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, Pennsylvania State University, University Park, Pennsylvania

Submitted to *Journal of Climate*
Accepted, 19 June 2012

Corresponding author address: Dr. Fuqing Zhang, Department of Meteorology, The Pennsylvania State University, University Park, PA 16802
E-mail: fzhang@psu.edu
Abstract

The NCEP/NCAR, NCEP/CFSR, ERA-40, and ERA-Interim reanalysis products are evaluated with sounding observations from an enhanced radiosonde network available every 6 hours 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 (that were not assimilated in any of the reanalyses), which represents for the first time 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 (NCEP/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 largest and highest plateau with an average elevation of over 4,500 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 Forecast (ECMWF) provide four widely used reanalysis datasets: the NCEP/NCAR Reanalysis Project (NNRP), the NCEP Climate Forecast System Reanalysis (CFSR), the ECMWF Reanalysis (ERA-40), and the ECMWF Interim Reanalysis (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 (Betts et al., 2009, Fan et al., 2008, Mao et al., 2010, Mooney et al., 2010, Zhao et al., 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 literature. Given the importance
of the Tibetan Plateau in the regional and global weather and climate while the scarce 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 rawinsondes observations from extended field experiments. Such assessments will also have directive 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 intense observing period (IOP) of TIPEX from 10 May to 9 August 1998. Over the three-month IOP, sounding observations were collected every 6 hours or 4 times per day at 11 locations covering a broad region of the Tibetan Plateau (Figure 1). For completely independent verifications, we exclude the 00 and 12 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, dew point
depression, wind direction, and wind speed) at 7 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 inter-comparison 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), cautions 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 hours. 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, and covers the period from mid-1957 to 2001 (including some ECMWF reanalysis ERA-15 data for 1979-1993). To produce analyses every six hours, the three-dimensional variational technique (3D-Var) was applied using the T159 (~125
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 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.

Figure 2a-d shows 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 and Figure 2e-h shows the corresponding standard deviations of these variables. Figure 2i-l shows the mean biases of the four reanalyses while Figure 2m-p shows the corresponding root-mean square (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 at 500 hPa to a peak of ~20 m/s at 200 hPa. The standard deviations of U (V) from five datasets increase from ~4.5 m/s (4 m/s) at 500 hPa to be peaked near the jet maximum (150-200 hPa) with a maximum of ~14 m/s (9 m/s) (Figures 2e-2f). Broadly speaking, the vertical profiles of the mean and standard derivation for U and V in all four reanalysis products closely follow those averaged over the verifying sounding...
observations indicating 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 throughout the vertical column (Figures 2i-2j). 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) (Figures 2i-2j).

The RMS error of U from the interpolated reanalysis verified against the soundings (Figure 2m) is the smallest for CFSR and ERA-Interim (<3.5 m/s at 500hPa and ~4.5 m/s above), slightly higher for ERA-40 (with a maximum of ~5m/s at 300hPa), and clearly the largest for NNRP (from ~4.2 m/s at 500 hPa to a maximum of ~6 m/s at 200 hPa). On average the NNRP RMS error of U is about 1m/s larger than those of CFSR and ERA-Interim. The RMS error of V (Figure 2n) is clearly the smallest for ERA-Interim (<4 m/s 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 ~1 m/s 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 (Figures 2m-2n) are nearly twice as large as the NCEP default observational error for radiosondes (also shown in Figures 2m-2n). 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 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 derivation of temperature of each reanalysis (Figure 2c,g) is hardly distinguishable from the verifying sounding mean throughout the vertical layers. The vertical profiles of standard deviation of temperature from the sounding and four reanalysis are very similar with each other with the maximum ~5°C around 300 hPa and the minimum 2 °C around 150 hPa (Figure 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 (Figure 2k). The ERA-40 has the least negative overall bias (~ -0.7°C averaged over all levels) whereas CFSR has the most negative bias (~ -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 (~ -1.9 °C). On the other hand, the NNRP is or nearly is the least negatively biased at top levels (between ~ -0.6 and -0.9°C at or above 250 hPa) but becomes the most negatively biased at 400 and 500 hPa (~ -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.

---

1 Refer to http://www.nco.ncep.noaa.gov/pmb/codes/nwprod/sorc/hwrf_v3.fd/var/obsproc/obserr.txt
For the relative humidity (RH), the 3-month mean of each reanalysis differs greatly and from
one another and from the verifying observed sounding mean (which decreases from ~57% at 500
hPa to ~40% at 150 hPa, Figure 2d). All reanalyses are more humid than the mean sounding
observations at lower levels but become drier than observations at higher levels (Figure 2l). In
terms of mean bias (Figure 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 ERA40 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 (Figure 2h). In comparison, the
difference of standard derivations of RH among different reanalyses is much smaller.

The RMS errors of RH for all reanalyses (Figure 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 weakness 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 (Figures 2m-2p). 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 (Figure 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, Toetoche and Nyingchi) between 90°E 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 Figure 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 (Figure 3 vs. Figure 2i-2l). 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 (Figure
Another example is the mean bias of temperature in ERA-40, which has the peak cold bias above 250 hPa in the western and central plateau but the larger cold bias is below 250 hPa in the eastern region. Overall, the 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 details 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 Figures 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 (Figure 4a) has a predominant diurnal negative peak of < -2.5 m/s at 06 UTC (14 BST or Beijing Standard Time) above 250 hPa but at the same time the lower-level bias is at its minimum (~ -0.3 to +0.2 m/s at 400-500 hPa). The low-level mean U-wind bias has a positive peak of ~ 0.5 at 18 UTC (02 BST) and a negative peak of ~ -0.5 m/s at 12 UTC (20 BST). The mean V-wind bias (Figure 4b) also has a different diurnal cycle at different pressure levels: at upper levels, the positive bias peaks at 12 UTC (20 BST) whereas the negative bias peaks at 18 UTC (02 BST). At lower levels, the positive bias of the V-wind peaks from 18 UTC at 500hPa to 00 UTC at 300-400 hPa. The peak negative bias of T (Figure 4c) occurs mostly at 12 UTC (20 BST) except for at 250hPa, where the peak is at 18 UTC (02 BST). A secondary negative peak occurs at 00 UTC (08 BST) at 150hPa. At lower levels, the weakest cold bias is centered at 00 UTC (08 BST). For the RH (Figure 4d), there is a predominant peak wet (positive) bias at lower levels during the daytime (00-12 UTC or 08-20 BST) and a peak dry (negative) bias at upper levels at 18 UTC (02 BST).

For ERA-Interim, there is also a strong diurnal variation of mean bias in that differs from variable to variable and between pressure levels (Figures 4e-4h). The mean U-wind bias (Figure 4e) has the negative diurnal peak at different levels among different times. The negative bias peaks at 12 UTC (20 BST) below 400 hPa, at 06 UTC (14 BST) for 200-300 hPa and at 00 UTC (08 BST) above 200 hPa; there is a minute positive bias (~0.2 m/s) at 06 and 18 UTC (14 BST and 02 BST) at 500 hPa. For the V-wind (Figure 4f) below 300 hPa, there is a negative peak at 06 UTC (14 BST) and a positive peak at 18 UTC (02 BST). At 100-200 hPa, the negative diurnal peak occurs at 18 UTC (02 BST). Consistent with CFSR, the cold bias of T (Figure 4g) in ERA-Interim also has a general diurnal peak at 12 UTC (20 BST) at all levels, and a relative
minimum at 00 UTC (08 BST). Also similar to CFSR, the wet bias of ERA-Interim (Figure 3h) generally peaks during the daytime (06-12 UTC; 14-20 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 (Figures 4c, 4d, 4g, and 4h).

There are also strong diurnal variations in the RMS error for both CFSR and ERA-Interim for different variables at different levels (Figure 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 (Figures 5a and 5e), the maximum RMS error in both reanalyses can exceed 4-5 m/s from 300 to 100 hPa at the diurnal peak times. For V (Figures 5b and 5f), the RMS error in both reanalyses has a diurnal peak at 18 UTC (02 BST) at almost all pressure levels with a maximum at around 250 hPa – this is not the case for the corresponding mean bias (Figures 4b and 4f). For T (Figures 5c and 5g), both reanalyses generally have a diurnal peak in RMS error (and cold/negative bias) at 12 UTC (20 BST) at all levels. For RH (Figures 5d and 5h), the maximum RMS errors in both reanalyses have peaks at 12 UTC (20 BST) at lower levels. At higher levels, the RMS error for CFSR shifts forward in time to peak at 18 UTC (02 BST) at 150-200hPa; the RMS error for ERA-Interim gradually shifts backward in time from peaking at 06 UTC (14 BST) at 250 hPa to 18 UTC (02 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 hours) during the Tibetan Plateau Experiment (TIPEX) conducted 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 can not 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 between 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.

Acknowledgments: We are grateful to Prof. Xiangde Xu at the Chinese Academy of
Meteorological Sciences for providing us the quality-controlled TIPEX radiosonde observations
that made this study feasible. Insightful review comments by three anonymous reviewers and
proofreading by Wei Li and Benjamin Green on earlier versions of the manuscript are also greatly
appreciated. This study is sponsored by NSF Grants 0840651 and 0904635.
References:


**Figure Captions**

**Figure 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.

**Figure 2.** Top row: vertical profiles of mean (a) U, (b) V, (c) T, and (d) RH averaged over all independent TIPEX IOP soundings during 10 May – 09 August 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 row: 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.

**Figure 3.** Top panels: 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 during 10 May – 09 August 1998, and the corresponding mean interpolated from the four reanalysis products. Middle panels: 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°E and 96°E (Lhasa, Nagqu, Nyingchi, and Toetoehe). Bottom panels: 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).

**Figure 4.** The diurnal variations of the mean biases of U, V, T, and RH for CFSR (top panels) and for ERA-Interim (bottom panels) at different pressure levels verifying against the TIPEX soundings.

**Figure 5.** The diurnal variations of the RMS errors of U, V, T, and RH for CFSR (top panels) and for ERA-Interim (bottom panels) at different pressure levels verifying against the TIPEX soundings.
Table 1: Correlation coefficients between each reanalysis and the sounding observations for different variables at different vertical levels.

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

<table>
<thead>
<tr>
<th>Pressure</th>
<th>NNRP</th>
<th></th>
<th></th>
<th></th>
<th>ERA-40</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>U</td>
<td>V</td>
<td>T</td>
<td>RH</td>
<td>U</td>
<td>V</td>
<td>T</td>
<td>RH</td>
<td></td>
</tr>
<tr>
<td>100 hPa</td>
<td>0.90</td>
<td>0.83</td>
<td>0.98</td>
<td></td>
<td>0.93</td>
<td>0.88</td>
<td>0.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>150 hPa</td>
<td>0.91</td>
<td>0.88</td>
<td>0.98</td>
<td></td>
<td>0.95</td>
<td>0.89</td>
<td>0.98</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>200 hPa</td>
<td>0.89</td>
<td>0.82</td>
<td>0.96</td>
<td></td>
<td>0.93</td>
<td>0.86</td>
<td>0.97</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>250 hPa</td>
<td>0.86</td>
<td>0.78</td>
<td>0.95</td>
<td></td>
<td>0.88</td>
<td>0.81</td>
<td>0.96</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>300 hPa</td>
<td>0.81</td>
<td>0.75</td>
<td>0.93</td>
<td>0.74</td>
<td>0.83</td>
<td>0.76</td>
<td>0.96</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>400 hPa</td>
<td>0.68</td>
<td>0.67</td>
<td>0.90</td>
<td>0.66</td>
<td>0.72</td>
<td>0.69</td>
<td>0.93</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>500 hPa</td>
<td>0.55</td>
<td>0.58</td>
<td>0.83</td>
<td>0.55</td>
<td>0.62</td>
<td>0.66</td>
<td>0.88</td>
<td>0.63</td>
<td></td>
</tr>
</tbody>
</table>
Figure 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.
Figure 2. Top panels: vertical profiles of mean (a) U, (b) V, (c) T, and (d) RH averaged over all independent TIPEX IOP soundings during 10 May – 09 August 1998, and the corresponding mean interpolated from the four reanalysis products. Second panels: 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 panels: 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 panels: 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.
Figure 3. Top panels: 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 during 10 May – 09 August 1998, and the corresponding mean interpolated from the four reanalysis products. Middle panels: 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°E and 96°E (Lhasa, Nagqu, Nyingchi, and Toetoehe). Bottom panels: 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).
Figure 4. The diurnal variations of the mean biases of U, V, T, and RH for CFSR (top panels) and for ERA-Interim (bottom panels) at different pressure levels verifying against the TIPEX soundings.
Figure 5. The diurnal variations of the RMS errors of U, V, T, and RH for CFSR (top panels) and for ERA-Interim (bottom panels) at different pressure levels verifying against the TIPEX soundings.