

## How Accurate Are Modern Atmospheric Reanalyses for the Data-Sparse Tibetan Plateau Region?

XINGHUA BAO

*State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing, China*

FUQING ZHANG

*Center for Advanced Data Assimilation and Predictability Techniques, and Department of Meteorology and Atmospheric Science, The Pennsylvania State University, State College, Pennsylvania*

(Manuscript received 22 October 2018, in final form 8 April 2019)

### ABSTRACT

More than 6000 independent radiosonde observations from three major Tibetan Plateau experiments during the warm seasons (May–August) of 1998, 2008, and 2015–16 are used to assess the quality of four leading modern atmospheric reanalysis products (CFSR/CFSv2, ERA-Interim, JRA-55, and MERRA-2), and the potential impact of satellite data changes on the quality of these reanalyses in the troposphere over this data-sparse region. Although these reanalyses can reproduce reasonably well the overall mean temperature, specific humidity, and horizontal wind profiles against the benchmark independent sounding observations, they have nonnegligible biases that can be potentially bigger than the analysis-simulated mean regional climate trends over this region. The mean biases and mean root-mean-square errors of winds, temperature, and specific humidity from almost all reanalyses are reduced from 1998 to the two later experiment periods. There are also considerable differences in almost all variables across different reanalysis products, though these differences also become smaller during the 2008 and 2015–16 experiments, in particular for the temperature fields. The enormous increase in the volume and quality of satellite observations assimilated into reanalysis systems is likely the primary reason for the improved quality of the reanalyses during the later field experiment periods. Besides differences in the forecast models and data assimilation methodology, the differences in performance between different reanalyses during different field experiment periods may also be contributed by differences in assimilated information (e.g., observation input sources, selected channels for a given satellite sensor, quality-control methods).

### 1. Introduction

Reanalyses provide comprehensive, gridded estimates of past atmospheric states at regular intervals over long time periods and have been widely used to study the global and regional climate trends (Bengtsson et al. 2004; Bosilovich 2013; Marshall 2003; Manzanas et al. 2014; Chang and Yau 2016; Kishore et al. 2016), especially in data-sparse regions (Nicolas and Bromwich 2014; Lavaysse et al. 2016; Robson et al. 2016). It has been noted that different atmospheric reanalysis systems may give considerably different results for the same diagnostic quantities due to difference in technical details (model characteristics, horizontal and vertical resolution, the top level, physical parameterizations, boundary conditions,

and assimilation scheme, etc.) or observation data assimilated in the reanalysis systems (Bao and Zhang 2013; Fujiwara et al. 2017). In particular, the input observation data assimilated in reanalyses, including conventional and satellite data, have generally become denser over time. On the other hand, the availability of these data is ever evolving, in particular with regard to the introduction and retirement of the observation instruments. Such changes may have strong impacts on the quality of the reanalyses (Fujiwara et al. 2017). Therefore, before using reanalyses in the study of weather and climate, in particular their trends and variability over long periods of time, quantitative uncertainty estimates are crucial (Thorne and Vose 2010; Parker 2016). Many recent studies have evaluated the performance and trends of various reanalyses from different sources for different regions (Serreze et al. 2012; Siam et al. 2013; Lindsay et al.

Corresponding author: Xinghua Bao, baoxh@cma.gov.cn

2014; Jones et al. 2016; Simmons et al. 2004, 2010, 2014, 2017). Limited by the lack of quality and independent observations that are not already used in the reanalyses, few studies have documented the accuracy and trends of the aboveground variables in various reanalyses over data-sparse regions (Bao and Zhang 2013; Chen et al. 2014; Dufour et al. 2016; Davis et al. 2017; Manney et al. 2017).

Satellite data represent the majority of the input observations assimilated in most reanalyses and its proportion continues to grow. For example, the percentage of the observation assimilated in National Aeronautics and Space Administration (NASA) Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2), that is measured by satellites increases from just over 60% in January 1980 to almost 90% in December 2014; meanwhile, the total observation count has increased more than 20 times (McCarty et al. 2016). So it is not surprising that reanalyses are sensitive to variations in the amount and type of satellite observations being assimilated (Kalnay et al. 1996; Bosilovich et al. 2011; Robertson et al. 2011, 2014). Bosilovich et al. (2017) evaluated the global average precipitation and evaporation from MERRA-2 and several other contemporary reanalyses, identifying discontinuities in some satellite data reanalyses associated with satellite data changes. For example, the National Centers for Environmental Prediction (NCEP) Climate Forecast System Reanalysis (CFSR) and MERRA precipitation increases sharply at the introduction of Advanced Microwave Sounding Unit-A (AMSU-A, the end of 1998). Furthermore, to identify differences among reanalyses and understand their underlying causes, the Stratosphere–Troposphere Processes and Their Role in Climate (SPARC) Reanalysis Intercomparison Project (S-RIP) is implemented by the World Meteorological Organization (WMO), International Council for Science (ICSU), and Intergovernmental Oceanographic Commission (IOC) of United Nations Educational, Scientific and Cultural Organization (UNESCO) (Fujiwara et al. 2017). As part of S-RIP, Long et al. (2017) intercompared the temperature and wind over the satellite era during 1979–2014, Davis et al. (2017) intercompared the upper troposphere and stratosphere water vapor and ozone from the five more recent reanalyses [CFSR, MERRA, European Centre for Medium-Range Weather Forecasts interim reanalysis (ERA-Interim), Japanese 55-year Reanalysis (JRA-55), and MERRA-2] and several older reanalyses (NCEP–NCAR, NCEP–DOE, ERA-40, and JRA-25). These S-RIP studies pointed out almost all the reanalyses have a temporal discontinuity at the end of 1998 when the observing system of microwave and infrared

sounders (whose observations are the most prevalent satellite data assimilated by reanalyses) transferred from Television Infrared Observation Satellite (TIROS) Operational Vertical Sounder (TOVS) to Advanced TIROS Operational Vertical Sounder (ATOVS). Long et al. (2017) indicated the temperature and wind variances among the reanalyses became smaller from the TOVS period to the ATOVS period. Moreover, the observations from the hyperspectral satellite instruments such as the Atmospheric Infrared Sounder (AIRS) and Infrared Atmospheric Sounding Interferometer (IASI) have varying degree of beneficial impacts on forecast quality of different data assimilation systems (Gelaro and Zhu 2009; Collard and McNally 2009; Singh et al. 2012). It is also reported that the assimilation of global positioning system (GPS) radio occultation (RO) observations as “anchor observations” not only directly helps reduce the temperature biases in ERA-Interim (and likely other reanalyses as well) but also provides better bias corrections for satellite radiances (Poli et al. 2010; Cucurull et al. 2014). However, these studies mostly examined the impacts of satellite data changes on reanalysis variables in the stratosphere or the global water cycle (precipitation and evaporation), with little attention to the aboveground variables in the troposphere over mountainous data-sparse regions.

The Tibetan Plateau over central Asia is the largest and highest plateau in the world. Due to its unique dynamic and thermodynamic forcing induced by the vast landmass and high terrains, Tibetan Plateau exerts significant influence on the regional and global climate (Ye 1981; Ye and Wu 1998; Molnar et al. 2010; Bao et al. 2011; Si and Ding 2013). In the past 20 years, three large Tibetan Plateau-related field experiments were conducted: the Second Tibetan Plateau Experiment in 1998 (TIPEX-II) (Xu et al. 2002), the China-Japan Meteorological Disaster Reduction Cooperation Research Center Project during 2005–09 (JICA/Tibet Project) (Zhang et al. 2012), and the Third Tibetan Plateau Atmospheric Scientific Experiment, which originally began in 2014 and is still ongoing (TIPEX-III) (Zhao et al. 2018). These three field experiments deployed their respective enhanced radiosonde observations collected during the intense observation periods (IOPs) over warm seasons of 1998, 2008, and 2015–16, respectively, almost all of which were not assimilated in any of the reanalyses (except that the field soundings during 1998 TIPEX-II were assimilated in JRA-55). Most notably, for these three field experiments, the TIPEX-II experiment in the summer of 1998 mostly belong to TOVS period, whereas the 2008 JICA/Tibet project and 2015–16 TIPEX-III are in the ATOVS period, as multiple other types of satellite observations such as AIRS, IASI,

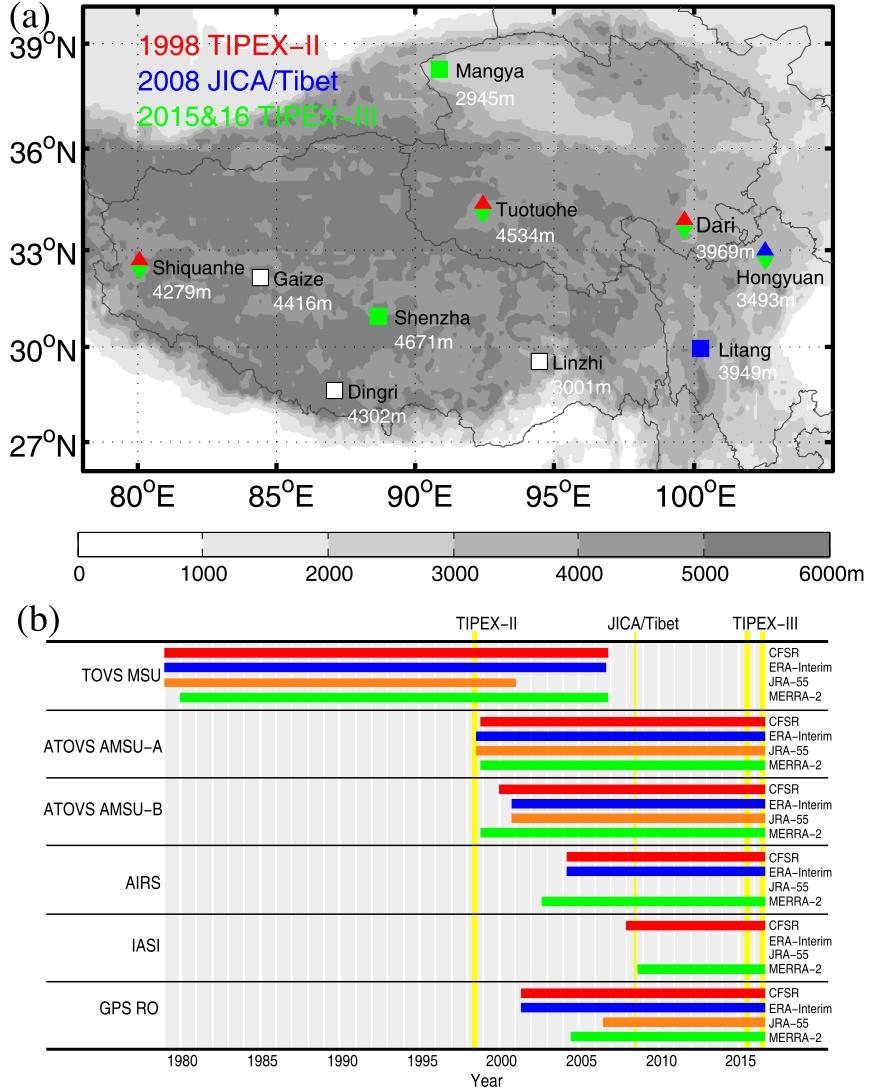


FIG. 1. (a) Locations of the independent radiosonde sites over the Tibetan Plateau during the 1998 TIPEX-II IOP, the 2008 JICA/Tibet Project IOP, and 2015–16 TIPEX-III IOPs. Gray shading represents topography. The white squares denote the same sounding stations for all 4 years, the red regular triangles, the blue regular triangles and the blue square, and the green inverted triangles and the green squares denote the other sounding stations during 1998 TIPEX-II IOP, 2008 JICA/Tibet Project IOP, and 2015–16 TIPEX-III IOPs, respectively. (b) Timelines of some primary satellite radiance observations assimilated by the CFSR (red), ERA-Interim (blue), JRA-55 (tan), and MERRA-2 (green) reanalysis systems [excerpt from Fujiwara et al. (2017)]. Yellow bars mark the IOPs of TIPEX-II, JICA/Tibet Project, and TIPEX-III.

and GPS RO anchoring observations were assimilated in these reanalyses successively since August 1998 (Fig. 1). Therefore, the unique field sounding dataset provides us a rare opportunity to assess the quality of several widely used modern atmospheric reanalysis products, and evaluate how sensitive they are to variations in the amount and type of the input satellite observations over Tibetan Plateau.

As an extension of our well-referenced earlier study of Bao and Zhang (2013) with regard to the number and extent of independent observations across different decades, here we further evaluate four leading atmospheric reanalysis datasets and discuss the impacts of the input satellite data variations on some aboveground variables of these reanalyses in the troposphere, using thousands of enhanced independent radiosonde observations

collected during the three major Tibetan Plateau field experiments that were conducted over three different decades: 1) CFSR (used during the periods of TIPEX-II and JICA/Tibet) (Saha et al. 2010) and NCEP Climate Forecast System, version 2, 6-hourly analysis products (CFSv2, used during the periods of TIPEX-III) (Saha et al. 2014); 2) ERA-Interim (Dee et al. 2011); 3) JRA-55 (Kobayashi et al. 2015); and 4) MERRA-2 (Gelaro et al. 2017).

## 2. Data and methodology

### a. Reanalysis data

In this study, four recent reanalysis products of NCEP CFSR/CFSv2, ERA-Interim, JRA-55, and NASA MERRA-2 are intercompared and evaluated. These reanalyses with high spatial resolution are the products of four major reanalysis centers. These reanalysis datasets are briefly described below.

NCEP CFSR and CFSv2 (Saha et al. 2010, 2014) are created by the second version of the NCEP Climate Forecast System, which is the first global reanalysis of the coupled atmosphere–ocean–sea ice system using three-dimensional variational data assimilation (3D-Var). CFSR covers the period from January 1979 to December 2010, while CFSv2 is from January 2011 to the present; they have the different horizontal grid spacing,  $0.3125^\circ$  (T382) for CFSR and  $0.2045^\circ$  (T574) for CFSv2, with the same 64 vertical levels up to  $\sim 0.266$  hPa. CFSv2 also has several changes in the physical parameterizations and data assimilation system relative to CFSR. Hereinafter CFSR and CFSv2 collectively referred to as CFSR for convenience.

ERA-Interim (Dee et al. 2011) is the product of ECMWF atmospheric reanalyses of the global climate from 1979 to the present. The data are produced using four-dimensional variational data assimilation (4D-Var) with a  $T_{L255}$  ( $\sim 79$  km) and L60 (60 vertical levels from the surface up to 0.1 hPa) spectral model.

JRA-55 (Kobayashi et al. 2015) is the second Japanese global atmospheric reanalysis conducted by JMA and covers the period from 1958 to the present. JRA-55 applies a 4D-Var data assimilation scheme. The spatial resolution of JRA-55 is  $\sim 55$  km ( $T_{L319}$ ), and the vertical levels include surface and 60 levels up to 0.1 hPa.

MERRA-2 (Gelaro et al. 2017) is the latest atmospheric reanalysis of the modern satellite era produced by NASA's GMAO covering 1980 to present. It is based on the Goddard Earth Observing System (GEOS), version 5.12.4, atmospheric data assimilation system, which uses 3D-Var assimilation with incremental analysis update (IAU) to constrain the analyses. MERRA-2

has the approximate horizontal resolution of  $0.5^\circ \times 0.625^\circ$  and 72 hybrid-eta levels from the surface to 0.01 hPa. Here we use the M216NPANA (inst6\_3d\_ana\_Np) data product available at 42 pressure levels. For more information about these reanalyses, please see Fujiwara et al. (2017).

Of note, the start times of ATOVS period are different for these reanalyses, related to when the AMSU-A radiances were introduced to the assimilation system. AMSU-A, one of three ATOVS instruments, is regarded as the most crucial satellite system with respect to tropospheric forecast skill scores (Gelaro and Zhu 2009). The assimilation of AMSU-A began in 28 October 1998 for CFSR, 2 August 1998 for ERA-Interim, 1 August 1998 for JRA-55, and 2 November 1998 for MERRA-2 (Fig. 1b).

### b. The intensive radiosonde data

To evaluate reanalyses, we use the independent radiosonde observation data collected during the IOP of three Tibetan Plateau experiments: TIPEX-II, JICA/Tibet Project, and TIPEX-III. For completely independent verifications, the radiosonde stations over the main body of the Tibetan Plateau we choose in this study are unconventional upper-air meteorological stations that do not belong to WMO Global Telecommunication System (GTS). Therefore, the intensive sounding observation data collected in these stations were not reported to WMO for international exchange. This means these sounding observations were not assimilated in any of the reanalyses. Figure 1a shows the location of these independent radiosonde stations during these three experiments.

The IOP of TIPEX-II was from 10 May to 9 August 1998. The independent sounding observations were from six unconventional sounding sites (Gaize, Dingri, Linzhi, Shiquanhe, Tuotuohe, and Dari), and were collected every 6 h or 4 times per day (0000, 0600, 1200, and 1800 UTC). It is worth noting that the 6-hourly off-line 1998 TIPEX-II dataset have been assimilated into JRA-55 but not into other three reanalyses. The Vaisala RS80 radiosonde was used for the sounding observations during TIPEX-II. Note that there was one week of overlap between the introduction of ATOVS for ERA-Interim and JRA-55 and the end of TIPEX-II. To show the difference of each reanalysis between TOVS and ATOVS periods more clearly, only the observations collected during 10 May–31 July 1998 are used to compare with that collected during other two campaigns.

The IOP of JICA/Tibet Project was from 20 June to 19 July 2008. Five independent sounding sites over Tibetan Plateau were designed with intensified upper air sounding activities at four times a day (0000, 0600, 1200, and 1800 UTC). Three types of radiosonde—GTS1

(Dingri and Hongyuan), 59–701 (Linzhi), and Vaisala RS92 (Gaize and Litang)—were used for the sounding observations during 2008 JICA/Tibet Project.

TIPEX-III implemented the intensive radiosonde observations at 3 times per day (0000, 0600, and 1200 UTC) over nine independent sounding sites (Gaize, Dingri, Linzhi, Shiquanhe, Tuotuohe, Dari, Hongyuan, Mangya, and Shenzha) during the 2015 IOP from 10 June to 31 August. Moreover, in 2016, the intensive radiosonde observing sites over Tibetan Plateau were six (Gaize, Dingri, Linzhi, Shiquanhe, Tuotuohe, and Shenzha) and the IOPs were from 1 June to 31 August. We combined the intensive sounding observations during the IOPs of 2015 and 2016 to evaluate some reanalyses in this study. Three types of radiosonde GTS1 (Dingri, Linzhi, Tuotuohe, Dari, Hongyuan, and Mangya), XGP-3 GZ sounder, and Vaisala RS92 (Gaize, Shiquanhe, and Shenzha) were used during 2015/16 TIPEX-III.

Quality-controlled data provided by the corresponding experiment team include the temperature  $T$ , dewpoint depression, wind direction, and wind speed at seven standard vertical levels (500, 400, 300, 250, 200, 150, and 100 hPa). As in Bao and Zhang (2013), we use these data to derive both components of the horizontal wind ( $U$  and  $V$ ). Besides, we calculate specific humidity ( $Q$ ) from temperatures and dewpoint temperatures according to the method described in Simmons et al. (1999), which is also used in the ERA-Interim forecast model (Dee et al. 2011). Here specific humidity is calculated with respect to liquid water when  $T > 0^\circ\text{C}$ , ice when  $T < -23^\circ\text{C}$ , and a mixed-phase function for temperatures in between.

The three most frequently used radiosonde types of Vaisala RS80, RS92, and GTS1 in three experiments had higher accuracy and better quality and stability compared with other types that used in the same period (Oakley 1998; Nash et al. 2006, 2011; Ingleby 2017). Vaisala RS80 with the GPS windfinding technology, made in Finland, was widely used in the world in the end of the twentieth century, then was gradually replaced by the newer Vaisala RS92 since 2003. Currently, the most used type also is Vaisala RS92, which is used by the Global Climate Observing System (GCOS) Reference Upper Air Network (GRUAN) as a reference radiosonde (Dirksen et al. 2014). Chinese-made GTS1 radiosonde with radar wind finding became operational in China since 2002 and represented a distinct improvement on the previous generation (59–701 mechanical radiosonde system). GTS1 showed good performance in the international radiosonde intercomparison (Nash et al. 2011; Ingleby 2017). These radiosonde systems have small interinstrument differences in temperature ( $<\pm 0.2^\circ\text{C}$ ) and wind speed ( $<\pm 0.2 \text{ m s}^{-1}$ ) in the mid–upper troposphere (Skrivankova 2004; Steinbrecht et al. 2008; Nash et al.

2011; Ingleby 2017), but they have different degrees of dry bias (underestimate the amount of water vapor) at low temperatures where the sensor response has historically been too slow, especially below  $-40^\circ\text{C}$  where RH measurements are unreliable (Miloshevich et al. 2001; Vömel et al. 2007; Bian et al. 2011). Therefore, the radiosonde humidity measurements above 300 hPa are often excluded from the assimilation in the reanalysis systems.

### c. Methods used in the intercomparison

For direct comparison of the gridded reanalysis with discrete soundings, as in Bao and Zhang (2013), we first interpolate these pressure coordinate reanalysis products (with simple bilinear interpolation) to each of the sounding locations at the same synoptic times and standard pressure levels when all of data sources are available. Then we reject the abnormal observation values that differ from any reanalysis under a subjective criterion (i.e., if  $\Delta U > 30 \text{ m s}^{-1}$ , and/or  $\Delta V > 30 \text{ m s}^{-1}$ , as well as  $\Delta T > 20^\circ\text{C}$ ,  $Q$  will be rejected when  $T$  is eliminated). Under these criteria, no more than 0.3% observations are rejected for each variable during three experiments. Next, we compare the radiosonde data to the four reanalysis datasets giving mean, bias, standard deviation (STD), and root-mean-square error (RMSE) reanalysis-minus-observation (R-O) statistics.

## 3. Major findings

### a. The climatological mean profile, mean variability, and year-to-year variation

At the first glance, the overall mean profiles for horizontal winds, temperature, and specific humidity from each analysis product follow well the sounding means over each of the three field experiment periods, that is, 1998 in Figs. 2a–c, 2008 in Figs. 2e–h, and 2015–16 in Figs. 3i–l. Typical of the warm-season climatology over this region, there is a moderate westerly mean flow maximum around 200 hPa. However, there is strong difference in the maximum values from the three different field campaign periods that vary from  $\sim 18 \text{ m s}^{-1}$  in 1998 to  $\sim 8 \text{ m s}^{-1}$  in 2008 and  $\sim 10 \text{ m s}^{-1}$  for 2015–16. The mean meridional flow is rather weak throughout the vertical profile with weak southerly flows near the ground and slightly larger northerly wind near the tropopause, as a reflection of the regional mean Hadley circulation over this longitude range. The mean temperature profiles nearly overlap with each other indicating that the mean dry static stability is well represented by each reanalysis. The mean sonde specific humidity from 1998 is dryer (smaller) than that from the later campaigns, which is  $\sim 3.5 \text{ g kg}^{-1}$  near the surface in 1998, close to  $5 \text{ g kg}^{-1}$  in 2008, and  $\sim 4.5 \text{ g kg}^{-1}$  in 2015–16. It may be

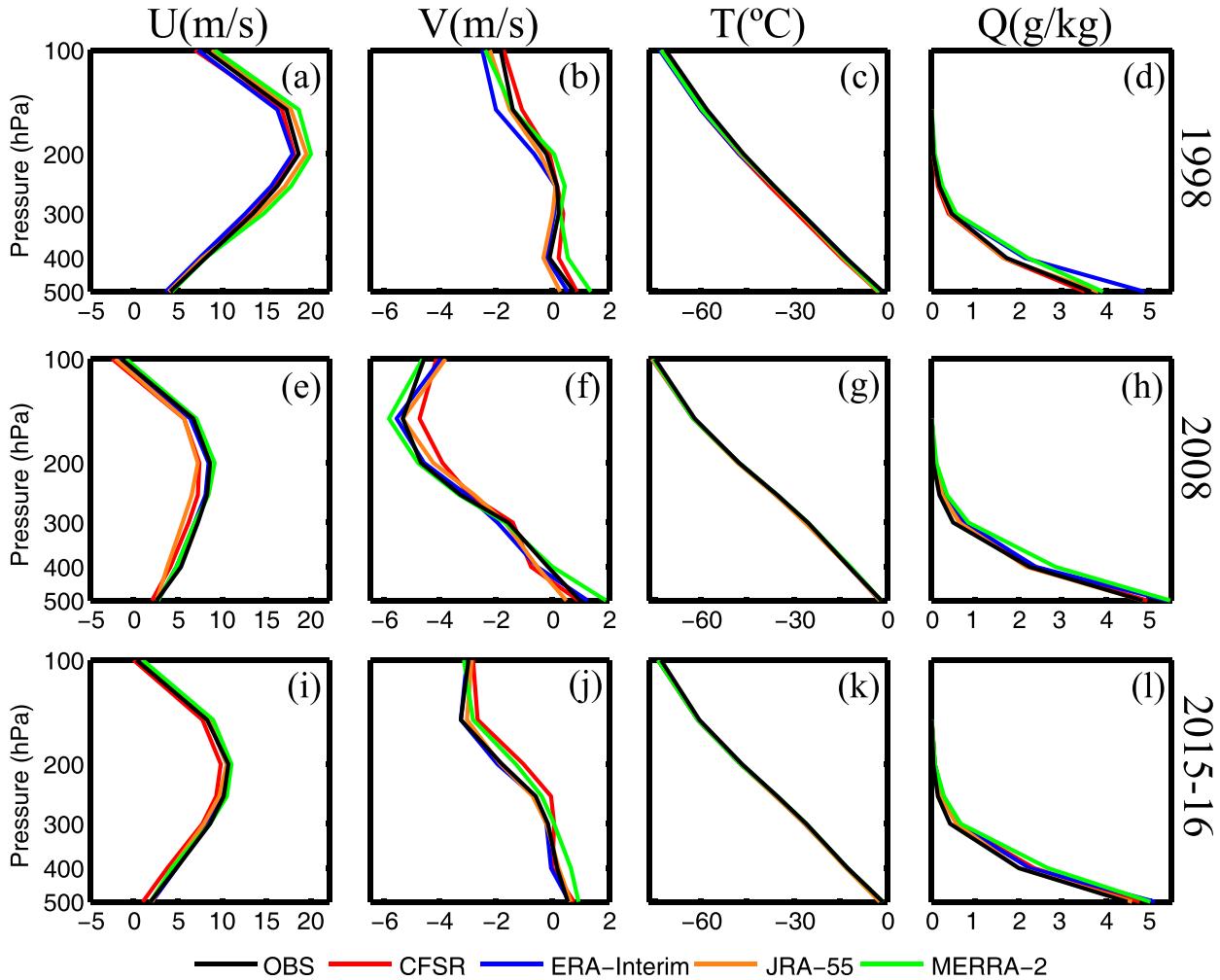


FIG. 2. Mean vertical profiles of  $U$  ( $\text{m s}^{-1}$ ),  $V$  ( $\text{m s}^{-1}$ ),  $T$  ( $^{\circ}\text{C}$ ), and  $Q$  ( $\text{g kg}^{-1}$ ) averaged over all independent soundings and the corresponding mean interpolated from the four reanalysis products during (top) 1998 TIPEX-II IOP, (middle) 2008 JICA/Tibet Project IOP, and (bottom) 2015–16 TIPEX-III IOPs.

related to the differences in the instrument precision, while RS80 have more dry bias than the later instruments RS92 and GTS1.

The mean natural variabilities of the temperature, specific humidity and winds measured in terms of standard deviation of all interpolated sounding profiles extracted from each reanalysis during each of the field experiment periods also well represent the observed sounding mean variabilities (Fig. 3). There are slightly stronger discrepancies by some reanalyses for horizontal winds variabilities at most levels (Figs. 3a,b,e,f,i,j) and also for temperature and specific humidity near the surface in this intercomparison (500 hPa) (Figs. 3c,d,g,h,k,l).

#### b. The mean biases and RMSE

For a more quantitative validation of each reanalysis, we calculated the mean bias and root-mean square error

(difference) between each interpolated reanalysis sounding and observations (R-O; Fig. 4).

#### 1) HORIZONTAL WIND

Although the mean horizontal wind profiles match well with sounding observations (Figs. 2a,b,e,f,i,j), non-negligible biases (R-O) do exist and vary greatly from reanalysis to reanalysis and for different periods (Figs. 4a,b,e,f,i,j). Both ERA-Interim and CFSR have a negative bias that increases with height in the westerly mean flow in 1998 ( $-0.6$  to  $-0.9 \text{ m s}^{-1}$  for ERA-Interim;  $-0.1$  to  $-1.2 \text{ m s}^{-1}$  for CFSR from 500 to 100 hPa). Such a bias is mostly diminished throughout all levels for the ERA-Interim reanalysis (around  $-0.5 \text{ m s}^{-1}$  in 2008,  $<\pm 0.2 \text{ m s}^{-1}$  in 2015–16) but enlarged below 300 hPa and reduced at the upper levels for CFSR with a similar bias amplitude of around  $-1 \text{ m s}^{-1}$  in the more recent

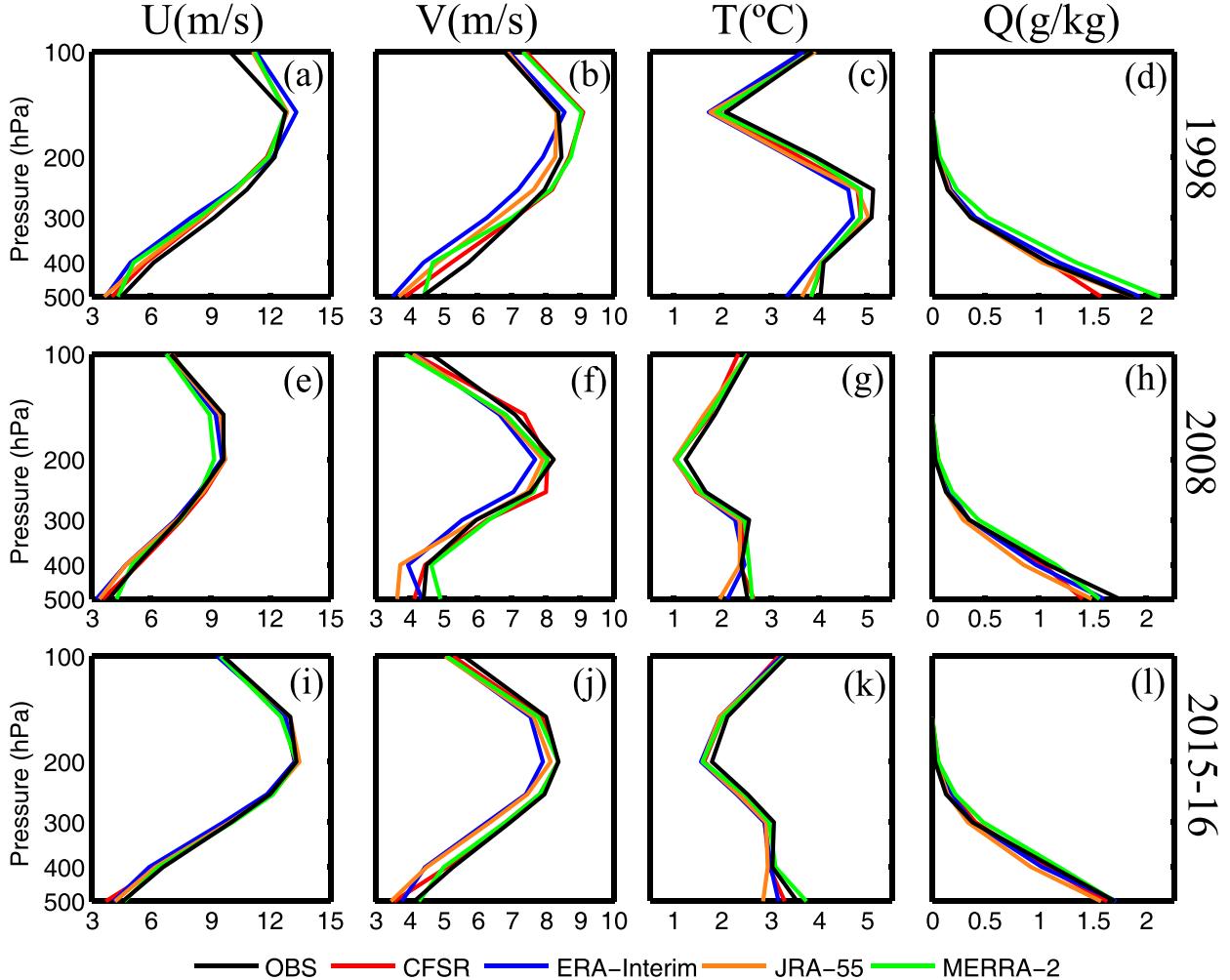


FIG. 3. Vertical profiles of standard deviation (STD) for  $U$  ( $\text{m s}^{-1}$ ),  $V$  ( $\text{m s}^{-1}$ ),  $T$  ( $^{\circ}\text{C}$ ), and  $Q$  ( $\text{g kg}^{-1}$ ) averaged over all independent soundings and the corresponding STDs for profiles interpolated from the four reanalysis products during (top) 1998 TIPEX-II IOP, (middle) 2008 JICA/Tibet Project IOP, and (bottom) 2015–16 TIPEX-III IOPS.

two field experiment periods (Figs. 4a,e,i). The JRA reanalysis has almost no zonal wind bias near the surface and a positive bias around  $0.7 \text{ m s}^{-1}$  above 300 hPa in 1998 (Fig. 4a), which becomes the largest (negative bias, from nearly zero at 500 hPa to a peak of  $-1.7 \text{ m s}^{-1}$  at 300 hPa, decreasing to  $-0.8 \text{ m s}^{-1}$  at 100 hPa) in 2008 (Fig. 4e) before returning to a much smaller bias ( $<\pm 0.6 \text{ m s}^{-1}$  with the negative value between 400 and 200 hPa and the positive value at 500 hPa as well as the upper levels) during the 2015–16 field experiment period (Fig. 4i). MERRA-2 has a markedly positive bias of zonal wind around  $1.5 \text{ m s}^{-1}$  above 400 hPa in 1998, and the negative bias at the lower levels and the positive bias at the upper levels, which is no larger than  $\pm 1 \text{ m s}^{-1}$  throughout two later field experiment periods (Figs. 4a,e,i). All reanalysis products have the mean biases in the meridional wind at different levels (mostly within  $\pm 1 \text{ m s}^{-1}$ ;

Figs. 4b,f,j) considerably smaller than the averaged RMSEs ( $3\text{--}6 \text{ m s}^{-1}$ ). This may be due to compensating biases among stations. Therefore, given the large variability of RMSEs at different levels, and the rather weak meridional flow, if we were to use these reanalyses to assess the mean climatological changes in the regional Hadley circulation, these seemingly small biases, and the variations from year to year may become nonnegligible as well.

The obvious variance of zonal and meridional wind among these reanalyses, which also changes during the time span of three field experiments, also are presented through RMSE (R-O) (Figs. 4a,b,e,f,i,j). The RMSEs of the meridional wind show much greater difference. CFSR has the largest RMSE with the peak up to  $5.5 \text{ m s}^{-1}$  at 250 hPa during 1998 IOP and around  $4.5 \text{ m s}^{-1}$  at 150 hPa during two later IOPs, while ERA-Interim and JRA-55 have relative smaller RMSEs with little variation

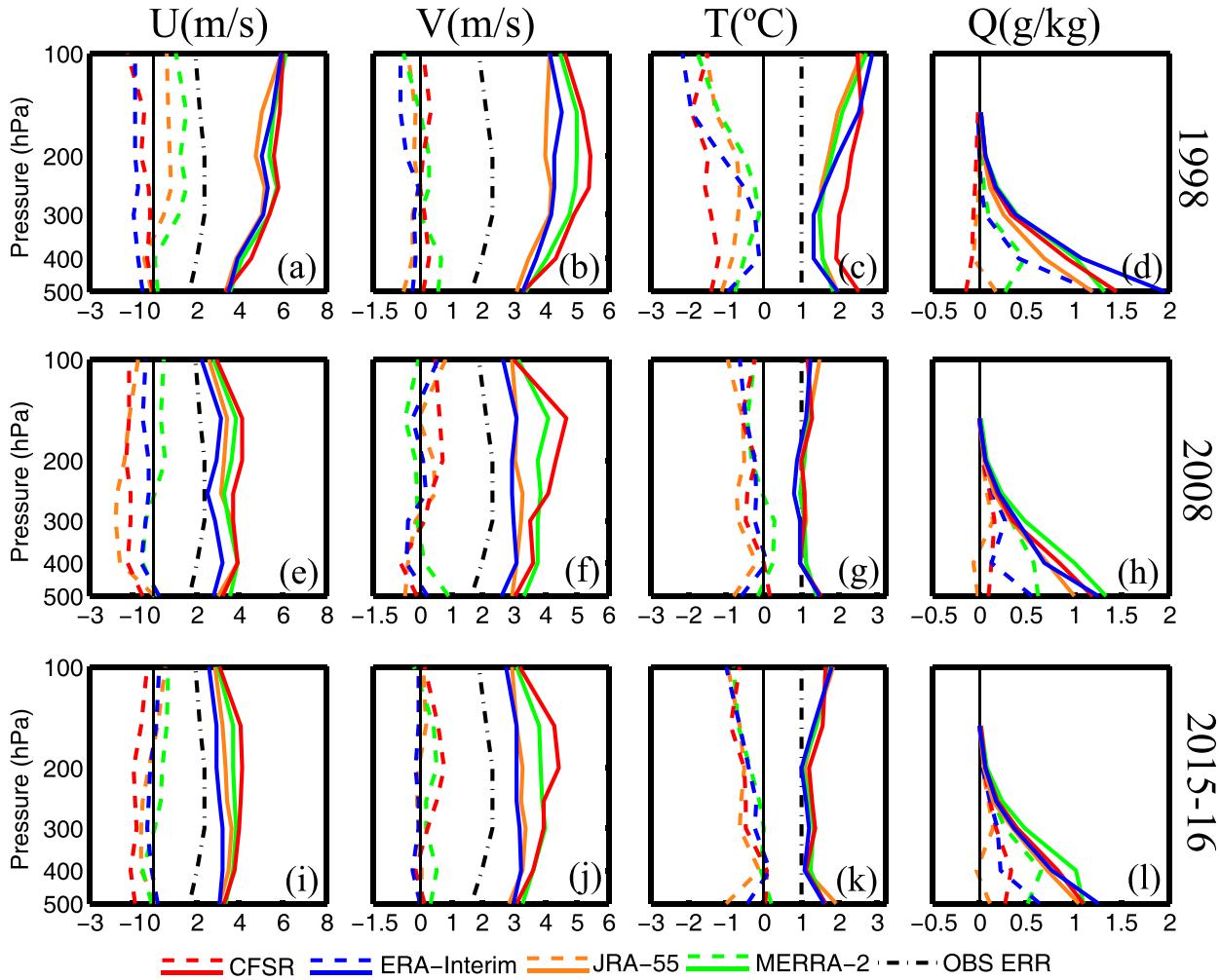


FIG. 4. Vertical profiles of the mean biases (dashed) and RMSEs (solid) of  $U$ ,  $V$ ,  $T$ , and  $Q$  averaged over all independent soundings for each reanalysis verifying against the observations (R-O) during (top) 1998 TIPEX-II IOP, (middle) 2008 JICA/Tibet Project IOP, and (bottom) 2015–16 TIPEX-III IOPs.

throughout the vertical column that are around  $4 \text{ m s}^{-1}$  during 1998 IOP reducing to  $3 \text{ m s}^{-1}$  during 2008 and 2015–16 IOPs. Overall, among the four reanalysis products, the CFSR reanalysis has the largest RMSEs for winds over all field periods while the ERA-Interim reanalysis (except for 1998) has the smallest uncertainty validating against the field radiosonde observations, which are consistent with previous validation studies (Bao and Zhang 2013; Simmons and Poli 2015; Manney et al. 2017). Note that the JRA reanalysis has the smallest RMSE for winds in 1998 (Figs. 4a,b) due to the assimilation of the validating soundings, and thus this is more of a measure of the fit of the reanalysis (and the respective data assimilation system) to the sounding observations rather than a measure of accuracy of the reanalysis. Furthermore, all reanalyses have greater RMSE in 1998 and less RMSE in 2008 and 2015–16, this improvement

should be because the transition from TOVS to ATOVS observations (Long et al. 2017).

## 2) TEMPERATURE

All reanalysis products have a negative (cold) bias comparing to sounding observations throughout the vertical temperature profiles over all three field periods, except for MERRA-2 between 400 and 300 hPa in 2008 (at 500 hPa in 2015–16) where the positive (warm) bias is around  $0.3^{\circ}\text{C}$  ( $0.2^{\circ}\text{C}$ ). The biases also vary greatly from reanalysis to reanalysis and for different periods (Figs. 4c,g,k). During the 1998 field period (Fig. 4c), the largest such bias is seen for CFSR, which has a negative bias of around  $-1.5^{\circ}\text{C}$  throughout the vertical column, whereas MERRA-2 has the smallest negative overall bias in 1998, which is close to zero around 300 hPa but increases quickly with height with the maximum of  $-1.7^{\circ}\text{C}$

at 100 hPa. Similar with MERRA-2, the bias in ERA-Interim temperature is almost negligible around 400 hPa but becomes the largest among the four reanalyses with a value of  $-2^{\circ}\text{C}$  at 100 hPa. These temperature biases in CFSR, MERRA-2, and ERA-Interim during the 2008 and 2015–16 field experiment periods remain or diminish to near negligible below 300 hPa but are greatly reduced to around  $-0.5^{\circ}\text{C}$  above 300 hPa (Figs. 4g,k). The JRA reanalyses have the most persistent negative biases across the vertical column for all periods, though with some noticeable reduction in more recent years.

The temperature RMSEs in four reanalyses have great differences in 1998 (Fig. 4c). The smallest overall RMSE is for JRA-55 and MERRA-2, both of which are very close to each other (from  $\sim 1.9^{\circ}\text{C}$  at 500 hPa to a minimum of  $\sim 1.5^{\circ}\text{C}$  around 250 hPa, and then increase to a maximum  $\sim 2.5^{\circ}\text{C}$  at 100 hPa); clearly the largest is for CFSR (from  $\sim 2.4^{\circ}\text{C}$  at 500 hPa to a peak of  $\sim 2.6^{\circ}\text{C}$  at 150 hPa). ERA-Interim has the smallest RMSE of temperature below 250 hPa but becomes the largest at 100 hPa ( $\sim 2.8^{\circ}\text{C}$ ). The temperature RMSEs among four reanalyses are found to agree very closely with each other in 2008 and 2015–16, and they decrease by one-half to one-third from 1998, especially in the upper levels (Figs. 4g,k). ERA-Interim and CFSR have the smallest and largest RMSE in temperature, respectively, but with minor difference in the more recent two field experiment periods.

In general, these temperature biases and RMSEs for all of four reanalyses largely diminished from 1998 to 2008 and 2015–16 coincident with the transition from TOVS to the ATOVS observation systems, as well as the introduction of other satellite radiance observations such as GPS-RO, AIRS, and IASI. In addition, the improvements in radiosonde instrument precisions may have also contributed to reduce the biases between reanalysis products and observations. Meanwhile, the variances of temperature biases as well as RMSEs among four reanalyses show the greatest disagreement occurs in 1998 and agreement improves in 2008 and 2015–16. These reanalyses agree more closely with each other after 1998 because there are fewer issues assimilating more quantity and higher quality observations (Bosilovich et al. 2017; Long et al. 2017).

### 3) SPECIFIC HUMIDITY

The biases in moisture content also have large differences in comparison to the mean profile for all reanalysis products and for different field campaign periods (Figs. 4d,h,l). The JRA-55 reanalysis has the overall smallest bias in specific humidity for all field periods. The bias in JRA is  $\sim 0.2 \text{ g kg}^{-1}$  near the surface and is close to zero above 500 hPa in 1998, while it is near zero below 300 hPa and smaller than  $0.2 \text{ g kg}^{-1}$  at the

upper levels in 2008 and 2015–16. The mean biases of specific humidity in CFSR vary from dry (negative) in 1998 to moist (positive) in 2008 and 2015–16. MERRA-2 are moist (positively) biased with minor changes among the three campaigns. ERA-Interim shows a noticeably bigger mean bias than the other three reanalyses at the lower level in 1998, while the positive maximum near the surface drastically decreases from  $\sim 1.2 \text{ g kg}^{-1}$  in 1998 to  $\sim 0.6 \text{ g kg}^{-1}$  in 2008 and 2015–16.

The specific humidity RMSEs from four reanalyses all decrease with height and have different degrees of reduction from 1998 to 2008 and 2015–16. Similar to temperature, the profiles of specific humidity RMSEs also present great difference among four reanalyses in 1998, and become much closer with each other in 2008 and especially 2015–16, which coincides with reduced differences in humidity amongst the reanalyses. ERA-Interim shows the biggest changes in humidity RMSE across the three campaigns, which decreases by one-third to one-half in later campaign periods from that in 1998. In 1998, ERA-Interim has the largest humidity RMSE with the maximum up to  $\sim 2 \text{ g kg}^{-1}$  near the surface, which is nearly twice that of JRA-55. Generally, JRA-55 has the smallest RMSEs in specific humidity with the smallest changes from campaign to campaign. Note that JRA-55 has much smaller RMSE and bias than other reanalyses in 1998, likely because the field campaign soundings were made available and assimilated into only the JRA reanalysis for that year's reanalysis.

Figure 5 presents mean biases and RMSEs for each of the reanalyses averaged over the three stations common to all three IOPs during three campaign periods, respectively. The most obvious differences between Figs. 4 and 5 are in the horizontal winds in 1998. To be specific, in Fig. 5, the RMSEs of zonal wind from all reanalyses at 100 hPa are larger than that in Fig. 4, and the mean meridional winds from CFSR and MERRA-2 during the 1998 campaign period have obvious positive biases ( $\sim 1 \text{ m s}^{-1}$ ) at almost all altitudes, whereas the smaller biases are limited in  $\pm 0.5 \text{ m s}^{-1}$  in Fig. 4. These imply the horizontal wind from reanalyses over these three stations in 1998 has bigger biases from the mean wind averaged over all stations. Nevertheless, the mean biases and RMSEs of horizontal wind in 2008 and 2015–16, temperature, and specific humidity in all three campaign periods averaged over these three stations are overall similar to those averaged over all stations, which indicates that the analyses are broadly consistent across this region, not just a few stations.

#### c. Variations of uncertainty and bias across the three campaigns

Figure 6 summarizes the campaign-to-campaign changes in bias spreads for each of the reanalyses. The results

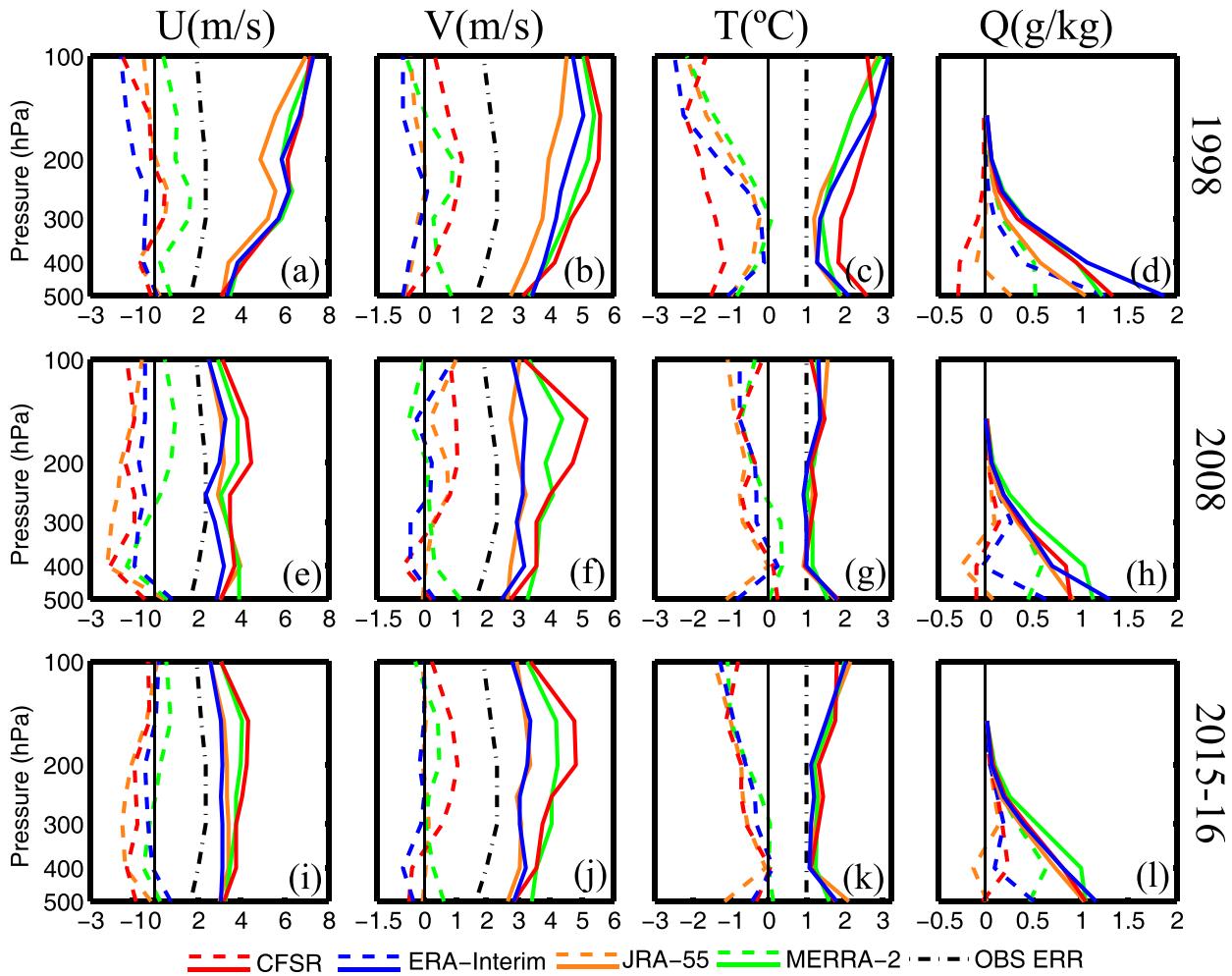


FIG. 5. Vertical profiles of the mean biases (dashed) and RMSEs (solid) of  $U$ ,  $V$ ,  $T$ , and  $Q$  averaged over the three stations common to all three IOPs (Gaize, Dingri, Linzhi) for each reanalysis verifying against the observations (R-O) during (top) 1998 TIPEX-II IOP, (middle) 2008 JICA/Tibet Project IOP, and (bottom) 2015–16 TIPEX-III IOPs.

show evidence of nonnegligible reduction in bias spread across the three campaigns. The zonal wind biases for each reanalysis significantly decrease from 1998 to 2008 and 2015–16. The reductions are more prominent in the upper troposphere; the bias span of zonal wind for all reanalyses in two later campaigns is cut to almost half of that in 1998 (Figs. 6a–d). The meridional wind biases present a relatively smaller trend (Figs. 6e–h). For example, the bias spread of  $V$  wind for CFSR at 150 hPa has little difference between 1998 and 2008 and becomes slightly smaller in 2015–16. Besides, JRA-55 shows the least notable variations in the bias spreads of meridional wind from campaign to campaign. The differences of horizontal winds among reanalysis products can also be found in these box-and-whisker plots. These differences are significant during the 1998 IOP but less evident during 2008 and 2015–16 IOPs. Overall, CFSR has

the largest bias variability of zonal and meridional wind during the three campaign IOPs, while ERA-Interim has the smallest bias variability in winds except for 1998. During the 1998 IOP, JRA-55, which assimilates the field campaign soundings, tends to have a slightly smaller bias variability in zonal wind than ERA-Interim and obviously has the smallest bias spread of meridional wind.

For temperature (Figs. 6i–l), the biases for these reanalyses mostly are negative at almost all altitude ranges with larger values in 1998 compared to 2008 and 2015–16. CFSR is the coldest compared with observations: 75%–90% temperature biases of CFSR are less than zero. By comparison, these temperature biases are close to a normal distribution with a smaller span during the two later campaigns. The bias spreads of temperature for the four reanalyses decrease from 1998 to 2008 and

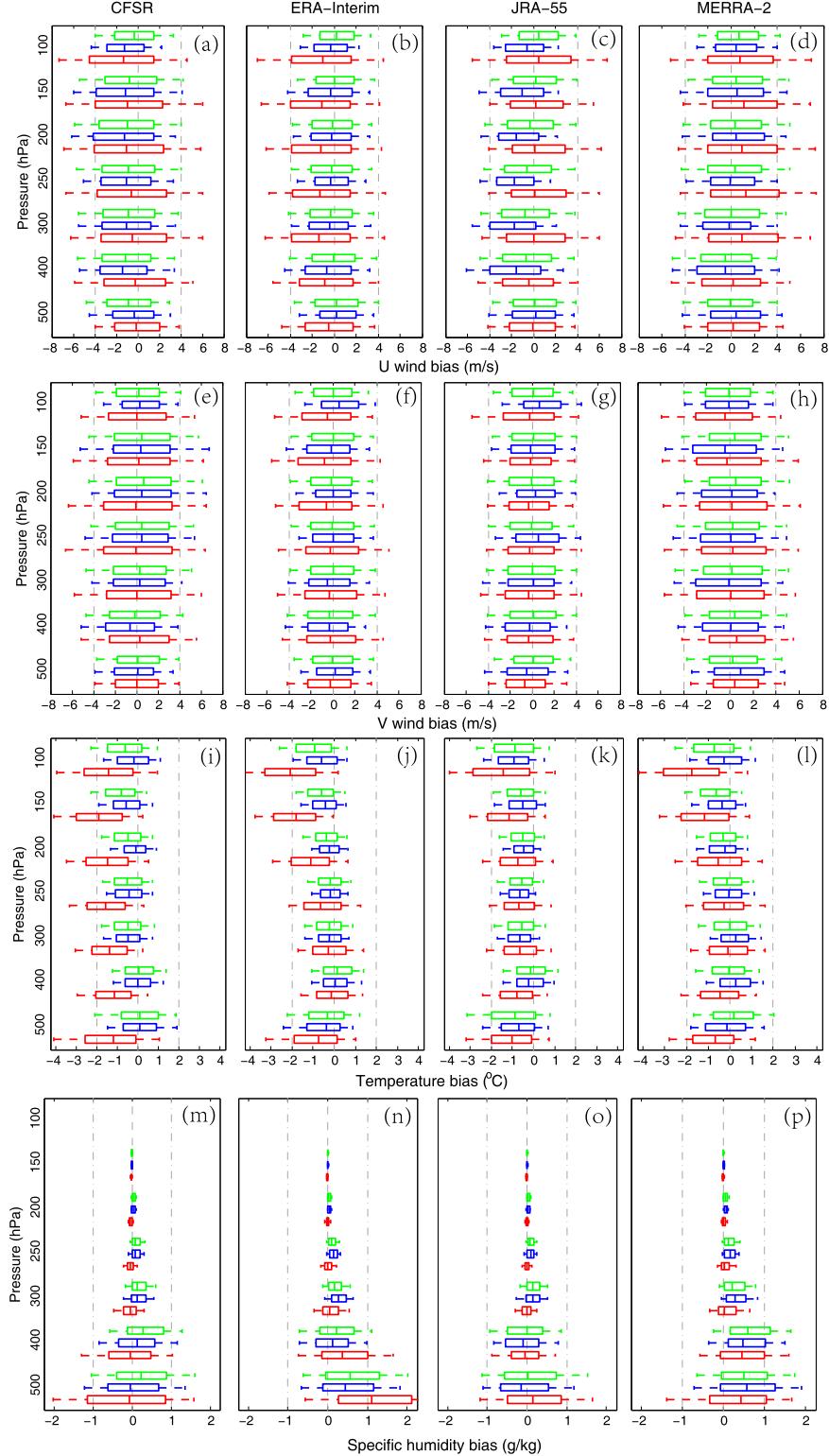


FIG. 6. Box-and-whisker plots of  $U$ ,  $V$ ,  $T$ , and  $Q$  bias for all independent soundings for each reanalysis verifying against the observations (R-O) during 1998 TIPEX-II IOP (red), 2008 JICA/Tibet Project IOP (blue), and 2015–16 TIPEX-III IOPs (green). The central mark is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the 10th and 90th percentiles.

2015–16, which is most pronounced at upper levels. The specific humidity biases also show a relatively small decreasing trend for each of reanalyses from campaign to campaign (Figs. 6m–p).

It is evident for winds, temperature, and specific humidity that the analysis uncertainty in terms of bias spread is larger (especially in the upper levels) in the earlier experiment period (1998) for each reanalysis product than in the later two field experiments (2008 and 2015–16). This suggests the transition from TOVS to ATOVS and the introductions of other satellite observations (AIRS, IASI, and GPS RO, etc.) may have contributed significantly to the improved reanalysis quality in more recent campaigns given the same data assimilation system is used for each reanalysis across different field campaigns. The differences between the last two field campaign periods, on the other hand, are rather small for the reanalysis quality of these variables.

#### *d. Diurnal variation of the mean bias*

In our earlier work based on the high-frequency 1998 TIPEX-II IOP soundings available multiple times daily (Bao and Zhang 2013), we found that there are strong diurnal variations in both RMSEs and mean biases of some variables by the newer-generation reanalyses (CFSR and ERA-Interim) and their predecessors [NCEP–NCAR Reanalysis Project (NNRP) and ERA-40]. In this study, these intense soundings collected during 1998 TIPEX-II IOP complemented by the two more recent field experiments (2008 JICA/Tibet Project IOP and 2015–16 TIPEX-III IOPs) offer us a rare opportunity to estimate the characteristics and changes in diurnal variations of the mean biases by different reanalysis products across three field campaign periods in the most recent two decades. It is worth noting that, unlike the two earlier field experiments in 1998 and 2008, the sounding observations during 2015–16 TIPEX-III IOPs were implemented at three times a day (0000, 0600, and 1200 UTC) instead of four times a day. Therefore, merely the variations of the mean biases at these three times are analyzed. It is clear from Figs. 7–10 that there are strong diurnal variations in the mean biases of almost reanalysis variables by four reanalyses across three field experiments. The degree of diurnal variation for each reanalysis also has a considerable difference among different IOPs.

#### 1) DIURNAL VARIATION OF THE MEAN HORIZONTAL WIND BIAS

For CFSR and MERRA-2, the degree of diurnal variations of the mean-*U*-wind bias weakens from campaign to campaign: the strongest variations occur during 1998 TIPEX-II IOP with significant peaks at

upper levels, while relatively weaker variations appear during the two later IOPs. For JRA-55, the diurnal mean *U*-wind bias has a positive peak (about  $1.7 \text{ m s}^{-1}$ ) between 250 and 150 hPa at 1800 UTC during the 1998 TIPEX-II IOP, and a slightly larger negative peak (about  $-2.2 \text{ m s}^{-1}$ ) between 400 and 300 hPa at 1200–1800 UTC during the JICA/Tibet IOP, as well as the weakest diurnal variations during the 2015–16 TIPEX-III IOP. ERA-Interim has the most indistinctive diurnal variations of the mean *U*-wind bias among the four analyses, and the smallest changes at diurnal variation amplitude from campaign to campaign (Fig. 7).

Unlike the mean *U*-wind bias, the strongest diurnal variations of the mean *V*-wind biases for these four reanalyses mostly appear during 2008 JICA/Tibet Project IOP except for MERRA-2, which has the strongest diurnal variations during 1998 TIPEX-II IOP. This is most pronounced at upper levels and might relate to the station distributions during 2008 JICA/Tibet, which are more closely aligned with the ridgeline of the upper tropospheric monsoon anticyclone (Nützel et al. 2016). Additionally, the weakest diurnal variations for each reanalysis occur during the 2015–16 TIPEX-III IOPs but with little difference from that for CFSR, ERA-Interim, and JRA-55 from that during 1998. ERA-Interim in general has the smallest changes on diurnal variations of mean *V*-wind biases across the three field campaign periods (Fig. 8).

#### 2) DIURNAL VARIATION OF THE MEAN TEMPERATURE BIAS AND THE MEAN SPECIFIC HUMIDITY BIAS

On the whole, all the four reanalyses tend to give the notable cold (negative) bias at each time throughout the vertical column during 1998 TIPEX-II IOP. The diurnal variation ranges of the mean temperatures for these reanalyses drastically reduce with some weak warm (positive) biases occurred at the low level (except for JRA-55) during the two later campaign periods. This reducing trend is more pronounced, in particular for ERA-Interim: the cold biases are far greater than  $-2^\circ\text{C}$  at the upper levels between 0600 and 1800 UTC during 1998 TIPEX-II, and they are reduced by one-third to two-thirds during 2008 JICA/Tibet and 2015–16 TIPEX-III. In addition, the diurnal variations of the cold bias for JRA-55 have little difference between three campaigns IOPs (Fig. 9).

Because specific humidity rapidly decreases with increase in height, the obvious diurnal variations of the mean specific humidity occur below 300 hPa. ERA-Interim and MERRA-2 present overall wetter bias during all three IOPs, but the wet peak of the mean specific humidity bias for ERA-Interim at 0600–1200 UTC near the surface decreases dramatically from

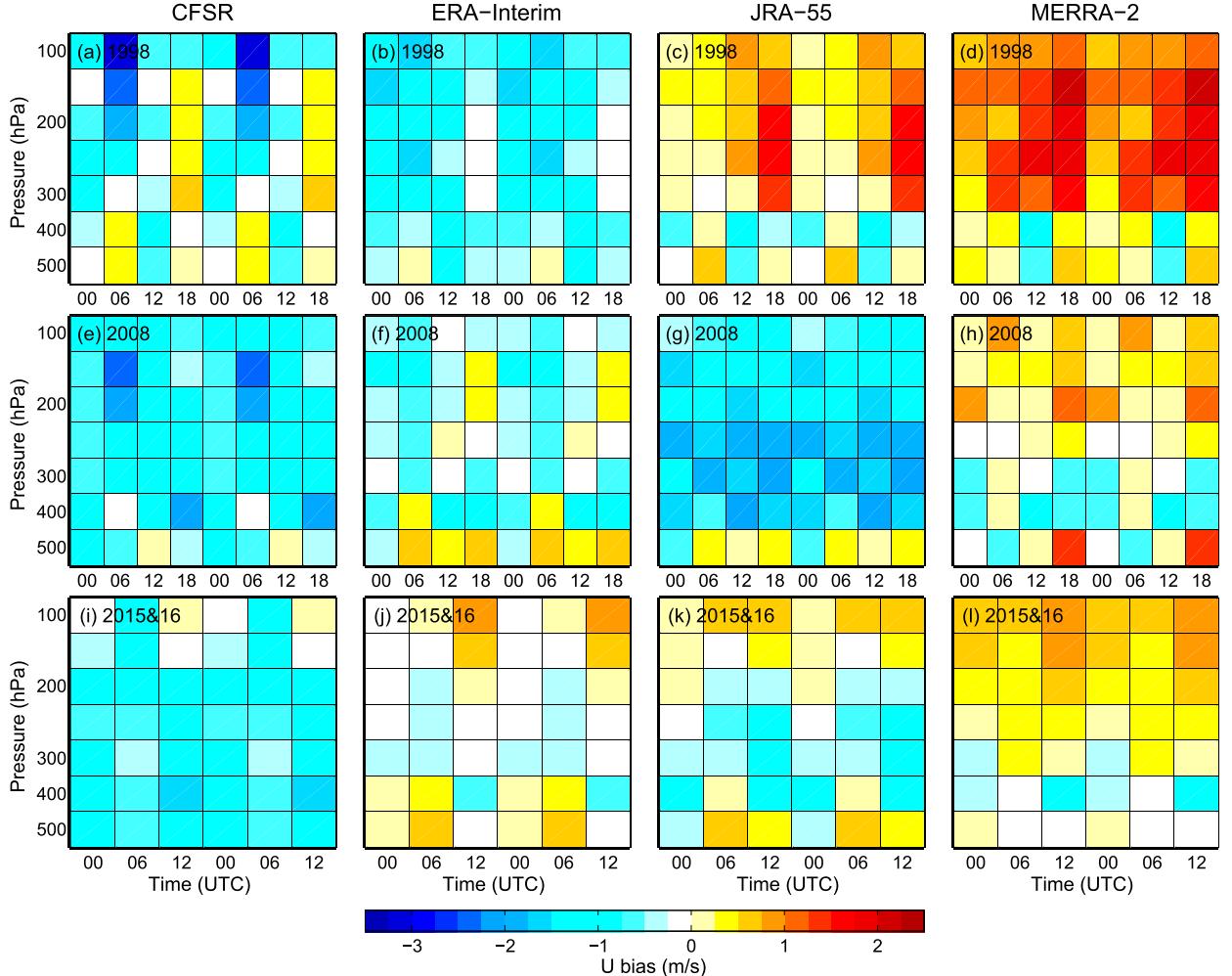


FIG. 7. The diurnal variations of the mean biases of  $U$  ( $\text{m s}^{-1}$ ) for CFSR, ERA-Interim, JRA-55, and MERRA-2 at different pressure levels validated against the soundings from (top) 1998 TIPEX-II IOP, (middle) 2008 JICA/Tibet Project IOP, and (bottom) 2015–16 TIPEX-III IOPs.

values larger than  $1.5 \text{ g kg}^{-1}$  during 1998 IOP to no more than  $1 \text{ g kg}^{-1}$  during 2008 and 2015–16 IOPs. For CFSR and JRA-55, the mean specific humidity biases generally present the wet peaks at 0600–1200 UTC and the dry peaks at 1800–0000 UTC below 400 hPa with the small reduction of diurnal variation range from campaign to campaign (Fig. 10).

To sum up, the mean biases of some available variables for the reanalyses present varying degrees of diurnal variations during the three campaign periods. These biases might affect the ability and quality of using the reanalyses to examine key aspects of the diurnal cycle (e.g., static stability, surface fluxes) in other applications where they might be used. The diurnal variation amplitudes in each variable bias for the four analyses show the overall decrease tendencies with different degrees from campaign to campaign. ERA-Interim has the weakest

diurnal variations in the mean horizontal wind biases across three campaign IOPs, but the strongest diurnal variations in the upper-level temperature and near-surface specific humidity during 1998 TIPEX-II IOP. JRA-55, which assimilates TIPEX-II intensive sounding observations, tends to give the relative weaker diurnal variations in each variable's bias during 1998 IOP; it also has small differences on the diurnal variation ranges of all variables across three IOPs.

There are several related reasons why these reanalyses might have higher quality and accuracy in 2008 and 2015–16 (ATOVS period) than in 1998 (TOVS period). TOVS consists of three instruments: the High Resolution Infrared Radiation Sounder (HIRS), which is an infrared temperature sounder and has a little impact on tropospheric forecast skill (Gelaro and Zhu 2009); the Stratospheric Sounding Unit (SSU), which

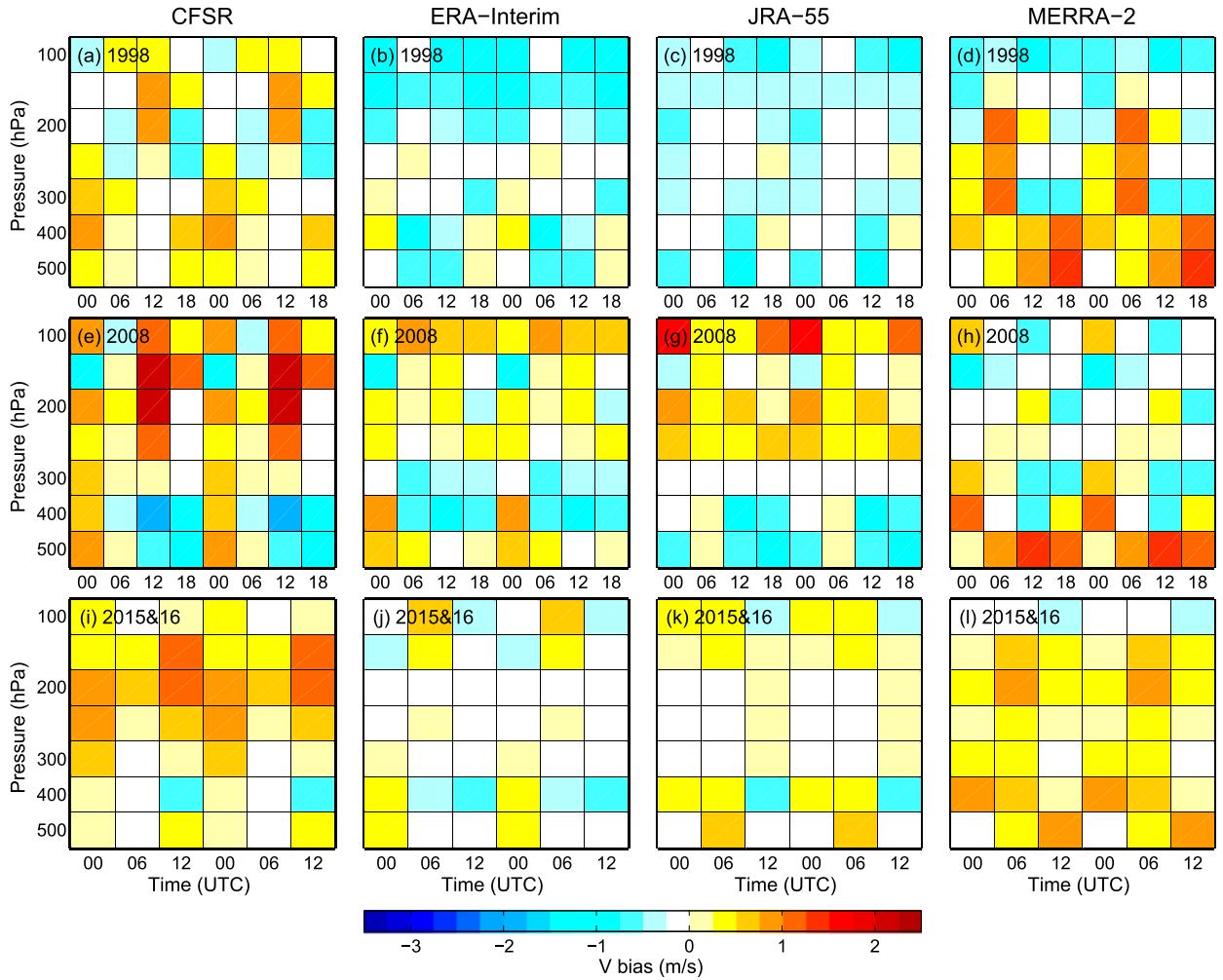


FIG. 8. The diurnal variations of the mean biases of  $V$  ( $\text{m s}^{-1}$ ) for CFSR, ERA-Interim, JRA-55, and MERRA-2 at different pressure levels validated against the soundings from (top) 1998 TIPEX-II IOP, (middle) 2008 JICA/Tibet Project IOP, and (bottom) 2015–16 TIPEX-III IOPs.

monitors the stratosphere thermal structure; and the Microwave Sounding Unit (MSU), which has four channels with deep vertical weighting functions and was the principal instrument measuring the atmospheric temperature spanning the surface through the stratosphere during 1978–98. In May 1998, ATOVS was flown on the *NOAA-15* satellite and is made up of three instruments as well: the next generation of HIRS, AMSU-A, and AMSU-B. Compared with the TOVS system, ATOVS provided a significant improvement for humidity and vertical resolution of temperature, particularly in cloudy areas (English et al. 2000). AMSU-A instrument combines the role of MSU and SSU, has 15 channels with finer vertical resolution, which is designed to retrieve the temperature profile from 3 hPa (45 km) to the surface. AMSU-B is a new five-channel microwave humidity sounder that contributes sounding information

on the water vapor profile in the troposphere and lower stratosphere (below about 10 km). AMSU-B radiances that observe moisture explicitly have beneficial impact on moisture processes and can provide water vapor products to study global or local precipitation and humidity (Bennartz et al. 2002; Brogniez and Pierrehumbert 2007).

Furthermore, the huge and various additional satellite radiances are assimilated in several reanalysis systems from 1999 to the present that have different degrees of influences on the quality of these reanalyses (Fig. 1b). For example, hyperspectral satellite instruments such as AIRS and IASI, both of which provide even near-real-time three-dimensional monitoring information on air and surface temperature, water vapor, greenhouse gases, and cloud properties in thousands of channels, add tremendous numbers of observations into several reanalysis systems from 2002 to the present (McCarty et al. 2016).

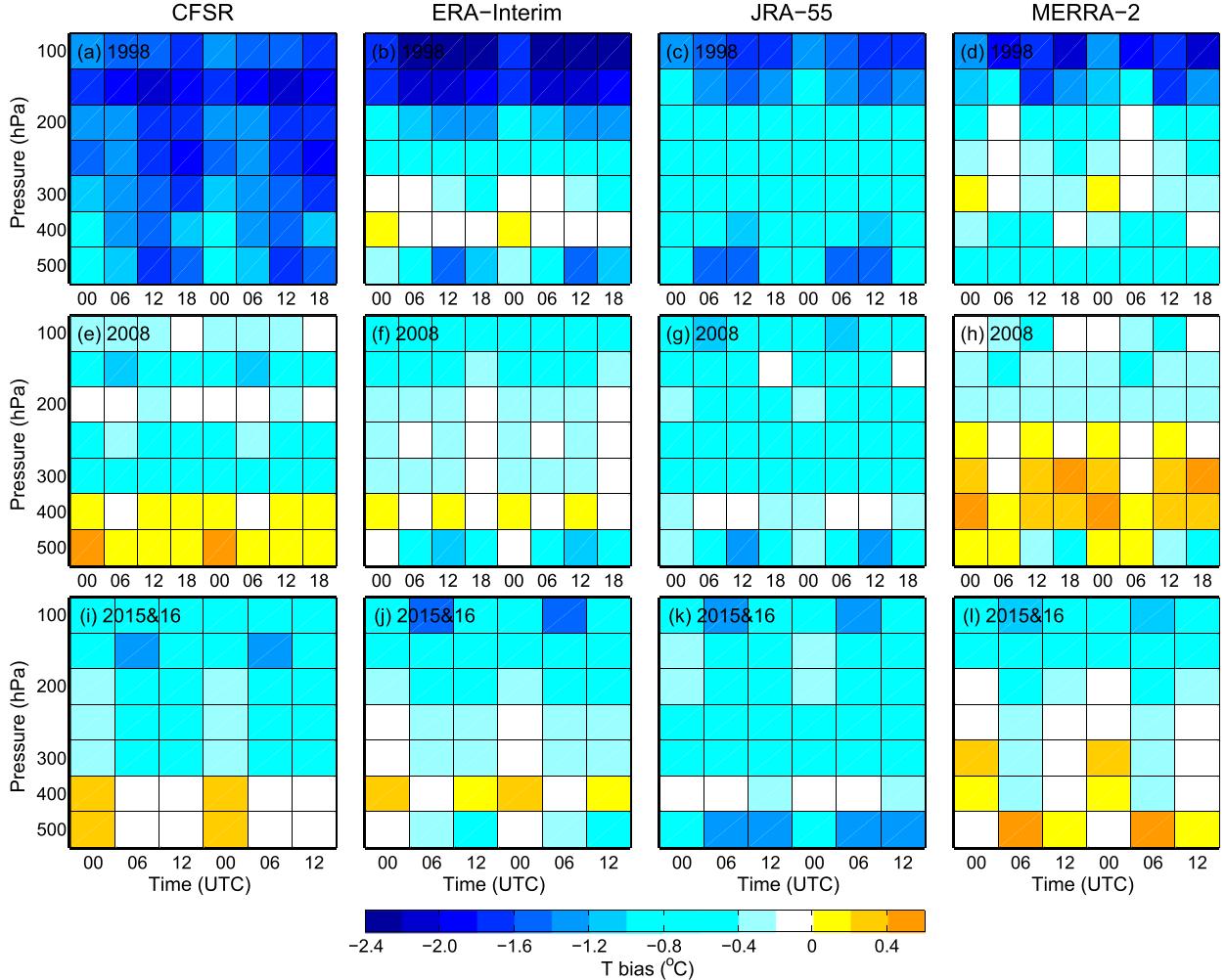


FIG. 9. The diurnal variations of the mean biases of  $T$  ( $^{\circ}$ C) for CFSR, ERA-Interim, JRA-55, and MERRA-2 at different pressure levels validated against the soundings from (top) 1998 TIPEX-II IOP, (middle) 2008 JICA/Tibet Project IOP, and (bottom) 2015–16 TIPEX-III IOPs.

AIRS have the second-largest contributions on improving tropospheric forecast skill after AMSU-A among various satellite systems (Gelaro and Zhu 2009). IASI also has a positive influence on the assimilation system (Collard and McNally 2009; Guidard et al. 2011). The other notable satellite observations are from GPS RO, which can obtain high precision and accuracy vertical profiles of the bending angle of radio wave trajectories and atmospheric refractivity in the neutral atmosphere (below the ionosphere) by receiving the radio signals from the GPS satellites (Anthes et al. 2008). Because the angle of refraction depends on pressure, temperature, and the amount of water vapor in the atmosphere, obvious reduction of the temperature bias and humidity bias can be found with the introduction of GPS RO observations in different global data assimilation or reanalysis systems (Poli et al. 2010; Cucurull et al. 2014).

It is worth noting that there are considerably large disagreements among reanalysis products, one of primary reasons is the sources of the input observations are not quite the same among these reanalysis systems (Fig. 1b). For example, the 1998 TIPEX-II sounding observations are assimilated into only JRA-55, AIRS is not assimilated in JRA-55, and IASI only is assimilated in CFSR and MERRA-2. Such differences can be found in numerous observation data not mentioned here. The other main reason may be the information assimilated from a given sensor differs from reanalysis to reanalysis (Fujiwara et al. 2017). Take AMSU-A, for instance, ERA-Interim assimilates its channels 5–14, CFSR assimilates channels 1–13 and channel 15, while JRA-55 and MERRA-2 assimilate channels 4–14. Such differences are not unique to AMSU-A (Table 1). In addition, the selection rules for excluding the same data can also

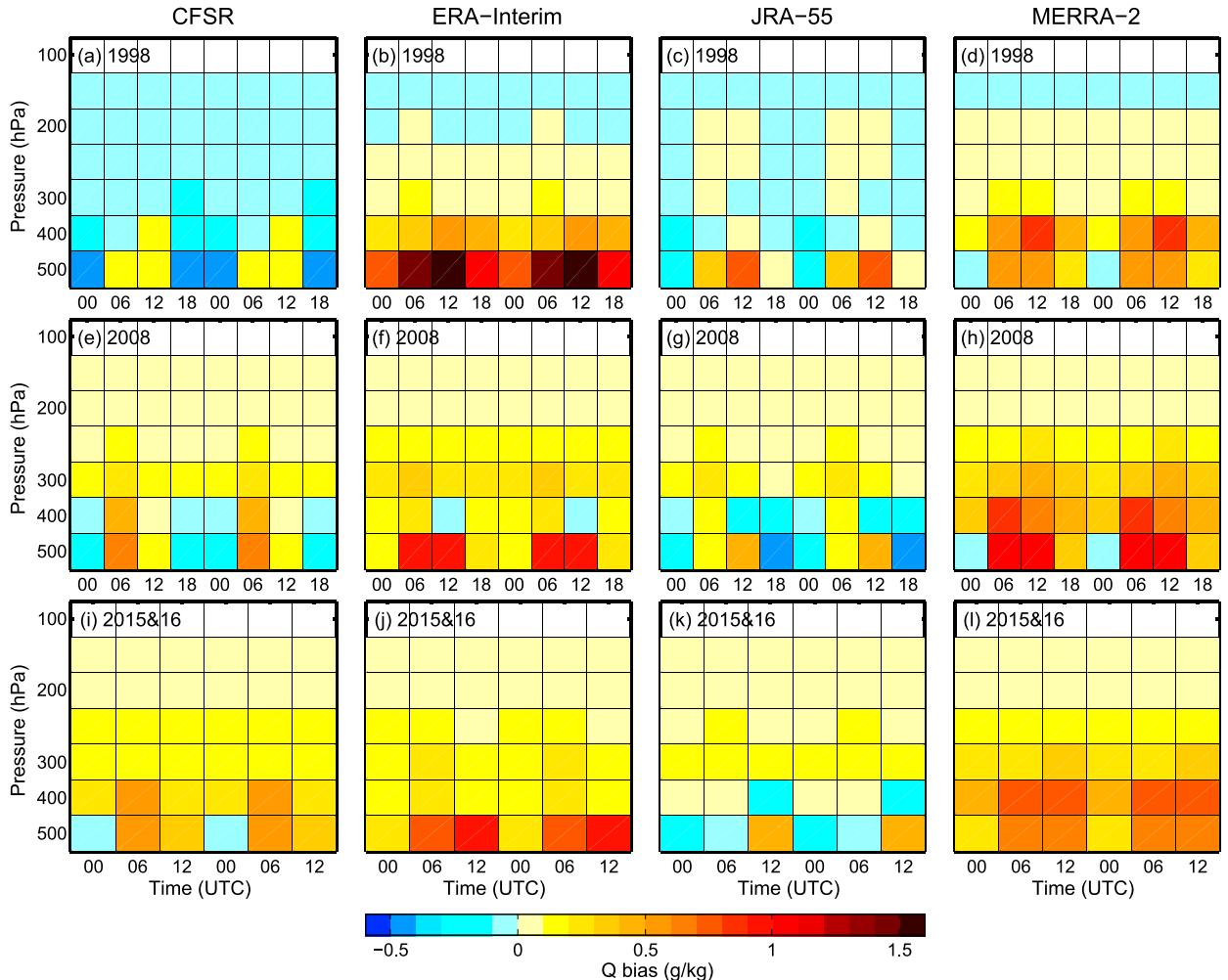


FIG. 10. The diurnal variations of the mean biases of  $Q$  ( $\text{g kg}^{-1}$ ) for CFSR, ERA-Interim, JRA-55, and MERRA-2 at different pressure levels validated against the soundings from (top) 1998 TIPEX-II IOP, (middle) 2008 JICA/Tibet Project IOP, and (bottom) 2015–16 TIPEX-III IOPs.

be different among these reanalyses. For example, in ERA-Interim, AMSU-A observations over high terrains for channels 5 and 6, or in rainy conditions, are rejected. JRA-55 excludes observations over land or sea ice for channels 4 and 5, over high terrains for channels 6 and 7, and in rainy conditions for channels 4–8. While only channels that peak above the surface are assimilated in MERRA-2, CFSR has a more complex blacklist for the AMSU-A data. The other well-known reason is the forecast models and data assimilation methodologies used in these analyses vary from each other.

#### 4. Concluding remarks

This study has two main objectives. One is to evaluate four leading modern atmospheric reanalysis products, namely, CFSR/CFSv2 produced by NCEP, ERA-Interim produced by ECMWF, JRA-55 produced by JMA, and

MERRA-2 produced by NASA, through validating against thousands of radiosonde observations from three major Tibetan Plateau experiments that were conducted respectively during the warm seasons (May–August) of 1998, 2008, and 2015–16 over three different decades. The other is to discuss how the satellite observing system changes might have influenced the quality of these reanalyses in the troposphere over the data-sparse Tibetan Plateau region. This large number of independent field campaign soundings offers us a rare opportunity to assess the quality of several state-of-the-art modern atmospheric reanalysis products. In particular, given these modern reanalysis products are widely used for understanding and detecting climate changes, it is imperative to assess whether they are adequately accurate for detecting regional climate trends, especially over data-sparse regions such as the Tibetan Plateau.

TABLE 1. The channels of microwave sounder observations used in four reanalyses.

Instrument	CFSR/CFSv2	ERA-Interim	JRA-55	MERRA-2
MSU	1-4	2-4	2-4	2-4
AMSU-A	1-13,15	5-14	4-14	4-14
AMSU-B	1-5	3-5	3-5	1-5

It is found that almost all reanalysis products can reproduce reasonably well the overall mean temperature, specific humidity, and horizontal wind profiles against the verifying independent sounding observations. However, there are nonnegligible mean biases that vary from reanalysis to reanalysis and from campaign to campaign that can be potentially comparable to or even bigger than the analysis-simulated mean regional climate trends in the study region (Xie et al. 2010; Ji et al. 2014). Large, diurnally and vertically varying systematic biases exist in the mean profiles of specific humidity and temperature in all reanalysis datasets, which suggests that extreme caution must be taken in using the reanalyses to assess regional climate changes in terms of atmospheric moisture and temperature. There are considerable differences in  $V$  wind and specific humidity among the reanalysis products, as well as relatively small disagreement in temperature and  $U$  wind. The mean biases and uncertainties of almost all reanalyses are reduced from 1998 IOP to the two later IOPs, especially in the upper levels for horizontal wind and temperature and near the surface for specific humidity. The RMSE profiles of temperature for the reanalyses are very close to each other during 2008 and 2015–16 IOPs. The variations of bias spreads in all variables for each reanalysis show reduced trend from 1998 IOP to 2008 and 2015–16 IOPs. Furthermore, the differences among these reanalysis products become smaller during the 2008 and 2015–16 IOPs than the 1998 IOP, especially for temperature. To be specific, JRA-55 and ERA-Interim have the smallest bias and RMSE of horizontal winds in 1998 and the two later campaigns, respectively, whereas CFSR has the biggest uncertainties with the largest vertical variations in winds during three campaigns, especially in meridional wind. For temperature, ERA-Interim has the smallest bias and RMSE at almost all vertical levels during the three campaign periods, although it has the largest value at the upper levels in 1998. However, ERA-Interim has the biggest bias and RMSE of specific humidity in 1998 and then decreases sharply in the two later campaigns, while JRA-55 has the least biases and uncertainties in specific humidity during the three campaigns. Notably, JRA-55 presents the relatively smaller mean bias, bias spread, and RMSE for each variable in 1998 as well as the smallest variations from campaign to campaign. This is likely because JRA-55

assimilated the sounding observations during the 1998 TIPEX-II IOP.

These mean biases, bias spread, and mean RMSE (R-O) greatly diminished after 1998, which is likely a direct consequence of changing observing systems, particularly the big changes in the satellite observations. In the 1998 IOP (TOVS period), MSU was the only instrument assimilated by reanalysis systems monitoring the troposphere thermal structure. In the 2008 and 2015–16 IOPs (ATOVS period), AMSU-A, as the updated instrument of MSU, can provide higher precision tropospheric temperature observations. AMSU-B, another ATOVS-suite instrument, is a new humidity sounder that can obtain humidity profiles below about 10 km. Second, the additional satellite radiances, made by other instruments such as AIRS, IASI, and GPS RO, also are assimilated in several reanalysis systems from 1999 to the present. The introduction of these satellite observations has made important contributions to improving troposphere forecast skill. That is to say, the huge increase in the volume and quality of satellite observations is likely the main reason why these reanalyses might have a higher quality and accuracy and are more similar to each other in the 2008 and 2015–16 IOPs (ATOVS period) than in the 1998 IOP (TOVS period). In addition, the improvement of radiosonde technologies should have positive impacts on reducing the bias between reanalysis products and observations.

There are several reasons for disagreements among reanalyses. One of the primary reasons is the assimilated observation data are not quite the same in these reanalysis systems. For instance, 1998 TIPEX-II sounding observations only are assimilated in JRA-55, but AIRS and IASI are not introduced in JRA-55. The second is the information assimilated from a given sensor may differ from reanalysis to reanalysis, which is caused by different selections of which channels can be introduced into the reanalysis systems and different quality-control methods to exclude passive observations for the same instrument. All in all, the reanalysis quality strongly depends not only on the forecast model and data assimilation methodology used in the respective analysis, it also critically depends on the assimilated information from ever-evolving observing systems, especially satellite radiances (e.g., observation input sources, selected channels for a given satellite sensor, quality-control methods), the latter of which are likely more difficult to remove from assessing regional climate change signals using any of the modern reanalyses that are designed to optimize the analysis accuracy using all applicable observations at the analysis time. One must be careful when using reanalysis temperatures and humidity for regional climate trends, in

particular over data-sparse regions such as the Tibetan Plateau, because their quality and accuracy may have varied over time.

**Acknowledgments.** This study was jointly supported by the National Natural Science Foundation of China (Grants 41775002 and 41605028), National Key Research and Development Program of China (2018YFC1505705), Basic Research Fund of the Chinese Academy of Meteorological Sciences (Grant 1920202231), and the U.S. National Science Foundation Grants AGS-1305798 and 1712290. We are grateful to Prof. Xiangde Xu at the Chinese Academy of Meteorological Sciences for providing us the quality-controlled radiosonde observations collected during the intense observation periods of TIPEX-II and JICA/Tibet Project, and to Yang Zhao for his assistance in processing reanalyses data. The 2015–16 quality-controlled radiosonde observations are provided by the TIPEX-III project team. The field experiment sounding observations are archived in the Information Service Center of the China Meteorological Administration and can be made available upon application for approval. CFSR and CFSv2 datasets are provided by the National Centers for Environmental Prediction (NCEP), the ERA-Interim dataset is provided by the European Centre for Medium-Range Weather Forecasts (ECMWF), the JRA-55 dataset is provided by Japan Meteorological Agency (JMA), and MERRA-2 is provided by the National Aeronautics and Space Administration (NASA).

## REFERENCES

Anthes, R. A., and Coauthors, 2008: The COSMIC/FORMOSAT-3 mission: Early results. *Bull. Amer. Meteor. Soc.*, **89**, 313–333, <https://doi.org/10.1175/BAMS-89-3-313>.

Bao, X., and F. Zhang, 2013: Evaluation of NCEP–CFSR, NCEP–NCAR, ERA-Interim, and ERA-40 reanalysis datasets against independent sounding observations over the Tibetan Plateau. *J. Climate*, **26**, 206–214, <https://doi.org/10.1175/JCLI-D-12-00056.1>.

—, —, and J. Sun, 2011: Diurnal variations of warm season precipitation east of the Tibetan Plateau over China. *Mon. Wea. Rev.*, **139**, 2790–2810, <https://doi.org/10.1175/MWR-D-11-0006.1>.

Bengtsson, L., S. Hagemann, and K. I. Hodges, 2004: Can climate trends be calculated from reanalysis data? *J. Geophys. Res.*, **109**, D11111, <https://doi.org/10.1029/2004JD004536>.

Bennartz, R., A. Thoss, A. Dybbroe, and D. B. Michelson, 2002: Precipitation analysis using the Advanced Microwave Sounding Unit in support of nowcasting applications. *Meteor. Appl.*, **9**, 177–189, <https://doi.org/10.1017/S1350482702002037>.

Bian, J. C., H. B. Chen, H. Vömel, Y. J. Duan, Y. J. Xuan, and D. R. Lu, 2011: Intercomparison of humidity and temperature sensors: GTS1, Vaisala RS80, and CFH. *Adv. Atmos. Sci.*, **28**, 139–146, <https://doi.org/10.1007/s00376-010-9170-8>.

Bosilovich, M. G., 2013: Regional climate and variability in NASA MERRA and recent reanalyses: U.S. summertime precipita-

tion and temperature. *J. Appl. Meteor. Climatol.*, **52**, 1939–1951, <https://doi.org/10.1175/JAMC-D-12-0291.1>.

—, F. R. Robertson, and J. Chen, 2011: Global energy and water budgets in MERRA. *J. Climate*, **24**, 5721–5739, <https://doi.org/10.1175/2011JCLI4175.1>.

—, —, L. Takacs, A. Molod, and D. Mocko, 2017: Atmospheric water balance and variability in the MERRA-2 reanalysis. *J. Climate*, **30**, 1177–1196, <https://doi.org/10.1175/JCLI-D-16-0338.1>.

Brogniez, H., and R. T. Pierrehumbert, 2007: Intercomparison of tropical tropospheric humidity in GCMs with AMSU-B water vapor data. *Geophys. Res. Lett.*, **34**, L17812, <https://doi.org/10.1029/2006GL029118>.

Chang, E. K. M., and A. M. W. Yau, 2016: Northern Hemisphere winter storm track trends since 1959 derived from multiple reanalysis datasets. *Climate Dyn.*, **47**, 1435–1454, <https://doi.org/10.1007/s00382-015-2911-8>.

Chen, G., T. Iwasaki, H. Qin, and W. Sha, 2014: Evaluation of the warm-season diurnal variability over East Asia in recent reanalyses JRA-55, ERA-Interim, NCEP CFSR, and NASA MERRA. *J. Climate*, **27**, 5517–5537, <https://doi.org/10.1175/JCLI-D-14-0005.1>.

Collard, A. D., and A. P. McNally, 2009: The assimilation of Infrared Atmospheric Sounding Interferometer radiances at ECMWF. *Quart. J. Roy. Meteor. Soc.*, **135**, 1044–1058, <https://doi.org/10.1002/qj.410>.

Cucurull, L., R. A. Anthes, and L. Tsao, 2014: Radio occultation observations as anchor observations in numerical weather prediction models and associated reduction of bias corrections in microwave and infrared satellite observations. *J. Atmos. Oceanic Technol.*, **31**, 20–32, <https://doi.org/10.1175/JTECH-D-13-00059.1>.

Davis, S. M., and Coauthors, 2017: Assessment of upper tropospheric and stratospheric water vapour and ozone in reanalyses as part of S-RIP. *Atmos. Chem. Phys.*, **17**, 12 743–12 778, <https://doi.org/10.5194/acp-17-12743-2017>.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.

Dirksen, R. J., M. Sommer, F. J. Immler, D. F. Hurst, R. Kivi, and H. Vömel, 2014: Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde. *Atmos. Meas. Tech.*, **7**, 4463–4490, <https://doi.org/10.5194/amt-7-4463-2014>.

Dufour, A., O. Zolina, and S. K. Gulev, 2016: Atmospheric moisture transport to the Arctic: Assessment of reanalyses and analysis of transport components. *J. Climate*, **29**, 5061–5081, <https://doi.org/10.1175/JCLI-D-15-0559.1>.

English, S. J., R. J. Renshaw, P. C. Dibben, A. J. Smith, P. J. Rayer, C. Poulsen, F. W. Saunders, and J. R. Eyre, 2000: A comparison of the impact of TOVS and ATOVS satellite sounding data on the accuracy of numerical weather forecasts. *Quart. J. Roy. Meteor. Soc.*, **126**, 2911–2931, <https://doi.org/10.1256/SMSQJ.56914>.

Fujiwara, M., and Coauthors, 2017: Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems. *Atmos. Chem. Phys.*, **17**, 1417–1452, <https://doi.org/10.5194/acp-17-1417-2017>.

Gelaro, R., and Y. Zhu, 2009: Examination of observation impacts derived from observing system experiments (OSEs) and adjoint models. *Tellus*, **61A**, 179–193, <https://doi.org/10.1111/j.1600-0870.2008.00388.x>.

—, and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *J. Climate*, **30**, 5419–5454, <https://doi.org/10.1175/JCLI-D-16-0758.1>.

Guidard, V., N. Fourrié, P. Brousseau, and F. Rabier, 2011: Impact of IASI assimilation at global and convective scales and challenges for the assimilation of cloudy scenes. *Quart. J. Roy. Meteor. Soc.*, **137**, 1975–1987, <https://doi.org/10.1002/qj.928>.

Ingleby, B., 2017: An assessment of different radiosonde types 2015/2016. ECMWF Tech. Memo. 807, 69 pp., <https://www.ecmwf.int/en/elibrary/17551-assessment-different-radiosonde-types-2015-2016>.

Ji, F., Z. Wu, J. Huang, and E. P. Chassignet, 2014: Evolution of land surface air temperature trend. *Nat. Climate Change*, **4**, 462–466, <https://doi.org/10.1038/NCLIMATE2223>.

Jones, R. W., I. A. Renfrew, A. Orr, B. G. M. Webber, D. M. Holland, and M. A. Lazzara, 2016: Evaluation of four global reanalysis products using in situ observations in the Amundsen Sea Embayment, Antarctica. *J. Geophys. Res.*, **121**, 6240–6257, <https://doi.org/10.1002/2015JD024680>.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).

Kishore, P., S. Jyothi, G. Basha, S. V. B. Rao, M. Rajeevan, I. Velicogna, and T. C. Sutterley, 2016: Precipitation climatology over India: Validation with observations and reanalysis datasets and spatial trends. *Climate Dyn.*, **46**, 541–556, <https://doi.org/10.1007/s00382-015-2597-y>.

Kobayashi, S., and Coauthors, 2015: The JRA-55 Reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48, <https://doi.org/10.2151/jmsj.2015-001>.

Lavaysse, C., C. Flamant, A. Evan, S. Janicot, and M. Gaetani, 2016: Recent climatological trend of the Saharan heat low and its impact on the West African climate. *Climate Dyn.*, **47**, 3479–3498, <https://doi.org/10.1007/s00382-015-2847-z>.

Lindsay, R. W., M. Wensnahan, A. Schweiger, and J. Zhang, 2014: Evaluation of seven different atmospheric reanalysis products in the Arctic. *J. Climate*, **27**, 2588–2606, <https://doi.org/10.1175/JCLI-D-13-00014.1>.

Long, C. S., M. Fujiwara, S. Davis, D. M. Mitchell, and C. J. Wright, 2017: Climatology and interannual variability of dynamic variables in multiple reanalyses evaluated by the SPARC Reanalysis Intercomparison Project (S-RIP). *Atmos. Chem. Phys.*, **17**, 14 593–14 629, <https://doi.org/10.5194/acp-17-14593-2017>.

Manney, G. L., and Coauthors, 2017: Reanalysis comparisons of upper tropospheric-lower stratospheric jets and multiple tropopauses. *Atmos. Chem. Phys.*, **17**, 11 541–11 566, <https://doi.org/10.5194/acp-17-11541-2017>.

Manzanas, R., L. K. Amekudzi, K. Preko, S. Herrera, and J. M. Gutiérrez, 2014: Precipitation variability and trends in Ghana: An intercomparison of observational and reanalysis products. *Climatic Change*, **124**, 805–819, <https://doi.org/10.1007/s10584-014-1100-9>.

Marshall, G. J., 2003: Trends in the southern annular mode from observations and reanalyses. *J. Climate*, **16**, 4134–4143, [https://doi.org/10.1175/1520-0442\(2003\)016<4134:TTSAM>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<4134:TTSAM>2.0.CO;2).

McCarty, W., and Coauthors, 2016: MERRA-2 input observations: Summary and assessment. NASA Tech. Rep. NASA/TM-2015-104606, Vol. 46, 51 pp.

Miloshevich, L. M., H. Vömel, A. Paukkunen, A. J. Heymsfield, and S. J. Oltmans, 2001: Characterization and correction of relative humidity measurements from Vaisala RS80-A radiosondes at cold temperatures. *J. Atmos. Oceanic Technol.*, **18**, 135–156, [https://doi.org/10.1175/1520-0426\(2001\)018<0135:CACORH>2.0.CO;2](https://doi.org/10.1175/1520-0426(2001)018<0135:CACORH>2.0.CO;2).

Molnar, P., W. R. Boos, and D. S. Battisti, 2010: Orographic controls on climate and paleoclimate of Asia: Thermal and mechanical roles for the Tibetan Plateau. *Annu. Rev. Earth Planet. Sci.*, **38**, 77–102, <https://doi.org/10.1146/annurev-earth-040809-152456>.

Nash, J., R. Smout, T. Oakley, B. Pathnack, and S. Kumosenko, 2006: WMO Intercomparison of High Quality Radiosonde Systems—Final Report, Vacoas, Mauritius, 2–25 February 2005. WMO Instruments and Observing Methods Rep. 83, 115 pp., [https://library.wmo.int/doc\\_num.php?explnum\\_id=9312](https://library.wmo.int/doc_num.php?explnum_id=9312).

—, T. Oakley, H. Vömel, and W. Li, 2011: WMO Intercomparison of High Quality Radiosonde Systems Yangjiang, China, 12 July–3 August 2010. WMO Instruments and Observing Methods Rep. 107, 249 pp., [https://library.wmo.int/doc\\_num.php?explnum\\_id=9467](https://library.wmo.int/doc_num.php?explnum_id=9467).

Nicolas, J. P., and D. H. Bromwich, 2014: New reconstruction of Antarctic near-surface temperatures: Multidecadal trends and reliability of global reanalyses. *J. Climate*, **27**, 8070–8093, <https://doi.org/10.1175/JCLI-D-13-00733.1>.

Nützel, M., M. Dameris, and H. Garny, 2016: Movement, drivers and bimodality of the South Asian High. *Atmos. Chem. Phys.*, **16**, 14 755–14 774, <https://doi.org/10.5194/acp-16-14755-2016>.

Oakley, T., 1998: Report by the rapporteur on radiosonde compatibility monitoring—Part A: WMO catalogue of radiosondes and upper-air wind systems in use by members (1998) and Part B: Compatibility of radiosonde geopotential measurements 1995, 1996 and 1997. WMO Instruments and Observing Methods Rep. 72, 195 pp., [https://library.wmo.int/doc\\_num.php?explnum\\_id=9701](https://library.wmo.int/doc_num.php?explnum_id=9701).

Parker, W. S., 2016: Reanalyses and observations: What's the difference? *Bull. Amer. Meteor. Soc.*, **97**, 1565–1572, <https://doi.org/10.1175/BAMS-D-14-00226.1>.

Poli, P., S. B. Healy, and D. P. Dee, 2010: Assimilation of global positioning system radio occultation data in the ECMWF ERA-Interim reanalysis. *Quart. J. Roy. Meteor. Soc.*, **136**, 1972–1990, <https://doi.org/10.1002/qj.722>.

Robertson, F. R., M. G. Bosisovich, J. Chen, and T. L. Miller, 2011: The effect of satellite observing system changes on MERRA water and energy fluxes. *J. Climate*, **24**, 5197–5217, <https://doi.org/10.1175/2011JCLI4227.1>.

—, —, J. B. Roberts, R. H. Reichle, R. Adler, L. Ricciardulli, W. Berg, and G. J. Huffman, 2014: Consistency of estimated global water cycle variations over the satellite era. *J. Climate*, **27**, 6135–6154, <https://doi.org/10.1175/JCLI-D-13-00384.1>.

Robson, J., P. Ortega, and R. Sutton, 2016: A reversal of climatic trends in the North Atlantic since 2005. *Nat. Geosci.*, **9**, 513–517, <https://doi.org/10.1038/ngeo2727>.

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057, <https://doi.org/10.1175/2010BAMS3001.1>.

—, and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185–2208, <https://doi.org/10.1175/JCLI-D-12-00823.1>.

Serreze, M. C., A. P. Barrett, and J. Stroeve, 2012: Recent changes in tropospheric water vapor over the Arctic as assessed from radiosondes and atmospheric reanalyses. *J. Geophys. Res.*, **117**, D10104, <https://doi.org/10.1029/2011JD017421>.

Si, D., and Y. Ding, Y., 2013: Decadal change in the correlation pattern between the Tibetan Plateau winter snow and the East Asian summer precipitation during 1979–2011. *J. Climate*, **26**, 7622–7634, <https://doi.org/10.1175/JCLI-D-12-00587.1>.

Siam, M. S., M. Demory, and E. A. B. Eltahir, 2013: Hydrological cycles over the Congo and upper Blue Nile basins: Evaluation of general circulation model simulations and reanalysis products. *J. Climate*, **26**, 8881–8894, <https://doi.org/10.1175/JCLI-D-12-00404.1>.

Simmons, A. J., and P. Poli, 2015: Arctic warming in ERA-Interim and other analyses. *Quart. J. Roy. Meteor. Soc.*, **141**, 1147–1162, <https://doi.org/10.1002/qj.2422>.

—, A. Untch, C. Jakob, P. Källberg, and P. Undén, 1999: Stratospheric water vapour and tropical tropopause temperatures in ECMWF analyses and multi-year simulations. *Quart. J. Roy. Meteor. Soc.*, **125**, 353–386, <https://doi.org/10.1002/qj.49712555318>.

—, and Coauthors, 2004: Comparison of trends and low-frequency variability in CRU, ERA-40 and NCEP/NCAR analyses of surface air temperature. *J. Geophys. Res.*, **109**, D24115, <https://doi.org/10.1029/2004JD005306>.

—, K. M. Willett, P. D. Jones, P. W. Thorne, and D. P. Dee, 2010: Low-frequency variations in surface atmospheric humidity, temperature, and precipitation: Inferences from reanalyses and monthly gridded observational datasets. *J. Geophys. Res.*, **115**, D01110, <https://doi.org/10.1029/2009JD012442>.

—, P. Poli, D. P. Dee, P. Berrisford, H. Hersbach, S. Kobayashi, and C. Peubey, 2014: Estimating low-frequency variability and trends in atmospheric temperature using ERA-Interim. *Quart. J. Roy. Meteor. Soc.*, **140**, 329–353, <https://doi.org/10.1002/qj.2317>.

—, P. Berrisford, D. P. Dee, H. Hersbach, S. Hirahara, and J.-N. Thpaut, 2017: A reassessment of temperature variations and trends from global reanalyses and monthly surface climatological datasets. *Quart. J. Roy. Meteor. Soc.*, **143**, 101–119, <https://doi.org/10.1002/qj.2949>.

Singh, R., C. M. Kishtawal, S. P. Ojha, and P. K. Pal, 2012: Impact of assimilation of Atmospheric InfraRed Sounder (AIRS) radiances and retrievals in the WRF 3D-Var assimilation system. *J. Geophys. Res.*, **117**, D11107, <https://doi.org/10.1029/2011JD017367>.

Skrivankova, P., 2004: Vaisala radiosonde RS92 validation trial at Prague–Libus. Vaisala News, Vol. 164, Vantaa, Finland, 4–8, [https://www.vaisala.com/sites/default/files/documents/VN164\\_Vaisala\\_Radiosonde\\_RS92\\_Validation\\_Trial\\_at\\_Prague-Libus.pdf](https://www.vaisala.com/sites/default/files/documents/VN164_Vaisala_Radiosonde_RS92_Validation_Trial_at_Prague-Libus.pdf).

Steinbrecht, W., H. Claude, F. Schönenborn, U. Leiterer, H. Dier, and E. Lanzinger, 2008: Pressure and temperature differences between Vaisala RS80 and RS92 radiosonde systems. *J. Atmos. Oceanic Technol.*, **25**, 909–927, <https://doi.org/10.1175/2007JTECHA999.1>.

Thorne, P., and R. S. Vose, 2010: Reanalyses suitable for characterizing long-term trends. *Bull. Amer. Meteor. Soc.*, **91**, 353–361, <https://doi.org/10.1175/2009BAMS2858.1>.

Vömel, H., and Coauthors, 2007: Radiation dry bias of the Vaisala RS92 humidity sensor. *J. Atmos. Oceanic Technol.*, **24**, 953–963, <https://doi.org/10.1175/JTECH2019.1>.

Xie, H., J. Ye, X. Liu, and E. Chongyi, 2010: Warming and drying trends on the Tibetan Plateau (1971–2005). *Theor. Appl. Climatol.*, **101**, 241–253, <https://doi.org/10.1007/s00704-009-0215-9>.

Xu, X., L. Chen, and M. Zhou, 2002: *The Second Tibetan Plateau Experiment of Atmospheric Sciences: TIPEX-GAME/TIBET*. China Meteorological Press, 236 pp.

Ye, D., 1981: Some characteristics of the summer circulation over the Qinghai-Xizang (Tibet) Plateau and its neighborhood. *Bull. Amer. Meteor. Soc.*, **62**, 14–19, [https://doi.org/10.1175/1520-0477\(1981\)062<0014:SCOTSC>2.0.CO;2](https://doi.org/10.1175/1520-0477(1981)062<0014:SCOTSC>2.0.CO;2).

—, and G. Wu, 1998: The role of the heat source of the Tibetan Plateau in the general circulation. *Meteor. Atmos. Phys.*, **67**, 181–198, <https://doi.org/10.1007/BF01277509>.

Zhang, R., X. Xu, T. Koike, Y. Ma, and K. Yang, 2012: A China–Japan cooperative JICA atmospheric observing network over the Tibetan Plateau (JICA/Tibet Project): An overview. *J. Meteor. Soc. Japan*, **90C**, 1–16, <https://doi.org/10.2151/jmsj.2012-C01>.

Zhao, P., and Coauthors, 2018: The third atmospheric scientific experiment for understanding the Earth–atmosphere coupled system over the Tibetan Plateau and its effects. *Bull. Amer. Meteor. Soc.*, **99**, 757–776, <https://doi.org/10.1175/BAMS-D-16-0050.1>.