bias pattern (Figure 6). Large positive free tropospheric relative humidity biases are generally found over the North Pacific, South Indian, Southeastern Pacific, and South Atlantic Oceans, while small negative free tropospheric relative humidity biases are seen over Amazonia, central America, the equatorial North Atlantic, south Asia including the Bay of Bengal, the equatorial Indian Ocean, the Maritime Continent, the equatorial Pacific, and middle and high latitudes (Figure 10). The tropical free tropospheric relative humidity biases (Figure 10) are mainly from the tropical free tropospheric specific humidity biases (Figure 6)

instead of the tropical free tropospheric air temperature biases (Figure 3) and related to the double-ITCZ bias in the models (Tian, 2015; Tian & Dong, 2020) (Figure 12). For all three CMIP phases, the global mean tropospheric relative humidity biases are positive at all pressure levels (Figure 9a). CMIP6 models have the largest while CMIP5 models have the smallest global mean tropospheric relative humidity biases except for 1000 hPa. Thus, from this perspective, there is no obvious improvement in CMIP6 models in comparison to CMIP3 or CMIP5 models. In the free troposphere, CMIP6 models have the highest pattern correlation and the smallest RMSD while CMIP3 models have the lowest pattern correlation and the largest RMSD relative to AIRS among all three CMIP phases (Figures 9c and 9d). This result indicates a possible improvement from CMIP3 to CMIP5 and then to CMIP6 in the free tropospheric relative humidity simulation. In the boundary layer, the situation is different and complicated and there is no obvious improvement in CMIP6 models in comparison to CMIP3 or CMIP5 models.

Data Availability Statement

References for the data sets used in this research are: Tian and Hearty (2020), Hersbach et al. (2020), Gelaro et al. (2017), Meehl et al. (2007), Taylor et al. (2012), and Eyring et al. (2016). The AIRS Obs4MIPs V2.1 data (Tian & Hearty, 2020) are available on the ESGF Obs4MIPs website (https://esgf-node.llnl.gov/projects/obs4mips/). The ERA5 data (Hersbach et al., 2020) are available on the Copernicus Climate Change Service (C3S) Climate Date Store (https://cds.climate.copernicus.eu/). The MERRA-2 data (Gelaro et al., 2017) are available on the NASA GSFC GMAO website (https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/data_access/). The CMIP3 (Meehl et al., 2007), CMIP5 (Taylor et al., 2012), and CMIP6 (Eyring et al., 2016) data are available on the ESGF CMIP3 (https://esgf-node.llnl.gov/search/cmip3/), CMIP5 (https://esgf-node.llnl.gov/search/cmip5/), and CMIP6 (https://esgf-node.llnl.gov/search/cmip6/) website, respectively.

Acknowledgments

This research was performed at the Jet Propulsion Laboratory (JPL), California Institute of Technology (Caltech) under a contract with NASA (80NM0018D0004) and supported by the NASA Science of Terra, Aqua, and Suomi NPP (TASNPP) program (NNH17ZDA001N-TASNPP) administered by Dr. Tsengdar Lee, the NASA Precipitation Measurement Mission Science Team (PMMST) program (NNH21ZDA001N-PMMST) administered by Dr. Will McCarty, the NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) program (NNH17ZDA001N-MEASURES) administered by Dr. Lucia Tsaoussi, and the AIRS Project and internal fund at JPL. We acknowledge the WCRP Working Group on Coupled Modeling (WGCM), which is responsible for CMIP, and the U. S. DOE's Program for Climate Model Diagnosis and Intercomparison (PCMDI), which provides coordinating support and leads development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. We thank the climate modeling groups around the world for producing and making available their model outputs. We also want to express our gratitude to Dr. Evan Fishbein for helpful discussions regarding the AIRS relative humidity calculation and three anonymous reviewers for their constructive comments and suggestions.

References

Aumann, H. H., Chahine, M. T., Gautier, C., Goldberg, M. D., Kalnay, E., McMillin, L., et al. (2003). AIRS/AMSU/HSB on the Aqua mission: Design, science objectives, data products, and processing systems. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(2), 253–264. https://doi.org/10.1109/tgrs.2002.808356

Bock, L., Lauer, A., Schlund, M., Barreiro, M., Bellouin, N., Jones, C., et al. (2020). Quantifying progress across different CMIP phases with the ESMValTool. *Journal of Geophysical Research*, *125*(21), e2019JD032321. https://doi.org/10.1029/2019jd032321

Chahine, M. T., Pagano, T. S., Aumann, H. H., Atlas, R., Barnet, C., Blaisdell, J., et al. (2006). AIRS: Improving weather forecasting and providing new data on greenhouse gases. *Bulletin America Meteorology Social*, *87*(7), 911–926. https://doi.org/10.1175/bams-87-7-911

Cinquini, L., Crichton, D., Mattmann, C., Harney, J., Shipman, G., Wang, F., et al. (2014). The Earth System Grid Federation: An open infrastructure for access to distributed geospatial data. *Future Generation Computer Systems*, *36*, 400–417. https://doi.org/10.1016/j.future.2013.07.002

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-interim reanalysis: Configuration and performance of the data assimilation system. *The Quarterly Journal of the Royal Meteorological Society*, *137*(656), 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., et al. (2006). Validation of Atmospheric Infrared Sounder temperature and water vapor retrievals with matched radiosonde measurements and forecasts. *Journal of Geophysical Research*, *111*(D9), D09s15. https://doi.org/10.1029/2005jd006116

Eyring, V., Bock, L., Lauer, A., Righi, M., Schlund, M., Andela, B., et al. (2020). Earth System Model Evaluation Tool (ESMValTool) v2.0-an extended set of large-scale diagnostics for quasi-operational and comprehensive evaluation of Earth system models in CMIP. *Geoscientific Model Development*, *13*(7), 3383–3438. https://doi.org/10.5194/gmd-13-3383-2020

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization [Dataset]. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Fasullo, J. T., & Trenberth, K. E. (2012). A less cloudy future: The role of subtropical subsidence in climate sensitivity. *Science*, *338*(6108), 792–794. https://doi.org/10.1126/science.1227465

Ferraro, R., Waliser, D. E., Gleckler, P., Taylor, K. E., & Eyring, V. (2015). Evolving Obs4MIPs to support Phase 6 of the Coupled Model Intercomparison Project (CMIP6). *Bulletin America Meteorology Social*, *96*(8), ES131–ES133. https://doi.org/10.1175/bams-d-14-00216.1

Gates, W. L., Boyle, J. S., Covey, C., Dease, C. G., Doutriaux, C. M., Drach, R. S., et al. (1999). An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bulletin America Meteorology Social*, *80*(1), 29–55. https://doi.org/10.1175/1520-0477(1999)080<0029:aootro>2.0.co;2

Gelaro, R., McCarty, W., Suarez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) [Dataset]. *Journal of Climate*, *30*(14), 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. *Journal of Climate*, *19*(23), 6104–6121. https://doi.org/10.1175/jcli3956.1

Gettelman, A., Weinstock, E. M., Fetzer, E. J., Irion, F. W., Eldering, A., Richard, E. C., et al. (2004). Validation of Aqua satellite data in the upper troposphere and lower stratosphere with in situ aircraft instruments. *Geophysical Research Letters*, *31*(22), L22107. https://doi.org/10.1029/2004gl020730

Hearty, T. J., Savtchenko, A., Tian, B., Fetzer, E., Yung, Y. L., Theobald, M., et al. (2014). Estimating sampling biases and measurement uncertainties of AIRS/AMSU-A temperature and water vapor observations using MERRA reanalysis. *Journal of Geophysical Research*, *119*(6), 2725–2741. https://doi.org/10.1002/2013jd021205

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis [Dataset]. *The Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

IPCC (2007). In S. Solomon, D. Qin, M. R. Manning, Z. Chen, et al. (Eds.), *Climate change 2007: The physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change* (p. 996). Cambridge University Press.

IPCC (2013). In T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, et al. (Eds.), *Climate change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change* (p. 1535). Cambridge University Press. https://doi.org/10.1017/CBO9781107415324

IPCC (2021). In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, et al. (Eds.), *Climate change 2021: The physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change*. Cambridge University Press. https://doi.org/10.1017/9781009157896

Jiang, J. H., Su, H., Wu, L. T., Zhai, C. X., & Schiro, K. A. (2021). Improvements in cloud and water vapor simulations over the tropical oceans in CMIP6 compared to CMIP5. *Earth and Space Science*, *8*(5), e2020EA001520. https://doi.org/10.1029/2020ea001520

Jiang, J. H., Su, H., Zhai, C. X., Perun, V. S., Del Genio, A., Nazarenko, L. S., et al. (2012). Evaluation of cloud and water vapor simulations in CMIP5 climate models using NASA "A-Train" satellite observations. *Journal of Geophysical Research*, *117*(D14), D14105. https://doi.org/10.1029/2011jd017237

John, V. O., & Soden, B. J. (2007). Temperature and humidity biases in global climate models and their impact on climate feedbacks. *Geophysical Research Letters*, *34*(18), L18704. https://doi.org/10.1029/2007gl030429

Lambrigtsen, B. H. (2003). Calibration of the AIRS microwave instruments. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(2), 369–378. https://doi.org/10.1109/tgrs.2002.808247

Lauer, A., Eyring, V., Bellprat, O., Bock, L., Gier, B. K., Hunter, A., et al. (2020). Earth System Model Evaluation Tool (ESMValTool) v2.0-diagnostics for emergent constraints and future projections from Earth system models in CMIP. *Geoscientific Model Development*, *13*(9), 4205–4228. https://doi.org/10.5194/gmd-13-4205-2020

Meehl, G. A., Boer, G. J., Covey, C., Latif, M., & Stouffer, R. J. (1997). Intercomparison makes for a better climate model. *Eos*, *78*(41), 445–451. https://doi.org/10.1029/97eo00276

Meehl, G. A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J. F. B., et al. (2007). The WCRP CMIP3 multimodel dataset - A new era in climate change research [Dataset]. *Bulletin America Meteorology Social*, *88*, (9), 1383–1394. https://doi.org/10.1175/bams-88-9-1383

Meehl, G. A., Covey, C., McAvaney, B., Latif, M., & Stouffer, R. J. (2005). Overview of the Coupled Model Intercomparison Project. *Bulletin America Meteorology Social*, *86*, 89–93. https://doi.org/10.1175/bams-86-1-89

Murphy, D. M., & Koop, T. (2005). Review of the vapour pressures of ice and supercooled water for atmospheric applications. *The Quarterly Journal of the Royal Meteorological Society*, *131*(608), 1539–1565. https://doi.org/10.1256/qj.04.94

Olsen, E. T., Fetzer, E. J., Hulley, G., Kalmus, P., Manning, E., Wong, S. et al. (2013). AIRS/AMSU/HSB version 6 level 2 product user guide. Retrieved from https://docserver.gesdisc.eosdis.nasa.gov/repository/Mission/AIRS/3.3_ScienceDataProductDocumentation/3.3.4_ProductGenerationAlgorithms/V6_L2_Product_User_Guide.pdf

Pierce, D. W., Barnett, T. P., Fetzer, E. J., & Gleckler, P. J. (2006). Three-dimensional tropospheric water vapor in coupled climate models compared with observations from the AIRS satellite system. *Geophysical Research Letters*, *33*(21), L21701. https://doi.org/10.1029/2006gl027060

Priestley, M. D. K., Ackerley, D., Catto, J. L., Hodges, K. I., McDonald, R. E., & Lee, R. W. (2020). An overview of the extratropical storm tracks in CMIP6 historical simulations. *Journal of Climate*, *33*(15), 6315–6343. https://doi.org/10.1175/jcli-d-19-0928.1

Righi, M., Andela, B., Eyring, V., Lauer, A., Predoi, V., Schlund, M., et al. (2020). Earth System Model Evaluation Tool (ESMValTool) v2.0-technical overview. *Geoscientific Model Development*, *13*(3), 1179–1199. https://doi.org/10.5194/gmd-13-1179-2020

Su, H., Jiang, J. H., Zhai, C. X., Shen, T. J., Neelin, J. D., Stephens, G. L., & Yung, Y. L. (2014). Weakening and strengthening structures in the Hadley Circulation change under global warming and implications for cloud response and climate sensitivity. *Journal of Geophysical Research*, *119*(10), 5787–5805. https://doi.org/10.1002/2014jd021642

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 Transactions on Geoscience and Remote Sensing*, *41*(2), 390–409. https://doi.org/10.1109/tgrs.2002.808236

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. *Journal of Applied Remote Sensing*, *8*(1), 084994. https://doi.org/10.1117/1.jrs.8.084994

Takahashi, H., Su, H., Jiang, J. H., Luo, Z. J., Xie, S., & Hafner, J. (2013). Tropical water vapor variations during the 2006-2007 and 2009-2010 El Ninos: Satellite observation and GFDL AM2.1 simulation. *Journal of Geophysical Research*, *118*(16), 8910–8920. https://doi.org/10.1002/jgrd.50684

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design [Dataset]. *Bulletin America Meteorology Social*, *93*(4), 485–498. https://doi.org/10.1175/bams-d-11-00094.1

Teixeira, J., Waliser, D., Ferraro, R., Gleckler, P., Lee, T., & Potter, G. (2014). Satellite observations for CMIP5: The genesis of Obs4MIPs. *Bulletin America Meteorology Social*, *95*(9), 1329–1334. https://doi.org/10.1175/bams-d-12-00204.1

Tian, B. (2015). Spread of model climate sensitivity linked to double-Intertropical Convergence Zone bias. *Geophysical Research Letters*, *42*(10), 4133–4141. https://doi.org/10.1002/2015gl064119

Tian, B., & Dong, X. (2020). The double-ITCZ bias in CMIP3, CMIP5, and CMIP6 models based on annual mean precipitation. *Geophysical Research Letters*, *47*(8), e2020GL087232. https://doi.org/10.1029/2020gl087232

Tian, B., Fetzer, E. J., Kahn, B. H., Teixeira, J., Manning, E., & Hearty, T. (2013a). Evaluating CMIP5 models using AIRS tropospheric air temperature and specific humidity climatology. *Journal of Geophysical Research*, *118*(1), 114–134. https://doi.org/10.1029/2012jd018607

Tian, B., Fetzer, E. J., & Manning, E. M. (2019). The atmospheric infrared sounder Obs4MIPs version 2 data set. *Earth and Space Science*, *6*, 324–333. https://doi.org/10.1029/2018ea000508

Tian, B., & Hearty, T. J. (2020). Estimating and removing the sampling biases of the AIRS Obs4MIPs V2 data [Dataset]. *Earth and Space Science*, *7*(12), e2020EA001438. https://doi.org/10.1029/2020ea001438

Tian, B., Manning, E., Fetzer, E. J., Olsen, E. T., Wong, S., Susskind, J. et al. (2013b). AIRS/AMSU/HSB Version 6 Level 3 product user guide. Retrieved from https://docserver.gesdisc.eosdis.nasa.gov/repository/Mission/AIRS/3.3_ScienceDataProductDocumentation/3.3.4_ProductGenerationAlgorithms/V6_L3_User_Guide.pdf

Tobin, D. C., Revercomb, H. E., Knuteson, R. O., Lesht, B. M., Strow, L. L., Hannon, S. E., et al. (2006). Atmospheric radiation measurement site atmospheric state best estimates for atmospheric infrared sounder temperature and water vapor retrieval validation. *Journal of Geophysical Research*, *111*(D9), D09s14. https://doi.org/10.1029/2005jd006103

Waliser, D., Gleckler, P. J., Ferraro, R., Taylor, K. E., Ames, S., Biard, J., et al. (2020). Observations for Model Intercomparison Project (Obs4MIPs): Status for CMIP6. *Geoscientific Model Development*, *13*(7), 2945–2958. https://doi.org/10.5194/gmd-13-2945-2020

Weigel, K., Bock, L., Gier, B. K., Lauer, A., Righi, M., Schlund, M., et al. (2021). Earth System Model Evaluation Tool (ESMValTool) v2.0-diagnostics for extreme events, regional and impact evaluation, and analysis of Earth system models in CMIP. *Geoscientific Model Development*, *14*(6), 3159–3184. https://doi.org/10.5194/gmd-14-3159-2021