Evaluation of NCEP–CFSR, NCEP–NCAR, ERA-Interim, and ERA-40 Reanalysis Datasets against Independent Sounding Observations over the Tibetan Plateau

XINGHUA BAO AND FUQING ZHANG

Department of Meteorology, The Pennsylvania State University, University Park, Pennsylvania

ABSTRACT

The NCEP–NCAR reanalysis, NCEP Climate Forecast System Reanalysis (CFSR), 40-yr ECMWF Re-Analysis (ERA-40), and interim ECMWF Re-Analysis (ERA-Interim) products are evaluated with sounding observations from an enhanced radiosonde network available every 6 h during the Tibetan Plateau Experiment (TIPEX) conducted from 10 May to 9 August 1998. This study uses more than 3000 high-quality, independent rawinsondes at 11 stations (which were not assimilated in any of the reanalyses), which represents the first time that such a comprehensive evaluation is performed to assess the quality of these four most widely used reanalysis products over this region, which is highest in the world and crucial to the global climate and weather.

Averaging over the entire three-month period, it is found that each reanalysis dataset produces mean values of temperature and horizontal winds consistent with the verifying soundings (indicating relatively small mean bias); however, there are considerable differences (biases) in the mean relative humidity. On average, except for temperature at higher levels, both newer-generation reanalyses (CFSR and ERA-Interim) have smaller root-mean-square (RMS) error and bias than their predecessors (NCEP–NCAR and ERA-40). With some exceptions, the RMS errors of all variables for both CFSR and ERA-Interim (verifying with soundings) are similar in magnitude to the RMS difference between these two reanalyses, all of which are approximately twice as large as the corresponding observation errors. It is also found that there are strong diurnal variations in both RMS error and mean bias that differ greatly among different reanalyses and at different pressure levels.

1. Introduction

The Tibetan Plateau (TP) over central Asia is the world's largest and highest plateau with an average elevation of over 4500 m and an extensive area of 2.5 million km^2. The Tibetan Plateau has crucial influence on the climate and weather over East Asia and around the whole world due to both the thermodynamic and dynamic effects induced by the high terrains (e.g., Ding and Chan 2005; Bao et al. 2011). However, the adverse weather and environmental conditions limit our ability to make direct in situ measurements in this region.

In recent years, several global reanalysis datasets with high spatial and temporal resolution have been used to compensate for the lack of direct observations in the TP. The National Centers for Environment Prediction (NCEP) and the European Centre for Medium-Range Weather Forecasts (ECMWF) provide four widely used reanalysis datasets: the NCEP–National Center for Atmospheric Research (NCAR) Reanalysis Project (NNRP), the NCEP Climate Forecast System Reanalysis (CFSR), the 40-yr ECMWF Re-Analysis (ERA-40), and the interim ECMWF Re-Analysis (ERA-Interim). Given the inherent uncertainties in the forecast model, input data, and data assimilation, it is essential to assess the quality of these reanalyses (Hodges et al. 2011) and the reliability of their use in evaluating variations in weather and climate and/or as surrogates of observations to be assimilated into climate models. Several studies have compared the reanalysis from different sources for different regions (e.g., Betts et al. 2009; Fan et al. 2008; Mao et al. 2010; Mooney et al. 2010; Zhao and Fu 2006). In particular, recent studies of Frauenfeld et al. (2005) and Wang and Zeng (2012) examined the quality of the reanalysis products on the surface variables over the Tibetan Plateau. However, to the best of our knowledge, systematic evaluation of the quality of these reanalyses above the ground over the Tibetan Plateau is hardly available in the literature. Given the importance of the

Corresponding author address: Dr. Fuqing Zhang, Department of Meteorology, The Pennsylvania State University, University Park, PA 16802.
E-mail: fzhang@psu.edu
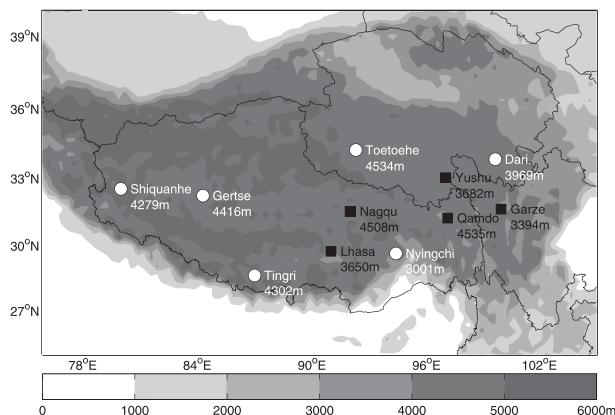
FIG. 1. Map plot of terrain elevations (shaded every 1000 m) over the Tibetan Plateau, and locations of the TIPEX IOP radiosonde sites. The black squares denote the routine sounding stations, and the white circles denote the added stations in TIPEX.

Tibetan Plateau in the regional and global weather and climate and the scarceness of observations in this region, it is very important to assess of the accuracy of different reanalyses at different pressure levels using high-quality observations such as those by rawinsonde observations from extended field experiments. Such assessments will also have direct significance for evaluating the quality and efficiency of different numerical weather prediction models and associated data assimilation systems over the Tibetan Plateau and elsewhere. Such assessments may also help the design of future generation observing systems and/or future field experiments over the Tibetan Plateau.

The enhanced radiosonde observations collected between 10 May and 9 August 1998 during the second Tibetan Plateau Experiment (TIPEX; Xu et al. 2002)—which were not assimilated in any of the reanalyses—provide a rare opportunity to verify independently the reliability of these reanalyses in this region, along with the diurnal variations in the data quality.

2. Data and methodology

The reanalysis products of NNRP, CFSR, ERA-40, and ERA-Interim are compared with independent sounding observations during the intensive observing period (IOP) of TIPEX from 10 May to 9 August 1998. Over the three-month IOP, sounding observations were collected every 6 h or 4 times per day at 11 locations covering a broad region of the Tibetan Plateau (Fig. 1). For completely independent verifications, we exclude the 0000 and 1200 UTC observations at Nagqu, Lhasa, Yushu, Garze, and Qamdu (which are assimilated in each reanalysis as part of the standard observing network). The TIPEX team (Xu et al. 2002) provided us

with quality-controlled observations from each sounding of four variables (temperature T, dewpoint depression, wind direction, and wind speed) at seven standard vertical levels (500, 400, 300, 250, 200, 150, and 100 hPa), from which we derived both components of the horizontal wind (U and V) as well as the relative humidity (RH). The TS-2A captive balloon radiosonde sensor was used for all the sounding observations during TIPEX. This is the same sensor as that used in regular sounding observations over China before 1999, the reliability of which over the Tibetan Plateau was assured through several intercomparison experiments before TIPEX formally started (Zhou et al. 2000). However, given that the radiosonde sensor used may not be particularly sensitive in the upper troposphere (such as above 400 hPa) in this region as noted by Bian et al. (2011), caution must be taken to interpolate the verification for relative humidity from reanalyses versus sounding observations.

For direct comparison of the gridded reanalysis with discrete soundings (Mooney et al. 2010), we first interpolate the reanalysis products (with simple bilinear interpolation) to each of the sounding locations at the same synoptic times and standard pressure levels. The NNRP, which was conducted at NCEP beginning in the early 1990s (Kalnay et al. 1996), is available for the period from 1948 to the present. The resolution of this global dataset is T62 (equivalent to 209 km) with 28 vertical sigma levels available every 6 h. An update of NNRP, CFSR, uses a high-resolution fully coupled model with the atmospheric component at T382 (38 km) resolution with 64 vertical levels from the surface to 0.26 hPa. It is available for the period from 1979 to 2009 (Saha et al. 2010). In collaboration with many institutions (Uppala et al. 2005), the ECMWF completed in 2002 the ERA-40 dataset, which covers the period from mid-1957 to 2001 [including some 15-yr ECMWF Re-Analysis (ERA-15) data for 1979–93]. To produce analyses every 6 h, the three-dimensional variational data assimilation (3D-Var) technique was applied using the T159 (~125 km) and L60 (60 vertical levels) vertical version of the Integrated Forecasting System. ERA-Interim (Dee et al. 2011) is the latest ECMWF global atmospheric reanalysis from 1979 to the present. Compared with ERA-40, ERA-Interim used an improved atmospheric model (including an increase in horizontal resolution to T255, or 80 km) and a more advanced assimilation system (4D-Var rather than 3D-Var). More details about each reanalysis can be found in the references cited above.

3. Overall RMS error and biases

The primary purpose of this study is to understand the quality and utility of the four reanalysis datasets over the
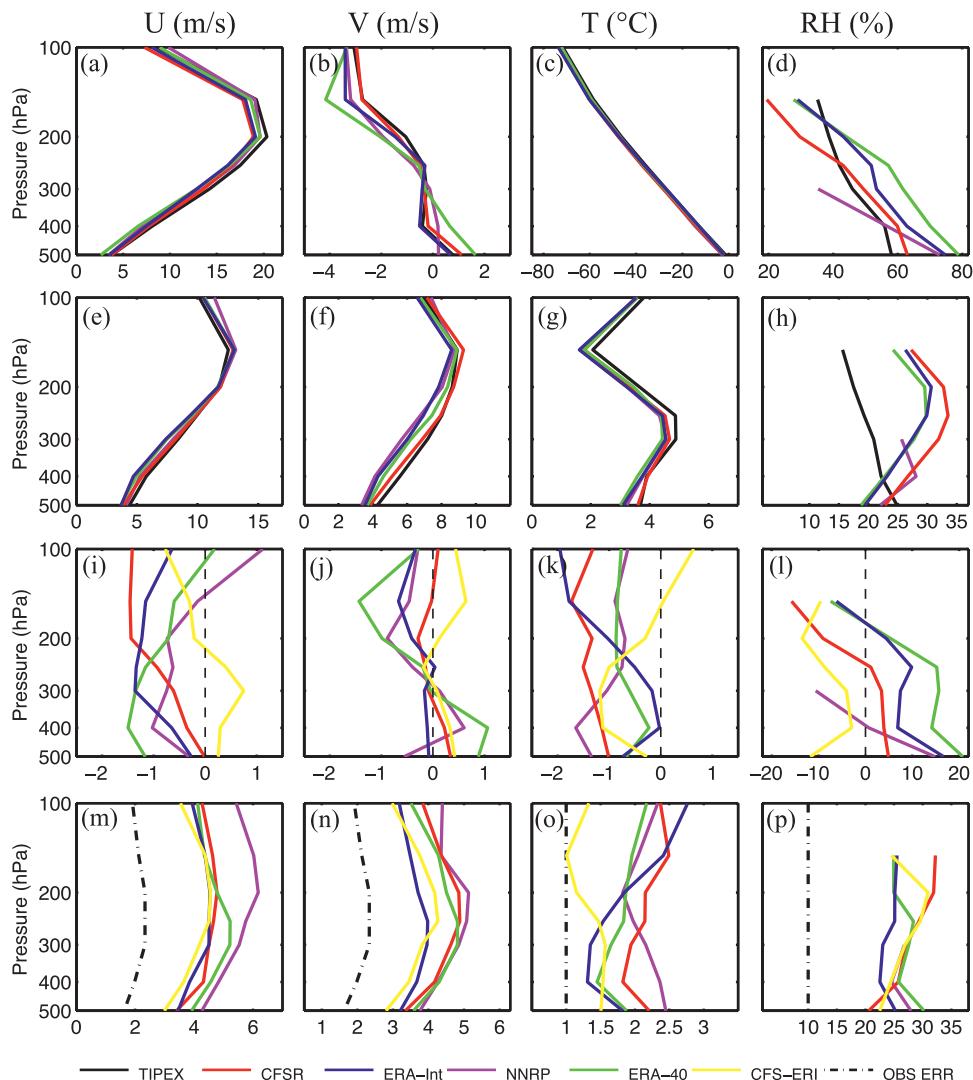
FIG. 2. (top) Vertical profiles of mean (a) *U*, (b) *V*, (c) *T*, and (d) RH averaged over all independent TIPEX IOP soundings during 10 May–9 Aug 1998, and the corresponding mean interpolated from the four reanalysis products. (second row) Vertical profiles of standard deviation (STD) (e) *U*, (f) *V*, (g) *T*, and (h) RH for the TIPEX IOP soundings and four reanalysis products. (third row) Vertical profiles of the mean biases for each reanalysis verifying against the TIPEX soundings, as well as the mean difference between CFSR and ERA-Interim for (i) *U*, (j) *V*, (k) *T*, and (l) RH. (bottom) Vertical profiles of the RMS errors for each reanalysis verifying against the TIPEX soundings for (m) *U*, (n) *V*, (o) *T*, and (p) RH, along with the NCEP standard sounding observation error for these variables.

TP in terms of root-mean-square (RMS) error and mean bias verified against all independent sounding observations obtained during the three-month TIPEX IOP.

Figures 2a–d show the mean vertical profiles of *U*, *V*, *T*, and RH averaged over all the TIPEX IOP soundings, and the corresponding averages for the interpolated soundings derived from each of the four reanalysis products; Figs. 2e–h show the corresponding standard deviations of these variables. Figures 2i–l show the mean biases of the four reanalyses while Figs. 2m–p show the

corresponding RMS error verified against the sounding observations. For this 3-month verification period, the averaged observed winds derived from the soundings are predominantly westerlies that increase from ~5 m s^{-1} at 500 hPa to a peak of ~20 m s^{-1} at 200 hPa. The standard deviations of *U* (*V*) from five datasets increase from ~4.5 m s^{-1} (4 m s^{-1}) at 500 hPa to be peaked near the jet maximum (150–200 hPa) with a maximum of ~14 m s^{-1} (9 m s^{-1}) (Figs. 2e,f). Broadly speaking, the vertical profiles of the mean and standard deviation for *U* and *V* in all

four reanalysis products closely follow those averaged over the verifying sounding observations, indicating that all reanalyses capture well the mean and variation of the horizontal wind fields.

The mean biases of U and V for each dataset are rather small and mostly within 1 m s^{-1} throughout the vertical column (Figs. 2i,j). All reanalyses have some small underestimation of the westerlies at all levels (negative bias in U, except for NNRP and ERA-40 at 100 hPa) and some small overestimation of northerlies above 300 hPa (positive bias below 300 hPa except for ERA-Interim and negative bias above). The V biases for both CFSR and ERA-Interim are extremely small (<0.5 m s^{-1}) (Figs. 2i,j).

The RMS error of U from the interpolated reanalysis verified against the soundings (Fig. 2m) is the smallest for CFSR and ERA-Interim (<3.5 m s^{-1} at 500 hPa and $\sim$4.5 m s^{-1} above), slightly higher for ERA-40 (with a maximum of $\sim$5 m s^{-1} at 300 hPa), and clearly the largest for NNRP (from $\sim$4.2 m s^{-1} at 500 hPa to a maximum of $\sim$6 m s^{-1} at 200 hPa). On average the NNRP RMS error of U is about 1 m s^{-1} larger than those of CFSR and ERA-Interim. The RMS error of V (Fig. 2n) is clearly the smallest for ERA-Interim (<4 m s^{-1} at all levels), the second smallest for ERA-40 and CFSR, both of which are very close to each other (similar to their RMS error of U), and again clearly the largest for NNRP (albeit $\sim$1 m s^{-1} smaller than its corresponding RMS error of U). Consistent with the RMS errors, the correlation coefficient between each reanalysis and the sounding observations for both U and V is generally high at upper levels but drops to no higher than 0.70 at 500 hPa for all reanalyses (Table 1). Both newer-generation reanalyses (CFSR and ERA-Interim) correlate better to the sounding observations than the corresponding older-generation reanalyses (NNRP and ERA-40).

Note that the RMS errors of both U and V for reanalyses (Figs. 2m,n) are nearly twice as large as the NCEP default observational error for radiosondes (also shown in Figs. 2m,n; see http://www.nco.ncep.noaa.gov/pmb/codes/nwprod/sorc/hwrf_v3.fd/var/obsproc/obserr.txt.) Thus, caution must be taken when using these reanalyses to verify daily weather. On the other hand, the relatively small overall biases for both U and V suggest that the reanalyses of the horizontal winds are very reliable for plateau-scale averages over seasonal or longer time scales.

Given its strong vertical gradient, the mean and standard deviation of temperature of each reanalysis (Fig. 2c,g) are hardly distinguishable from the verifying sounding mean throughout the vertical layers. The vertical profiles of standard deviation of temperature from the

TABLE 1. Correlation coefficients between each reanalysis and the sounding observations for different variables at different vertical levels.

Pressure (hPa)	CFSR				ERA-Interim			
	U	V	T	RH	U	V	T	RH
100	0.93	0.86	0.99		0.93	0.89	0.99	
150	0.94	0.88	0.98	0.86	0.95	0.92	0.98	0.92
200	0.93	0.84	0.96	0.83	0.94	0.90	0.97	0.90
250	0.91	0.81	0.97	0.80	0.91	0.86	0.97	0.87
300	0.87	0.79	0.96	0.77	0.88	0.83	0.97	0.83
400	0.74	0.71	0.93	0.70	0.78	0.76	0.94	0.75
500	0.70	0.69	0.85	0.67	0.69	0.70	0.89	0.68

Pressure (hPa)	NNRP				ERA-40			
	U	V	T	RH	U	V	T	RH
100	0.90	0.83	0.98		0.93	0.88	0.99	
150	0.91	0.88	0.98		0.95	0.89	0.98	0.89
200	0.89	0.82	0.96		0.93	0.86	0.97	0.86
250	0.86	0.78	0.95		0.88	0.81	0.96	0.81
300	0.81	0.75	0.93	0.74	0.83	0.76	0.96	0.75
400	0.68	0.67	0.90	0.66	0.72	0.69	0.93	0.68
500	0.55	0.58	0.83	0.55	0.62	0.66	0.88	0.63

sounding and four reanalysis are very similar with each other with the maximum $\sim$5°C around 300 hPa and the minimum 2°C around 150 hPa (Fig. 2g). Correspondingly, the correlation coefficients between the interpolated temperature and the verifying soundings are similar and high among each reanalysis (Table 1). The higher the altitude, the stronger the correlation with the correlation coefficient above 0.90 at all levels for each reanalysis except for 500 hPa.

Nevertheless, all reanalyses have some degree of cold bias, albeit with considerable variation from level to level (Fig. 2k). The ERA-40 has the least negative overall bias ($\sim$−0.7°C averaged over all levels) whereas CFSR has the most negative bias ($\sim$−1.3°C averaged over all levels). The ERA-Interim has the least negative bias below 250 hPa but becomes the most negatively biased at 100 hPa ($\sim$−1.9°C). On the other hand, the NNRP is or nearly is the least negatively biased at top levels (between $\sim$−0.6° and −0.9°C at or above 250 hPa) but becomes the most negatively biased at 400 and 500 hPa ($\sim$−1.6°C). Compared with their predecessors, both newer reanalyses have a smaller bias in T at lower levels but a larger bias at upper levels.

For the relative humidity, the 3-month mean of each reanalysis differs greatly and from one another and from the verifying observed sounding mean (which decreases from $\sim$57% at 500 hPa to $\sim$40% at 150 hPa; Fig. 2d). All reanalyses are more humid than the mean sounding observations at lower levels but become drier than observations at higher levels (Fig. 2l). In terms of mean

bias (Fig. 2l), CFSR is the smallest overall ranging from ~5% at 500 hPa to near zero at 250 hPa and to −15% at 150 hPa. Both of the ERA reanalyses have considerably more positive bias than CFSR at or below 250 hPa (with 12%–16% for ERA-40 and 8%–12% for ERA-Interim) but are closer to observations at 200 and 150 hPa. The standard deviations of RH in each of the reanalyses are considerably larger than those estimated from the soundings at nearly all levels with the biggest differences observed at 200–300 hPa (Fig. 2h). In comparison, the difference of standard deviations of RH among different reanalyses is much smaller.

The RMS errors of RH for all reanalyses (Fig. 2p) are greater than 20% at all vertical levels. CFSR actually has both the smallest (~21% at 500 hPa) and largest (~32% at 200 and 150 hPa) RMS errors of all reanalyses. The ERA-Interim has the smallest overall RMS error in RH ranging without much change in magnitude (23%–26%) throughout the vertical column. The RMS errors between the reanalyses and the sounding observations are large, suggesting that the quality of the moisture analysis may be highly uncertain. Correspondingly, the correlation coefficients between the interpolated RH and the verifying soundings are generally the weakest for each reanalysis among all variables at all levels (Table 1). However, part of the large RMS errors and weak correlations may be due to the quality of the observations themselves, given the 10% assumed observation error for RH at and above 500 hPa.

It is worth noting that, as with the horizontal wind field, the RMS errors of RH for both CFSR and ERA-Interim (verifying with soundings) are similar in magnitude to the RMS difference between these two reanalyses, all of which are approximately twice as large as the observation errors (Figs. 2m–p). The RMS error of temperature for CFSR is considerably larger than that of ERA-Interim and the RMS difference between these two reanalyses at lower levels. At upper levels, the RMS errors of most variables for both CFSR and ERA-Interim are similar in magnitude, both of which are slightly larger than the RMS difference between the two reanalyses or about twice the observation error.

Given strong inhomogeneity in the density of these 11 stations, and the lack of regular sounding observations in the entire western Tibetan Plateau, we further examine the subregional dependence of the mean bias of these four reanalyses (Fig. 3). We first subdivide the 11 stations into three groups: 1) the western plateau group to the west of 90°E that includes Shiquanhe, Gertse, and Tingri with no regular sounding stations; 2) the central plateau group for the four stations (Lhasa, Nagqu, Toetoehe, and Nyingchi) between 90° and 96°E; and 3) the eastern plateau group for the four stations (Qamdo,

Yushu, Dari, and Garze) to the east of 96°E (refer to Fig. 1). Not surprisingly, although the overall structure between different reanalyses is grossly consistent with the plateau-wise averages, there are considerable differences in biases among different subregions for some reanalyses and some variables (cf. Figs. 3 and 2i–l).

For example, the mean U bias for NNRP is negative in the western TP, nearly zero in the central TP, but overall positive in the eastern TP, indicating a systematic shift of the upper-tropospheric jet in the NNRP reanalysis, and also to some extent in ERA-40 while the two reanalyses (CFSR and ERA-Interim) have considerably less subregional variability in U (Figs. 3a,e,i). Another example is the mean bias of temperature in ERA-40, which has a peak cold bias above 250 hPa in the western and central plateau whereas the larger cold bias is below 250 hPa in the eastern region. Overall, ERA-Interim has the smallest subregional variability, CFSR has slightly more, while both older-generation reanalyses (NNRP and ERA-40) have the largest subregional variability of mean biases. Nevertheless, despite the regional dependence in mean biases, the difference of the RMS error of each variable for each reanalysis (including NNRP and ERA-40) among the three subregions is much less evident (not shown). We also divide the verification period into three monthly periods and examine the variability of the mean bias and RMS error among different subperiods. Overall the difference is small for all variables and all reanalyses, for both the mean bias and the error (not shown), and thus will not be discussed in detail here.

4. Diurnal variations in the RMS errors and biases

The high-frequency TIPEX IOP soundings also provide a rare opportunity to evaluate the diurnal variations of the RMS error and bias at different levels by different reanalysis products. This will further add to our understanding of the uncertainties in the reanalysis as a surrogate of observations, as well as in the reliability of using this analysis for examining the regional-scale diurnal cycles (e.g., He and Zhang 2010; Bao et al. 2011). Here we focus only on the two newer-generation reanalysis products, CFSR and ERA-Interim. It is clear from Figs. 4 and 5 that there are strong diurnal variations in both bias and RMS error in both reanalyses. The degree of diurnal variation also differs greatly at different pressure levels.

For CFSR, the mean U-wind bias (Fig. 4a) has a predominant diurnal negative peak of <-2.5 m s^{-1} at 0600 UTC [1400 Beijing standard time (BST)] above 250 hPa but at the same time the lower-level bias is at its minimum (~-0.3 to $+0.2$ m s^{-1} at 400–500 hPa). The
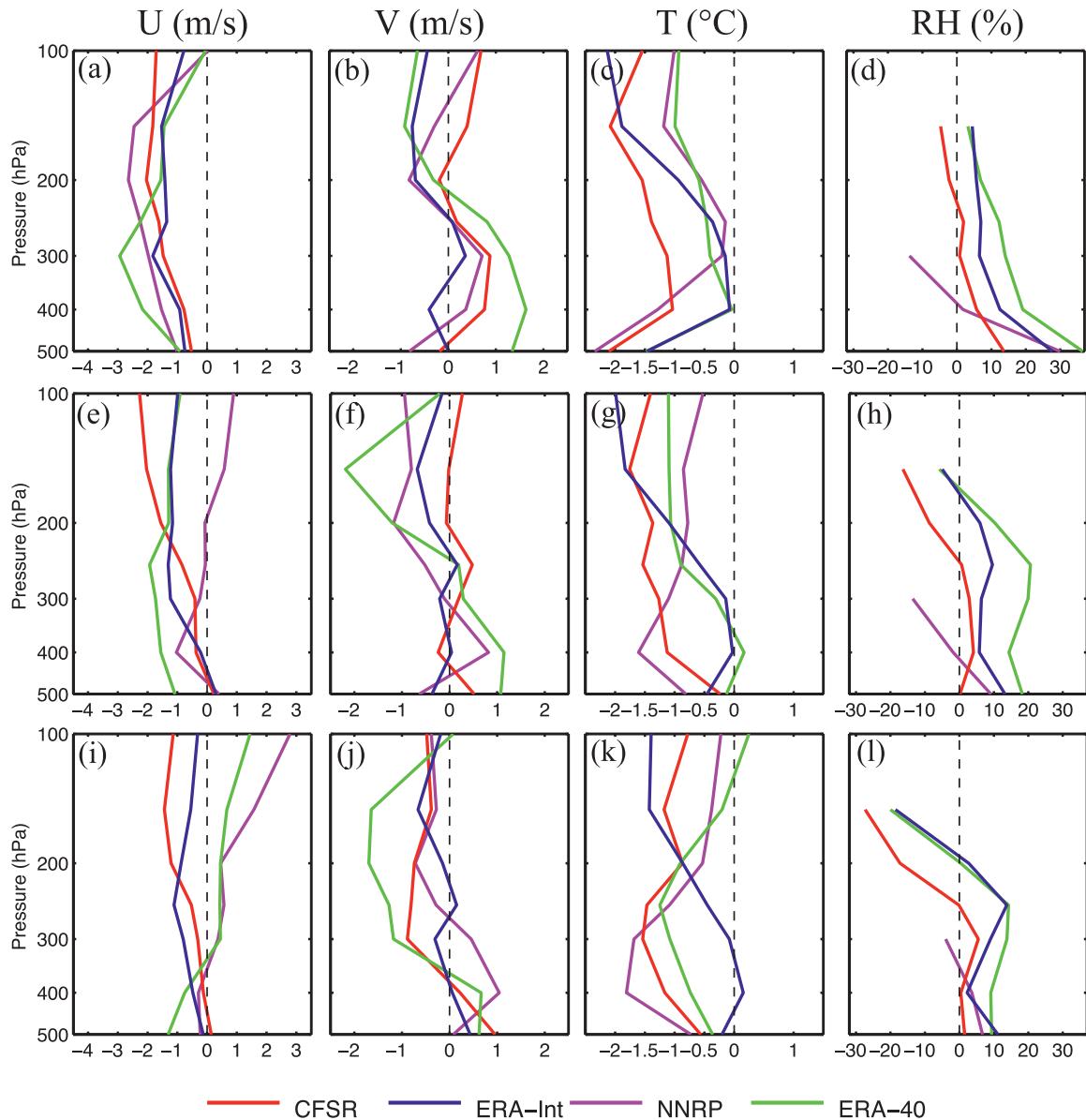
FIG. 3. (Top) Vertical profiles of the mean bias of (a) *U*, (b) *V*, (c) *T*, and (d) RH averaged over three stations in the western plateau to the west of 90°E (Shiquanhe, Gertse, and Tingri) from independent TIPEX IOP soundings. (middle) Vertical profiles of the mean bias of (e) *U*, (f) *V*, (g) *T*, and (h) RH averaged over four stations in the central plateau between 90° and 96°E (Lhasa, Nagqu, Nyingchi, and Toetoehe). (bottom) Vertical profiles of the mean bias of (i) *U*, (j) *V*, (k) *T*, and (l) RH averaged over four stations in the eastern plateau to the east of 96°E (Qamdo, Yushu, Dari, and Garze).

low-level mean *U*-wind bias has a positive peak of ~0.5 at 1800 UTC (0200 BST) and a negative peak of ~−0.5 m s^{-1} at 1200 UTC (2000 BST). The mean *V*-wind bias (Fig. 4b) also has a different diurnal cycle at different pressure levels: at upper levels, the positive bias peaks at 1200 UTC (2000 BST) whereas the negative bias peaks at 1800 UTC (0200 BST). At lower levels, the positive bias of the *V*-wind peaks from 1800 UTC at 500 hPa to 0000 UTC at 300–400 hPa. The peak negative bias of *T* (Fig. 4c) occurs mostly at 1200 UTC

(2000 BST) except for at 250 hPa, where the peak is at 1800 UTC (0200 BST). A secondary negative peak occurs at 0000 UTC (0800 BST) at 150 hPa. At lower levels, the weakest cold bias is centered at 0000 UTC (0800 BST). For the RH (Fig. 4d), there is a predominant peak wet (positive) bias at lower levels during the daytime (0000–1200 UTC or 0800–2000 BST) and a peak dry (negative) bias at upper levels at 1800 UTC (0200 BST).

For ERA-Interim, there is also a strong diurnal variation of mean bias in that differs from variable to variable
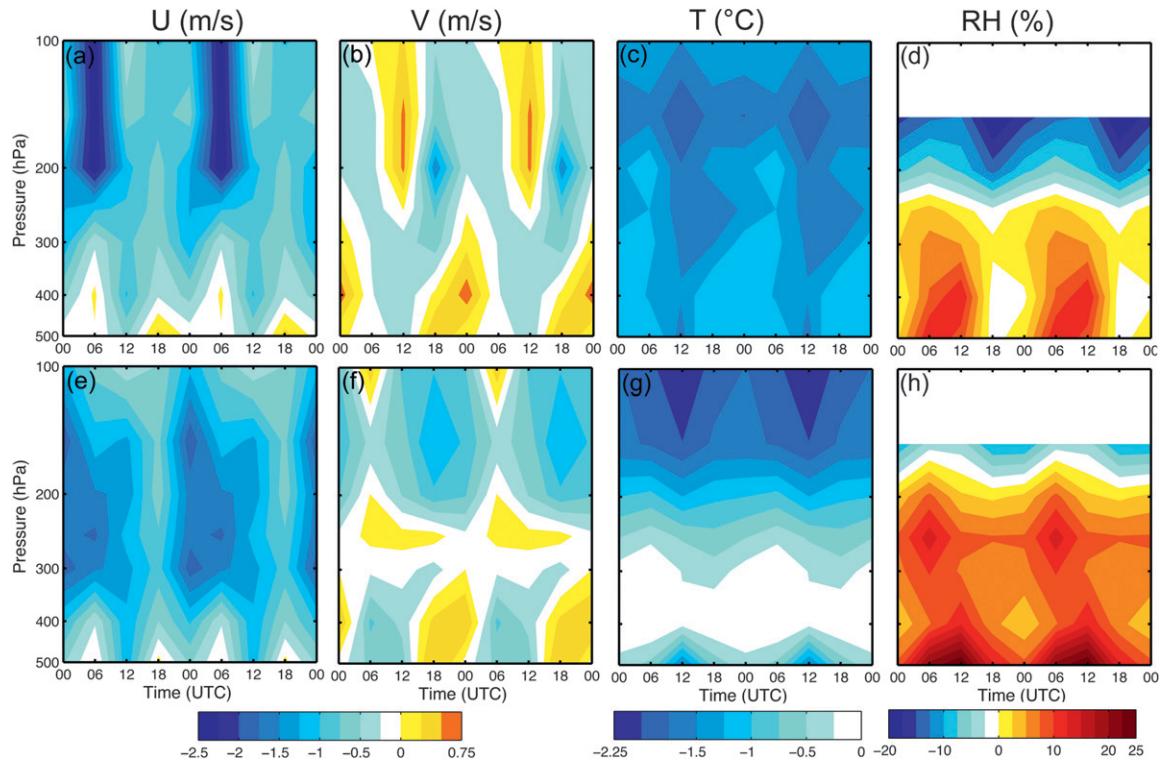
FIG. 4. The diurnal variations of the mean biases of *U*, *V*, *T*, and RH for (top) CFSR and (bottom) ERA-Interim at different pressure levels verifying against the TIPEX soundings.

and between pressure levels (Figs. 4e–h). The mean *U*-wind bias (Fig. 4e) has a negative diurnal peak at different levels at different times. The negative bias peaks at 1200 UTC (2000 BST) below 400 hPa, at 0600 UTC (1400 BST) for 200–300 hPa, and at 0000 UTC (0800 BST) above 200 hPa; there is a minute positive bias ($\sim$0.2 m s^{-1}) at 0600 and 1800 UTC (1400 and 0200 BST) at 500 hPa. For the *V*-wind (Fig. 4f) below 300 hPa, there is a negative peak at 0600 UTC (1400 BST) and a positive peak at 1800 UTC (0200 BST). At 100–200 hPa, the negative diurnal peak occurs at 1800 UTC (0200 BST). Consistent with CFSR, the cold bias of *T* (Fig. 4g) in ERA-Interim also has a general diurnal peak at 1200 UTC (2000 BST) at all levels, and a relative minimum at 0000 UTC (0800 BST). Also similar to CFSR, the wet bias of ERA-Interim (Fig. 3h) generally peaks during the daytime (0600–1200 UTC; 1400–2000 BST). In both CFSR and ERA-Interim, there appears to be some correlation between the biases of *T* and RH for reasons that are beyond the scope of this study (Figs. 4c,d,g,h).

There are also strong diurnal variations in the RMS error for both CFSR and ERA-Interim for different variables at different levels (Fig. 5). For *U* and *T* (and to a lesser extent in RH), it is apparent that the diurnal variations of the mean bias contribute strongly to the

diurnal variations of the RMS error in both reanalyses. For *U* (Figs. 5a,e), the maximum RMS error in both reanalyses can exceed 4–5 m s^{-1} from 300 to 100 hPa at the diurnal peak times. For *V* (Figs. 5b,f), the RMS error in both reanalyses has a diurnal peak at 1800 UTC (0200 BST) at almost all pressure levels with a maximum at around 250 hPa—this is not the case for the corresponding mean bias (Figs. 4b,f). For *T* (Figs. 5c,g), both reanalyses generally have a diurnal peak in RMS error (and cold/negative bias) at 1200 UTC (2000 BST) at all levels. For RH (Figs. 5d,h), the maximum RMS errors in both reanalyses have peaks at 1200 UTC (2000 BST) at lower levels. At higher levels, the RMS error for CFSR shifts forward in time to peak at 1800 UTC (0200 BST) at 150–200 hPa; the RMS error for ERA-Interim gradually shifts backward in time from peaking at 0600 UTC (1400 BST) at 250 hPa to 1800 UTC (0200 BST) at 150 hPa.

5. Concluding remarks

The quality and reliability of the NCEP–NCAR, NCEP CFSR, ERA-40, and ERA-Interim reanalysis products are compared to sounding observations from an enhanced radiosonde network (11 sites, every 6 h) during the Tibetan Plateau Experiment (TIPEX) conducted
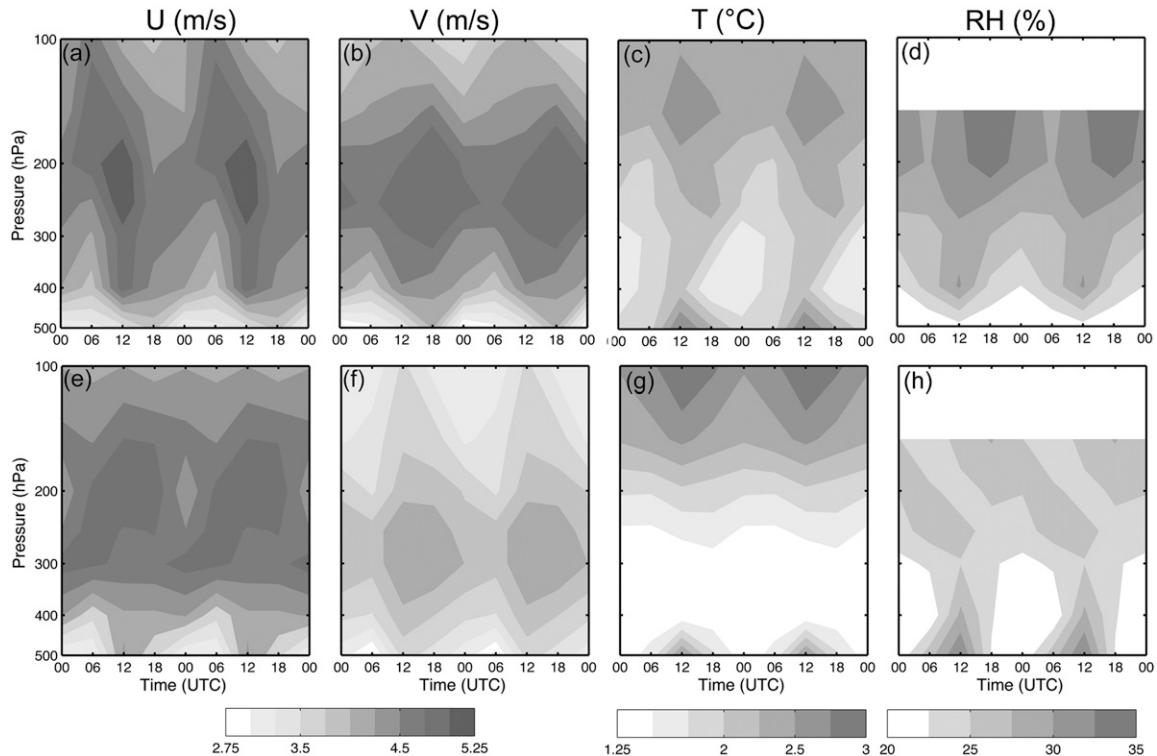
FIG. 5. As in Fig. 4, but for the RMS errors.

from 10 May to 9 August 1998. These more than 3000 soundings at 11 stations are independent of the reanalyses because only those that are not assimilated in any of the reanalyses are used for verification. It is found that, averaged over the entire three-month period, each reanalysis dataset produces mean values consistent with the verifying soundings for temperature and horizontal winds (corresponding to relatively small mean bias), but with large differences (and thus biases) in relative humidity. On average, except for temperature at upper levels, both newer-generation reanalyses (CFSR and ERA-Interim) have smaller RMS error and bias than their predecessors (NNRP and ERA-40), consistent with recent studies (e.g., Betts et al. 2009; Mao et al. 2010; Mooney et al. 2010; Hodges et al. 2011). With some exceptions, the RMS errors of all variables for both CFSR and ERA-Interim (verifying with soundings) are similar in magnitude to the RMS difference between these two reanalyses, and are approximately twice as large as the corresponding observation errors. This suggests that with a lack of independent high-quality verifying observations, the difference between two independent reanalyses can be used to approximate the analysis error of these reanalyses; the same cannot be generalized for estimating the mean bias because each reanalysis appears to have unique, albeit small, biases.

It is also found that there are strong diurnal variations in both RMS error and mean bias that differ greatly among different reanalyses and pressure levels. It is obvious that the diurnal variations in the mean bias may have contributed considerably to the diurnal variations in the RMS error. The reasons for the strong RMS error and bias, as well as their difference in diurnal variations are beyond the scope of the current study.

Despite the enhanced independent high-quality and high-frequency sounding observations we used, one must be cautious not to generalize the current error statistics to regions outside the mountainous Tibetan Plateau. It also remains unclear whether such error statistics can be generalized to other seasons or other climate regimes. Future studies will use both the sounding and reanalyses datasets to examine the regional-scale weather and climate processes over the Tibetan Plateau.

REFERENCES

Bao, X., F. Zhang, and J. Sun, 2011: Diurnal variations of warm-season precipitation east of the Tibetan Plateau over China. *Mon. Wea. Rev.,* **139,** 2790–2810.

Betts, A. K., M. Köhler, and Y. Zhang, 2009: Comparison of river basin hydrometeorology in ERA-Interim and ERA-40 reanalysis with observations. *J. Geophys. Res.,* **114,** D02101, doi:10.1029/2008JD010761.

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, doi:10.1007/s00376-010-9170-8.

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.

Ding, Y., and J. C. L. Chan, 2005: The East Asian summer monsoon: An overview. *Meteor. Atmos. Phys.,* **89,** 117–142.

Fan, X., J. E. Walsh, and J. R. Krieger, 2008: A one-year experimental Arctic reanalysis and comparisons with ERA-40 and NCEP/NCAR reanalyses. *Geophys. Res. Lett.,* **35,** L19811, doi:10.1029/2008GL035110.

Frauenfeld, O. W., T. Zhang, and M. C. Serreze, 2005: Climate change and variability using European Centre for Medium-Range Weather Forecasts reanalysis (ERA-40) temperatures on the Tibetan Plateau. *J. Geophys. Res.,* **110,** D02101, doi:10.1029/2004JD005230.

He, H., and F. Zhang, 2010: Diurnal variations of warm-season precipitation over Northern China. *Mon. Wea. Rev.,* **138,** 1017–1025.

Hodges, K. I., R. W. Lee, and L. Bengtsson, 2011: A comparison of extratropical cyclones in recent reanalyses ERA-Interim, NASA MERRA, NCEP CFSR, and JRA-25. *J. Climate,* **24,** 4888–4906.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

Mao, J., X. Shi, L. Ma, D. P. Kaiser, Q. Li, and P. E. Thornton, 2010: Assessment of reanalysis daily extreme temperatures with China's homogenized historical dataset during 1979–2001 using probability density functions. *J. Climate,* **23,** 6605–6623.

Mooney, P. A., F. J. Mulligan, and R. Fealy, 2010: Comparison of ERA-40, ERA-Interim and NCEP/NCAR reanalysis data with observed surface air temperatures over Ireland. *Int. J. Climatol.,* **31,** 545–557, doi:10.1002/joc.2098.

Saha, S., and Coauthors, 2010: The NCEP climate forecast system reanalysis. *Bull. Amer. Meteor. Soc.,* **91,** 1015–1057.

Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.,* **131,** 2961–3012.

Wang, A., and X. Zeng, 2012: Evaluation of multireanalysis products with in situ observations over the Tibetan Plateau. *J. Geophys. Res.,* **117,** D05102, doi:10.1029/2011JD016553.

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

Zhao, T., and C. Fu, 2006: Comparison of products from ERA-40, NCEP-2, and CRU with station data for summer precipitation over China. *Adv. Atmos. Sci.,* **23,** 593–604.

Zhou, M., X. Xu, L. Bian, J. Chen, H. Liu, Y. Zhang, S. Li, and Y. Zhao, 2000: *Observations, Diagnostics, and Dynamics of the Atmospheric Boundary Layer over the Tibetan Plateau during TIPEX 1998* (in Chinese). China Meteorological Press, 125 pp.