Estimation of Uncertainty in Temperature Observations Made at Meteorological Stations Using a Probabilistic Spatiotemporal Approach*

ARTICLE *in* JOURNAL OF APPLIED METEOROLOGY AND CLIMATOLOGY · JUNE 2014

Impact Factor: 2.54 · DOI: 10.1175/JAMC-D-13-0179.1

4 AUTHORS, INCLUDING:

Maogui Hu
State Key Laboratory of Resources and E…
24 PUBLICATIONS **180** CITATIONS

Qingxiang Li
China Meteorological Administration
41 PUBLICATIONS **957** CITATIONS

Assessment of Reanalysis Daily Extreme Temperatures with China's Homogenized Historical Dataset during 1979–2001 Using Probability Density Functions

JIAFU MAO AND XIAOYING SHI

Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee

LIJUAN MA

National Climate Center, China Meteorological Administration, Beijing, China

DALE P. KAISER

Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee

QINGXIANG LI

China Meteorological Administration, Beijing, China

PETER E. THORNTON

Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee

(Manuscript received 16 December 2009, in final form 16 June 2010)

ABSTRACT

Using a recently homogenized observational daily maximum (T_{MAX}) and minimum temperature (T_{MIN}) dataset for China, the extreme temperatures from the 40-yr ECMWF Re-Analysis (ERA-40), the Japanese 25-year Reanalysis (JRA-25), the NCEP/Department of Energy Global Reanalysis 2 (NCEP-2), and the ECMWF's ERA-Interim (ERAIn) reanalyses for summer (June–August) and winter (December–February) are assessed by probability density functions for the periods 1979–2001 and 1990–2001. For 1979–2001, no single reanalysis appears to be consistently accurate across eight areas examined over China. The ERA-40 and JRA-25 reanalyses show similar representations and close skill scores over most of the regions of China for both seasons. NCEP-2 generally has lower skill scores, especially over regions with complex topography. The regional and seasonal differences identified are commonly associated with different geographical locations and the methods used to diagnose these quantities. All the selected reanalysis products exhibit better performance for winter compared to summer over most regions of China. The T_{MAX} values from the reanalysis tend to be systematically underestimated, while T_{MIN} is systematically closer to observed values than T_{MAX}. Comparisons of the reanalyses to reproduce the 99.7 percentiles for T_{MAX} and 0.3 percentiles for T_{MIN} show that most reanalyses tend to underestimate the 99.7 percentiles in maximum temperature both in summer and winter. For the 0.3 percentiles in T_{MIN}, NCEP-2 is relatively inaccurate with a $-12°C$ cold bias over the Qinghai–Tibetan Plateau in winter. ERA-40 and JRA-25 generally overestimate the extreme T_{MIN}, and the extreme percentage differences of ERA-40 and JRA-25 are quite similar over all of the regions. The results are generally similar for 1990–2001, but in contrast to the other three reanalysis products the newly released ERAIn is very reasonable, especially for wintertime T_{MIN}, with a skill score greater than 0.83 for each region of China. This demonstrates the great potential of this product for use in future impact assessments on continental scales where those impacts are based on extreme temperatures.

Corresponding author address: Jiafu Mao, Oak Ridge National Laboratory, MS6301, P.O. Box 2008, Oak Ridge, TN 37831-6301.
E-mail: maoj@ornl.gov

DOI: 10.1175/2010JCLI3581.1

1. Introduction

Extreme weather and climatic events such as intense heat or cold waves, severe storms, and prolonged droughts are responsible for a disproportionately large part of the climate-related damage to human society and the natural environment (Meehl et al. 2000; Parmesan 2006). There is little doubt that the frequency and intensity of certain types of extreme events will increase with the elevated concentrations of anthropogenic greenhouse gases and expected climate warming (Trenberth et al. 2007). Research on temperature extremes in the twentieth century has been carried out extensively, and on the daily time scale all studies show patterns of changes in extremes consistent with an overall warming (Solomon et al. 2007). For the land areas sampled, the distributions of maximum temperature (T_{MAX}) and minimum temperature (T_{MIN}) have not only shifted to higher values, but the cold extremes have warmed more than the warm extremes (Alexander et al. 2006). Due to the considerable regional variations in the trends and increased concerns regarding regional-scale impacts, there is a great demand for improving our understanding of the regional features and causes of extreme temperature events within the context of global change (Griffiths et al. 2005; Alexander and Arblaster 2009). Like most of the world over the past few decades, China has experienced varying degrees of regional warming, and there was an overall tendency of greater warming rates for minimum and winter temperatures as compared to maximum and summer temperatures (Zhai et al. 1999; Yan et al. 2002; Zhai and Pan 2003; Gong et al. 2004; Liu et al. 2006; Qian and Qin 2005; Fu et al. 2008).

One effective way of gaining a better understanding of the extreme temperatures and their relation to other climate factors is through analysis of high quality data products with consistent physical processes at suitable temporal and spatial resolutions (Griffiths et al. 2005; Vose et al. 2005). Observations, especially global-scale observed daily datasets, remain somewhat problematic due to gaps in temporal continuity and geographical coverage. However, various reanalysis products, based on previously observed climate data, provide consistent, long-term gridded meteorological datasets using modern, state-of-the-art data assimilation systems developed for numerical weather prediction (Kalnay et al. 1996). They play a crucial role in quantifying and understanding atmospheric features, seasonal prediction, and extreme events, while also helping to assess the ability of climate models to simulate the average climate and its variations. Furthermore, they can be used in identifying deficiencies in representations of physical processes that produce climate model errors (Uppala et al. 2005, 2008; Onogi et al. 2007).

Although reanalysis data are the most reliable atmospheric analysis data, they are of course not without inaccuracies. Systematic biases and uncertainties in the climatological variables (especially in surface fluxes), and poor estimates of climatological variations and trends, limit the ability of these products to represent the real state of the weather and climate (Trenberth and Olson 1988; Bosilovich et al. 2008). Therefore, comparison of the capabilities and limitations of current reanalysis datasets with observations from independent meteorological stations will be of value for determining the best uses of our current reanalysis products for scientific and practical purposes (Smith et al. 2001; Ma et al. 2009; Zhao and Fu 2009). The latest global or regional evaluations provide detailed assessments of the strengths and weaknesses of reanalysis products based principally on monthly, seasonal, and annual time scales (Kanamitsu et al. 2002; Uppala et al. 2005; Onogi et al. 2007; Ma et al. 2008, 2009; Zhao and Fu 2006, 2009). Given that atmospheric conditions on time scales of days have direct impacts on human health and human activities, an assessment of the capacity of reanalysis to represent conditions on daily or finer time scales is clearly valuable (Li et al. 2005; Pitman and Perkins 2009).

Daily extreme temperatures at 1000 hPa and 2-m height from three global reanalysis datasets were first comprehensively evaluated by Pitman and Perkins (2009) between the reanalyses and over regions where daily observed T_{MAX} and T_{MIN} were available. They highlighted the widespread differences between the selected reanalyses due to the different methods used to diagnose these quantities. However, because high quality, land-based observations were not always available, their evaluation was primarily limited to intercomparisons between the reanalyses on global and continental scales. Furthermore, the analysis and comparisons at seasonal time scales, which may help isolate the mechanisms leading to these differences, were not stressed, especially over East Asia. Applications of extreme temperatures from reanalysis products in China are not new. For example, Zhou et al. (2004) used monthly T_{MAX} and T_{MIN} from surface stations and the National Centers for Environmental Prediction (NCEP)/Department of Energy (DOE) Global Reanalysis 2 (NCEP-2) over southern China to analyze differences in air temperature trends to explore the impacts of urbanization on south China temperatures. Gong et al. (2006) used the daily T_{MAX} and T_{MIN} of NCEP-2 from 1979 to 2000 to evaluate the robustness of the "weekend effect" from the observations. In these studies, however, the reanalyzed extreme temperatures are commonly used as if they were observations. Few studies have directly evaluated the ability of the reanalysis to represent the extreme temperatures, especially on daily time scales. Furthermore, traditional assessments of historical reanalysis products such as pressure, precipitation, humidity and winds, 2-m surface temperature, and radiation fluxes over China are based principally on monthly, seasonal, or longer average time scales (Xu et al. 2001; Shi et al. 2006; Ma et al. 2008, 2009; Zhao and Fu 2006, 2009). While valid, this tends to hide biases or systematic errors that are identifiable in the daily data (Pitman and Perkins 2009).

Some questions therefore arise: To what extent can we trust the quality of the present daily extreme temperatures from the latest reanalysis over continental China? Are there more useful metrics and higher quality ground-based observations capable of better assessing the extremes? A probability density function (PDFs) derived skill score proposed by Perkins et al. (2007) and homogenized extreme temperature data for all of China (Li et al. 2009b) provide us with an unprecedented opportunity to directly compare multiple reanalyzed trends with observed trends in extremes. The skill score is a useful means of measuring the skill of daily reanalysis products or model outputs using the observed dataset. It was found to be robust against data limitations and to be a clear and straightforward way of comparing the entire modeled and observed dataset (Perkins et al. 2007; Perkins and Pitman 2008; Maxino et al. 2007; Pitman and Perkins 2009). In China, a newly developed higher quality daily surface air temperature dataset has become available to the scientific community. Compared to previous products, it has a more dense observational network and more consistent observing practices with comprehensive error corrections and quality control (Li et al. 2006; Ma et al. 2008; Li et al. 2009a,b).

This paper focuses mainly on the reliability assessment over continental China of daily T_{MAX} and T_{MIN} derived from four latest-generation reanalysis products such as NCEP-2, the 40-year European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40), the Japanese 25-year Reanalysis (JRA-25), and the ECMWF reanalysis data from the ERA Interim project (ERAIn). We use PDF skill scores to assess the seasonal performance of selected reanalyses in terms of their capacity to represent the summer and winter daily extreme temperatures for the periods 1979–2001 and 1990–2001. We also compare the four reanalyses' capacities to represent the 99.7th percentile for T_{MAX} (T_{MAX99}) and the 0.3 percentile for T_{MIN} (T_{MIN03}), which are extreme values and may typically be observed perhaps once per year (Pitman and Perkins 2009). Possible reasons for the differences between the four products, and potential mechanisms behind the biases compared to the observations, will also be discussed.

The remainder of this paper explains our methodology (section 2), describes our results (section 3), discusses our results (section 4), and states our conclusions (section 5).

2. Methodology

a. Data sources

1) HOMOGENIZED HISTORICAL TEMPERATURE DATASET OF CHINA

Raw observed data used for long-term climate research may be affected by inhomogeneities, which include changes in instrument exposure, changes in observing time, and station relocations. Hence, homogenization is considered to be an important part of the quality control process (Vincent et al. 2002). Since the 1950s, the China Meteorological Administration (CMA) has recorded the daily T_{MAX} and T_{MIN} at its meteorological stations, which have greatly increased in number, especially since the mid-1970s. Released by the CMA in December 2006, the China Homogenized Historical Temperature (CHHT) dataset (1951–2004), version 1.0, data are first examined for internal consistency and homogeneity (Li et al. 2004, 2006, 2009b). This dataset was developed using observations of the daily mean, T_{MAX}, and T_{MIN} from a total of 731 national weather stations distributed throughout China. Figure 1 shows the locations of the stations with topography. The density of the stations is lower in the sparsely populated high-mountain and desert areas of west and northwest China, and higher in eastern and, especially, southeastern China, which has experienced rapid urbanization and dramatic economic growth since 1978.

The 24-h days for T_{MAX} and T_{MIN} are from 2000 Beijing time (BT, which is 8 h earlier than UTC) on the previous day to 2000 BT on the current day. The daily T_{MAX} and T_{MIN} are the true maxima and minima of the day, and are recorded by different instruments. A basic logic verification performed on daily extreme temperatures in the CHHT1.0 dataset includes a daily maximum air temperature that must be higher than or equal to any values recorded at the regular observing times and a daily maximum minus daily minimum air temperature that must be less than or equal to 24°C (Li et al. 2009b), by convention of the World Meteorological Organization (WMO 1993). CHHT1.0 has been widely used in climate change detection in China and has greatly improved the accuracy of regional and local climate trends (Zhou et al. 2004; Zhai et al. 2004; Jones et al. 2008; Ma et al. 2008; Li et al. 2009a).

2) REANALYSIS DATA

The reanalyses data used in this paper were obtained as follows. The NCEP-2, described by Kanamitsu et al. (2002) and Roads (2003), was downloaded from the National Oceanic and Atmospheric Administration/Earth Systems Research Laboratory (NOAA/ERSL) Web site (http://www.esrl.noaa.gov/psd/data/gridded/data.

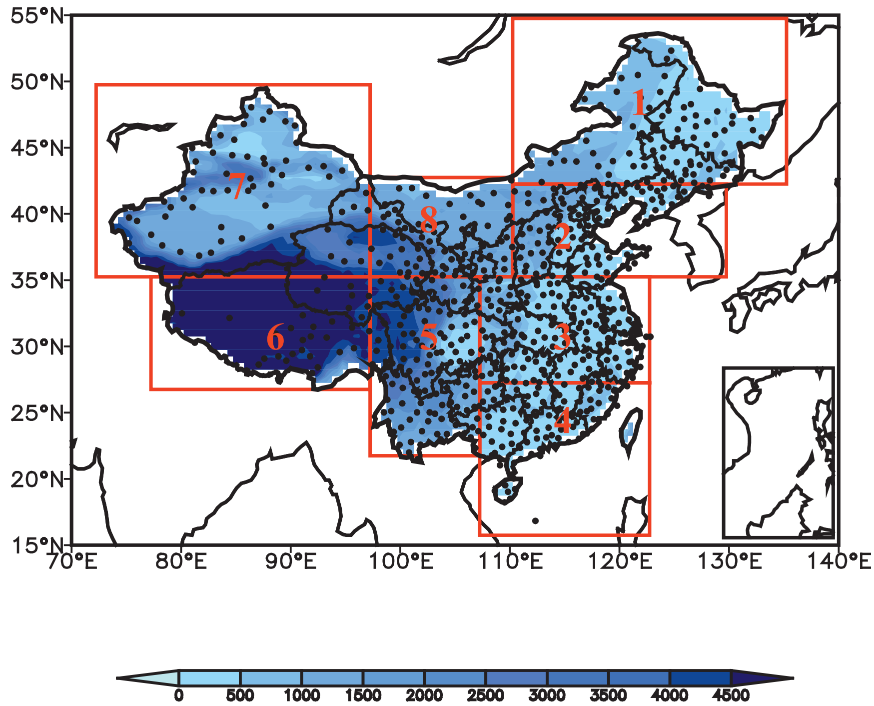

FIG. 1. Locations of the stations in China with their topography (m) and regionalization.

ncep.reanalysis2.html) for 2-m air temperature. This dataset corrects known errors in the NCEP–National Center for Atmospheric Research (NCAR) 40-Year Reanalysis Project (NCEP-1; Kalnay et al. 1996; Kistler et al. 2001), and can serve as a basic verification for the second Atmosphere Model Intercomparison Project (AMIP-II). The original data covers January 1979 to the present, and were sourced at a resolution of approximately 1.875° × 1.875° (T62 Gaussian grid points). The ERA-40 (Uppala et al. 2005) data provided a very high quality reference atmosphere state from September 1957 to August 2002. The daily 2-m air temperature was sourced from ECMWF (information online at http://www.ecmwf.int/research/era/do/get/era-40) on a 2.5° × 2.5° latitude–longitude grid. The JRA-25 has been released recently for general use (Onogi et al. 2007). The 2-m temperatures were sourced from the Japan Meteorological Agency Web site (http://jra.kishou.go.jp/JRA-25/index_en.html) and were available at 2.5° × 2.5°. In 2006, ECMWF started to develop a new interim reanalysis system derived from the latest version of their operational system. This system provides a bridge between ECMWF's previous reanalysis, ERA-40 (1957–2002), and the next-generation extended reanalysis envisaged at ECMWF (Simmons et al. 2007a,b; Uppala et al. 2008). The ERAIn reanalysis starts in 1989 and is available up to the present. It was downloaded from the ECMWF Web site (http://data-portal.ecmwf.int/data/d/interim_daily/) at a resolution of 1.5° × 1.5°.

b. Data processing and PDF skill scores

Each of the NCEP-2, ERA-40, JRA-25, and ERAIn daily datasets contained four records per day. The original 6-hourly, 2-m air temperature outputs are for 0000, 0600, 1200, and 1800 UTC, which correspond to 0800, 1400, 2000, and 0200 (the following morning) BT. Hence, for comparisons with the observed dataset, the 1800 (the previous day), 0000, 0600, and 1200 UTC datasets (0200, 0800, 1400, and 2000 BT) were selected, and the highest value was defined as T_{MAX} and the lowest was defined as T_{MIN}.

Limited by the common time period covered by the observations and all of the reanalyses, the basic comparisons will be performed over the period 1979–2001 in this paper. To further assess ERAIn reanalysis data together with NCEP-2, ERA-40 and JRA-25 data, evaluations will be also conducted over the time period 1990–2001. Since the resolutions of the reanalyzed grid boxes do not match each other, the daily extreme temperatures from the reanalyzed grid were all transformed to a finer 0.5° × 0.5° latitude–longitude grid using bilinear interpolation. In terms of the observed dataset, we employed an improved inverse distance weighting (IDW) method, which considers the influences of latitude, longitude, and elevation differences between input and output grids, to convert the station data into the 0.5° × 0.5° regular grid (Ma et al. 2008).

TABLE 1. Latitude and longitude boundaries for all eight regions over China, with the region names.

Region No.	Lat (°N)	Lon (°E)	Region name
1	42.25–54.75	110.25–135.25	Northeast China
2	35.25–42.25	110.25–129.75	North China
3	27.25–35.25	107.25–122.75	Jianghuai
4	15.75–27.25	107.25–122.75	South China
5	21.75–35.25	97.25–107.25	Southwest China
6	26.75–35.25	77.25–97.25	Tibetan Plateau
7	35.25–49.75	72.25–97.25	West of northwest China
8	35.25–42.75	97.25–110.25	East of northwest China

In evaluating the results of the daily T_{MIN} and T_{MAX} from reanalyses over continental China, we used a measure of skill proposed by Perkins et al. (2007). The skill score calculates the cumulative minimum value of two distributions of each binned value in a given PDF, thereby measuring the common area between two PDFs. The common area between two PDFs equals 1.0 for a perfect match and 0.0 where the two PDFs are independent (Perkins et al. 2007; Pitman and Perkins 2009). It thus provides a very simple but useful metric of the amount of overlap between an observed and a reanalyzed PDF, which allows a comparison across the whole PDF distribution (Perkins et al. 2007; Perkins and Pitman 2008; Maxino et al. 2007; Pitman and Perkins 2009).

For the purpose of performing a quantitative comparison in the next section, the Chinese mainland is divided into eight subareas primarily based on administrative divisions and the characteristics of the monsoon climate of China (Shi and Xu 2007; Fig. 1, Table 1). All observed and reanalyzed extreme temperatures within each area were concatenated and then used to derive the PDF.

3. Results

a. Temporal and spatial comparisons

To make a general assessment of the reanalysis data against ground-based measurements across China, the annual cycles of the zonal-averaged daily reanalysis extreme temperatures for 1979–2001 and 1990–2001 are first examined. As there are similar spatial and temporal patterns of ERA-40, JRA-25, and NCEP-2 for the two periods, we mainly choose the period between 1990 and 2001 for analysis in this section.

Figure 2 shows the time–latitude cross sections of the mean annual cycle from daily observed and reanalysis extreme temperatures over China. Figure 3 shows the differences between the reanalyses and the observations. In general, the latitudinal daily changes in the reanalysis extreme temperatures agree well with the features reflected in the observations for different zones (Fig. 2). The reanalysis datasets can reproduce the two separate high temperature centers in the midlatitudes (40°–45°N) and the low latitudes (19°–23°N) during the summer season (days 160–240), and the two cold regions during the winter season. Compared to the observational dataset, the reanalysis products, while exhibiting similar latitudinal daily patterns, tend to systematically underestimate the T_{MAX}, particularly over latitudes between 30° and 40°N where the Qinghai–Tibetan Plateau (QTP) is located (Figs. 3a, 3c, 3e, and 3g). In terms of the T_{MIN}, the errors are smaller than those of the T_{MAX} (Figs. 3b, 3d, 3f, and 3h), and except for the latitude bands near the QTP, the reanalyzed T_{MIN} values are closer to or a bit higher than the observations. Among the four reanalysis datasets, the JRA-25 T_{MIN} has the smallest differences over most of the latitudes and seasons (Fig. 3d), and the NCEP-2 has the most obvious seasonal variations in the errors (Fig. 3f). The ERA-40 and ERAIn T_{MIN} differences with the observations reveal weaker seasonal variations and show better consistency in comparison with JRA-25 and NCEP-2 (Figs. 3b and 3h).

b. T_{MAX}

Figure 4 shows the summer PDFs of T_{MAX} during 1979–2001 over eight regions of continental China. With the exception of QTP (region 6, Fig. 4f), the observations indicate that for all other regions the highest probability for T_{MAX} is about 30°C. Southeastern China (region 4) has the highest probability of T_{MAX} exceeding 32°C (Fig. 4d). Over region 6, the probability distributions of ERA-40, JRA-25, and NCEP-2 shift to the left, and show clear differences in skill (0.43 for ERA-40, 0.57 for JRA-25, and 0.32 for NCEP-2). All three reanalyses most closely resemble the observations in region 8 (Fig. 4h), and the skill score is more than 0.8 (Fig. 8a). Generally, the ERA-40 and JRA-25 data exhibit similar representations and have close skill scores over most of the regions for summer, while the NCEP-2 is anomalous compared to the other two products, especially over regions 4–6, where it shows lower skill. For winter during 1979–2001, northeastern China has a high probability of T_{MAX} below 0°C (region 1, Fig. 5a), and southern China (region 4) has the highest T_{MAX} probability, around 20°C (Fig. 5d). All three products have similar descriptions over most regions and the skill scores of regions 1, 2, 3, 4, 5 and 8 are all greater than 0.7 (Fig. 6b). Compared to ERA-40 and JRA-25, NCEP-2 has lower skill scores over region 6 (the area with the highest topography; Fig. 6b). For this region, however, all three products perform better for winter than summer, and each of them has a higher skill score (Figs. 4f, 5f, and 6a,b). The skill scores of ERA-40 and JRA-25 over regions 1–6 are higher for winter than summer, and the winter skill scores for NCEP-2 over regions 3, 4, 5, and 7 are also higher than

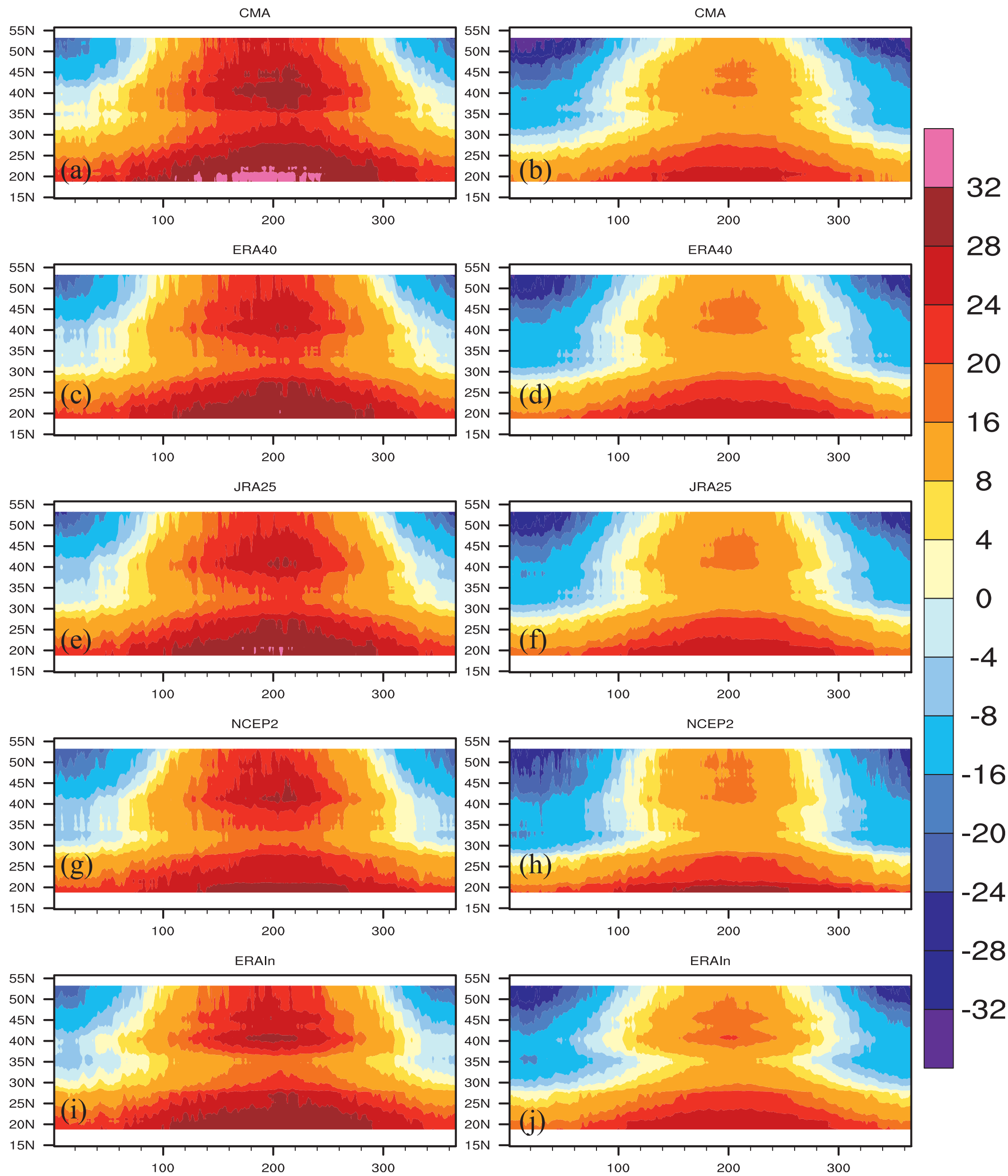

FIG. 2. The 1990–2001 annual cycle of zonally averaged daily maximum temperature (°C) for (a) observations, (c) ERA-40, (e) JRA-25, (g) NCEP-2, and (i) ERAIn; and minimum temperature (°C) for (b) observations, (d) ERA-40, (f) JRA-25, (h) NCEP-2, and (j) ERAIn.

those during the summer. For north-central China (region 8), the skill scores of the three products are all lower during winter than summer (Figs. 4h, 5h, and 6a,b).

Due to the addition of the ERAIn product since 1989, we selected the period from 1990 to 2001 for further comparisons. PDF distributions of ERA-40, JRA-25, and NCEP-2 for 1990–2001 are quite similar to those of 1979–2001 for both June–August (JJA) and December–February (DJF) (figures omitted). For JJA, ERAIn has the highest skill score over regions 3, 4, 7, and 8, and the second highest score over regions 1, 5, and 6 (Fig. 6c). In contrast with the other three products, for DJF ERAIn shows the highest levels of skill in regions 1, 3, 4, 5, 7, and 8, and has the second highest scores over the remaining regions (regions 2 and 6; Fig. 6d).

In general, all four reanalysis products exhibit better performance in winter than summer for most of the regions during 1990–2001 (Figs. 6c and 6d). These products

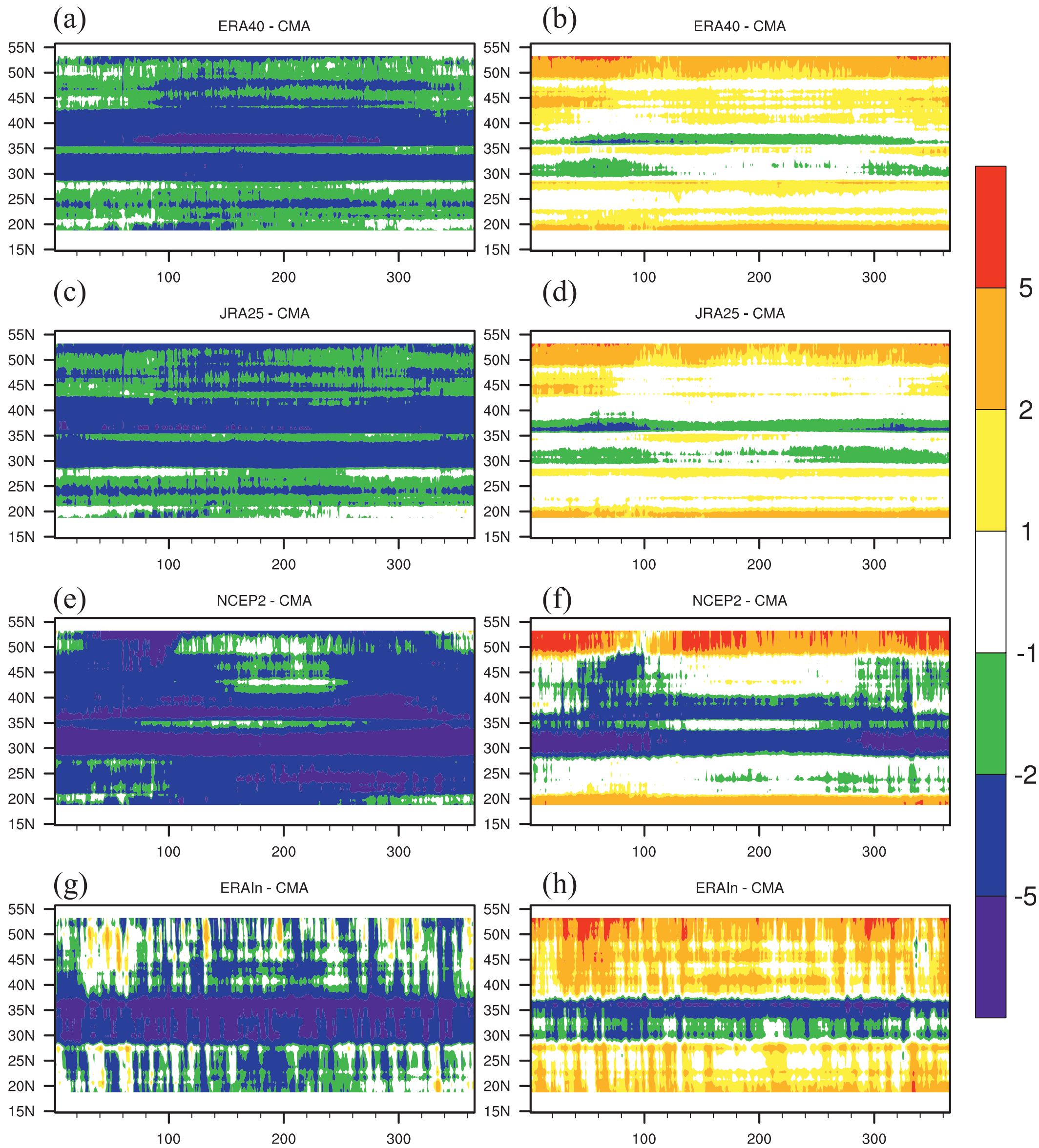

FIG. 3. The 1990–2001 annual cycle of zonally averaged daily maximum temperature (°C) differences between (a) ERA-40, (c) JRA-25, (e) NCEP-2, (g) ERAIn, and the observations; and the daily minimum temperature (°C) differences between (b) ERA-40, (d) JRA-25, (f) NCEP-2, (h) ERAIn, and the observations.

show systematic underestimations of the T_{MAX} over most regions for JJA and DJF, which is consistent with results shown in Figs. 2 and 3. Spatially, these errors are evident over regions 5–7 (Figs. 4e–g and 5e–g), and are likely related to the more complex terrain and less dense observations (Fig. 1).

c. T_{MIN}

Figure 7 shows the probability density functions of T_{MIN} for each region for JJA over 1979–2001. In contrast to T_{MAX} (Fig. 4), a general feature of the T_{MIN} analysis is a systematically stronger similarity to the observations, which is again consistent with previous findings (Figs. 2 and 3). Very tight observed PDFs (regions 1, 2, 3, 4, 5, and 8) are well captured by all the reanalysis products, with overlap statistics exceeding 0.75 (Figs. 7 and 9a). In regions 6 and 7, the reanalysis products tend to overestimate the probability of colder T_{MIN}. It is clear from Figs. 7f and 7g that NCEP-2 is poorer in these two regions for T_{MIN}, with overlap statistics of 0.55 and 0.67 (Fig. 9a). Figure 8 presents the DJF PDFs of the observed and reanalyzed T_{MIN} for each region from 1979 to 2001. Overall, most reanalysis products describe the PDF of observed T_{MIN} well. ERA-40 and JRA-25 capture the changes in location

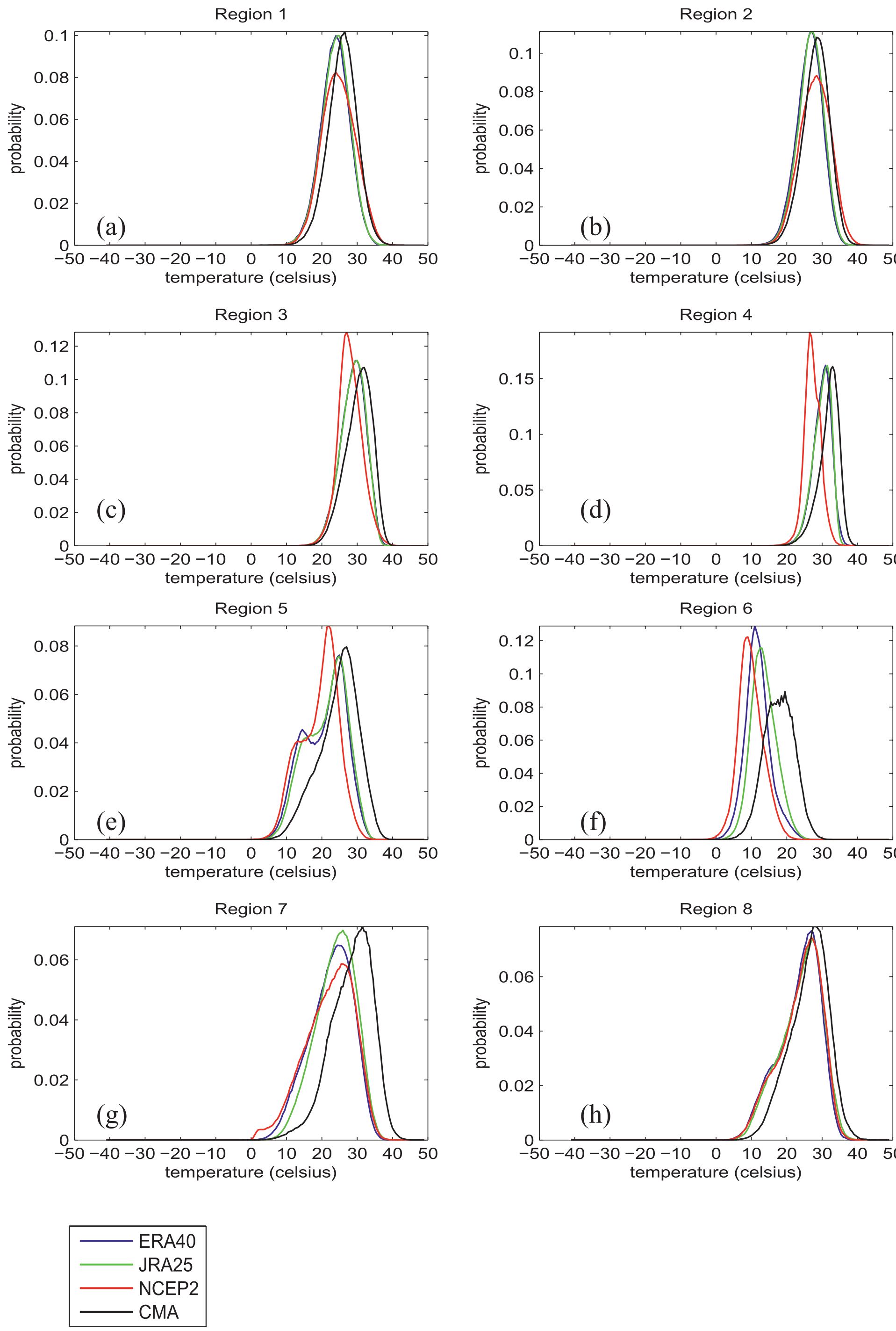

FIG. 4. JJA PDFs (1979–2001) of maximum temperature for regions 1–8.

(with respect to the x axis) and shape of the observed PDFs between most regions, and the skill scores are greater than 0.83 for all eight regions (Fig. 9b). Contrasted with the other reanalysis representations, NCEP-2 is highly competitive in regions 1, 2, 4, 7, and 8. However, there is a tendency toward underestimation in regions 3, 5, and, especially, 6 (Fig. 8). As with T_{MAX} during 1990–2001, there is a suggestion that most of the reanalysis can capture the changing shape of the T_{MIN} PDF with season (PDF distributions are omitted; Figs. 9c and 9d), and at least visually the shapes of the models' PDFs for DJF match with the observed data better than that for JJA.

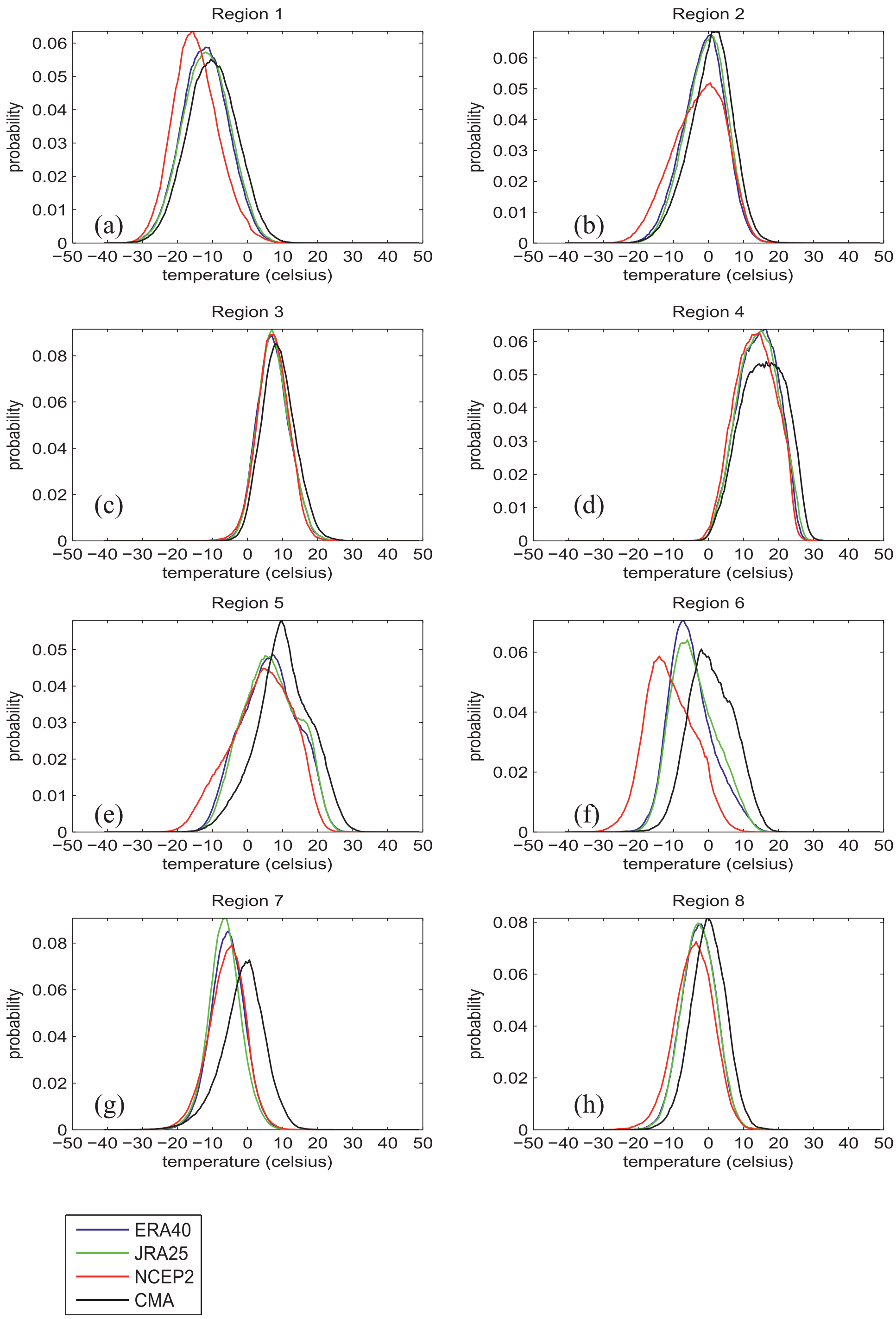

FIG. 5. As in Fig. 4, but for DJF.

During 1990–2001, the order of the scores for ERA-40, JRA-25, and NCEP-2 for both seasons is almost the same as that between 1979 and 2001 (Fig. 9). In contrast with the other three reanalysis products, ERAIn is very competitive over both seasons, especially for DJF, with a skill score greater than 0.83 for each region (Fig. 9d).

d. T_{MAX99} *and* T_{MIN03}

We used T_{MAX99} and T_{MIN03} as additional measures of how well the reanalysis could describe these very rare values that are not easily interpreted from the figures showing the PDFs. This provides one way of discriminating

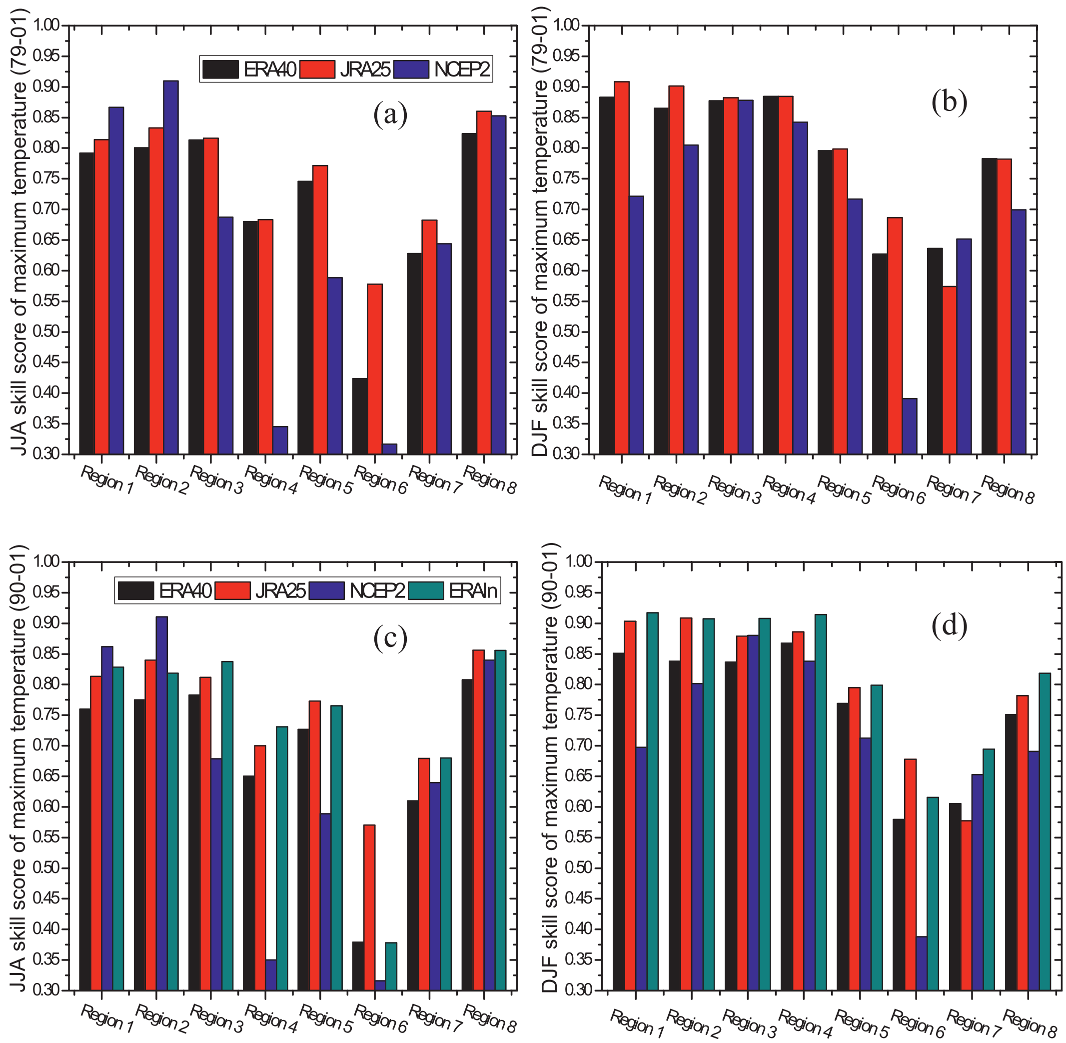

FIG. 6. (left) JJA and (right) DJF PDFs skill scores of maximum temperature for (a),(b) 1979–2001 and (c),(d) 1990–2001 across the eight regions of China.

between reanalyses in representing more extreme values that are important in impact assessments (Maxino et al. 2007). Figures 10a and 10c show the capacity of the reanalyses to produce the 99.7 percentiles for T_{MAX} in different seasons during 1979–2001 and 1990–2001. The reanalyses tend to underestimate extreme T_{MAX} for JJA and DJF over all eight regions, although NCEP-2 is an exception over regions 1–3 during JJA (Fig. 10a). Compared to ERA-40 and JRA-25, NCEP-2 has lower extreme T_{MAX} for all seasons over regions 4–6, and the percentile differences of ERA-40 and JRA-25 are quite similar over all regions (Fig. 10a). For 1990–2001, all products again underestimate the warmest temperatures in summer and winter over most regions, and there is little sense that there is a variation in the seasonal skill of the reanalysis (Fig. 10c). For all regions and both seasons, only ERAIn appears to be performing well, especially for JJA. Results of NCEP-2 during JJA, however, are impressively close to the observations over regions 1 and 3 (Fig. 10c). Percentage differences over region 8 are evident; JRA-25 being 4.8°C warmer for DJF and NCEP-2 being 8.2°C colder for JJA (Fig. 10c).

The results for T_{MIN} are quite variable (Figs. 10b and 10d). NCEP-2 does relatively poorly, with lower estimates of T_{MIN} for most regions. This underestimation is even obvious over the QTP (region 6) during DJF, with a -12°C cold bias (Fig. 10b). ERA-40 and JRA-25 generally overestimate the extreme T_{MIN}, and there is a maximum of 7°C difference for JRA-25 over region 7 during DJF (Fig. 10b). Percentage differences of T_{MIN} for ERA-40, JRA-25, and NCEP-2 during 1990–2001 are close to those during 1979–2001 (Figs. 10c and 10d). NCEP-2 is excessively cold and the other three reanalyses are warmer than the observed data. As with T_{MAX}, extreme T_{MIN}

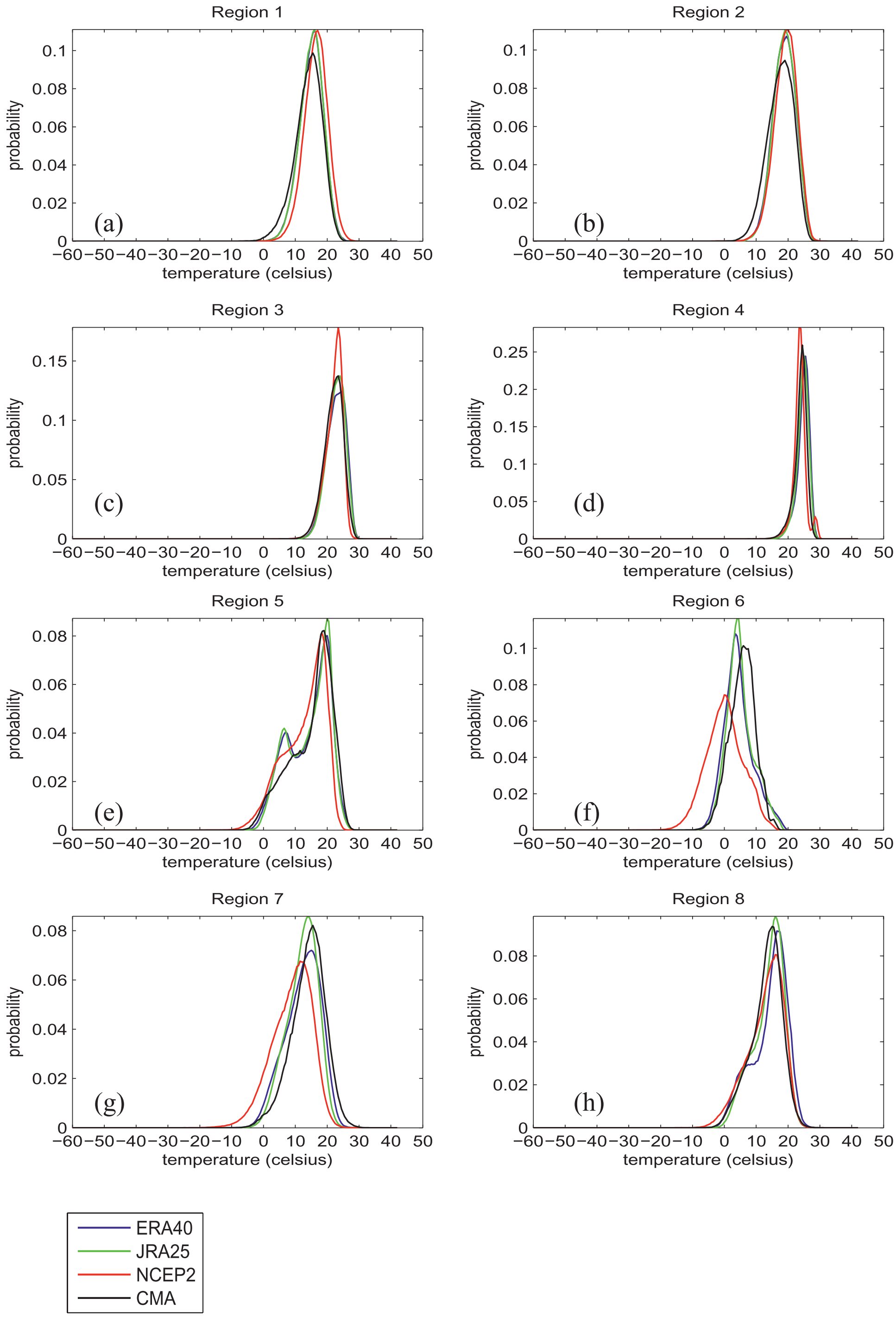

FIG. 7. JJA PDFs (1979–2001) of minimum temperature for regions 1–8.

differences of ERAIn are impressively close to the observations in both seasons over all of the regions (Fig. 10d).

4. Discussion

One advantage of the PDF-based criteria used here is that it is comparable between reanalyses. It therefore provides a means for ranking the reanalyses variable by variable or overall by averaging over regions (Perkins et al. 2007). Tables 2 and 3 show the rankings of the reanalysis products for T_{MAX} and T_{MIN} over all regions of China. For 1979–2001, in terms of T_{MAX}, the best to worst performance rankings are JRA-25, ERA-40, and NCEP-2 both

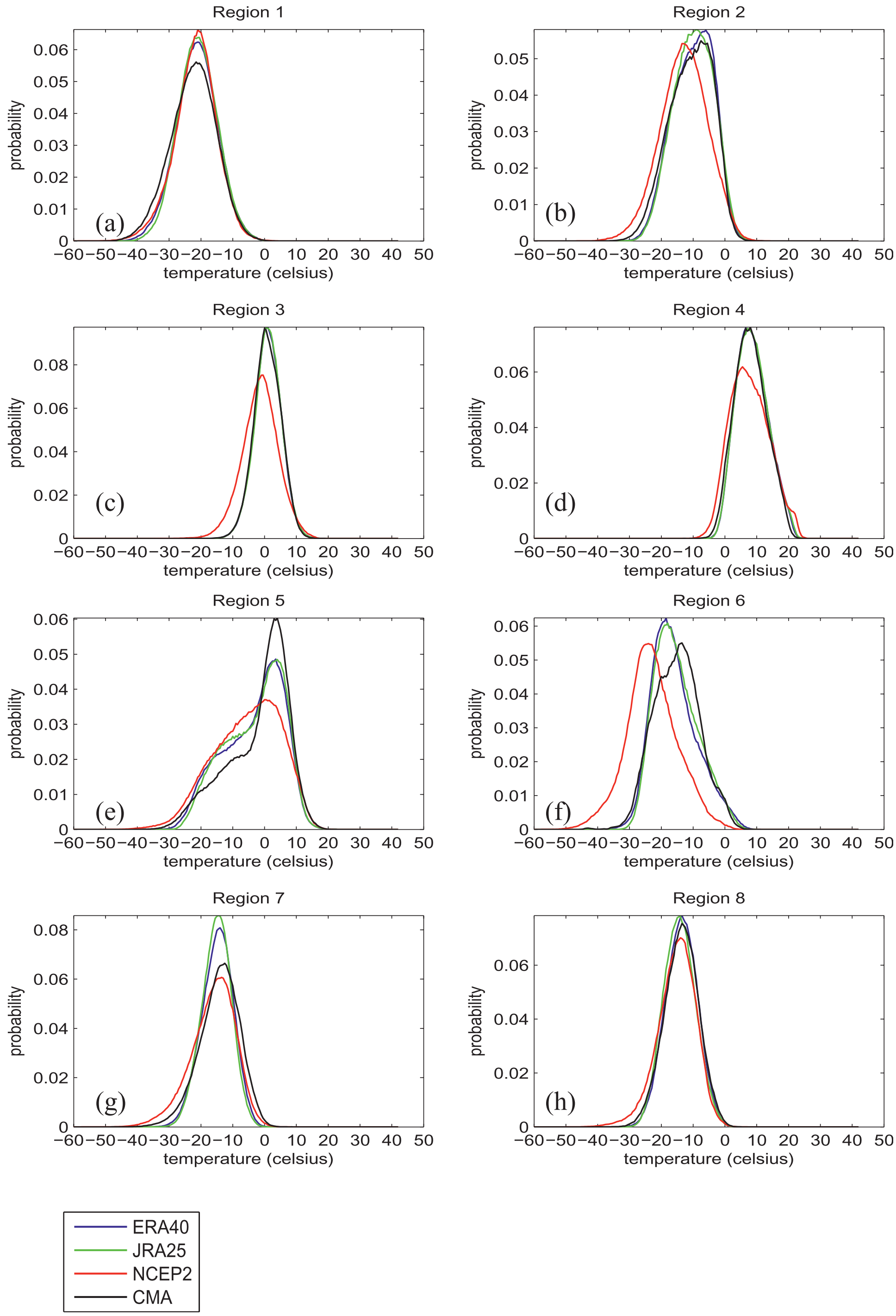

FIG. 8. As in Fig. 7, but for DJF.

for JJA and DJF. The average ranks for the whole summer and winter seasons follow the same order. Skill scores for T_{MIN} from each product are higher than those for T_{MAX} in both seasons, and the best result for JJA is JRA-25, while ERA-40 has the best performance in winter. Each reanalysis product for T_{MIN} has a skill score above 0.8 averaged over China and, as for T_{MAX}, they are, in order, JRA-25, ERA-40, and NCEP-2. During 1990–2001, the order of T_{MAX} in Table 3 is ERAIn, JRA-25, ERA-40, and NCEP-2 for both winter and the seasonal

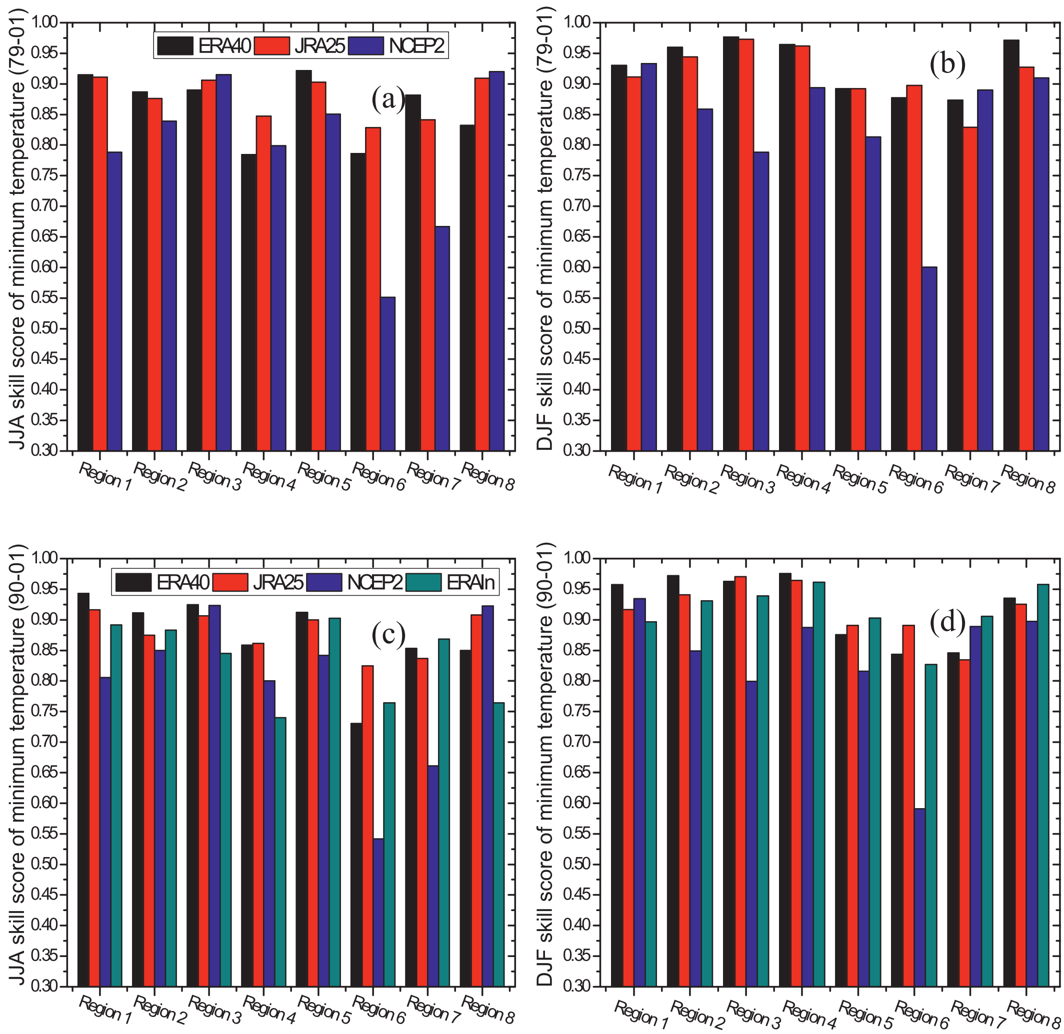

FIG. 9. As in Fig. 6, but for minimum temperature.

average. In terms of T_{MIN}, the best performer for the seasonal average is JRA-25, followed by ERA-40, ERAIn, and NCEP-2. Like the 1979–2001 rankings in Table 2, each reanalysis performs better for winter than summer for both T_{MAX} and T_{MIN}. The skill score of T_{MIN} for each product is higher than that of T_{MAX} over both seasons.

The better performance for DJF than JJA is similar to the conclusions of Zhao and Fu (2009), when they calibrated the reanalysis performance on 2-m temperature over China. They showed that levels of the 2-m temperature quality displayed by ERA-40, NCEP-2, and JRA-25 are usually better in winter than summer in China according to climate mean, interannual variation and variability, and climate trends after 1979. Climate variables including precipitation, temperature, and extreme temperatures over China are strongly influenced by the East Asian monsoon (Ding 2004). In winter, the climate is mostly cold and dry, and extreme temperature variations, particularly in northern China, are strongly sensitive to Siberian high activity (Qian and Qin 2005; Gong et al. 2006). Gong et al. (2006) showed that the mean strength and position of the Siberian high experienced an abrupt change in the 1980s, which coincided with an increase in winter temperature, particularly in northern China. The warmer tendency in the mean temperature has been driven by more high-temperature events and fewer low-temperature events. Therefore, the higher consistency of the reanalyzed extreme temperatures in DJF implies that the extremely high- and low-temperature climate and weather features over China could be better produced by the reanalysis products for DJF than for JJA.

Regional and seasonal analyses all suggest that JRA-25, ERA-40, and ERAIn show marked similarities in the PDFs with observations, while NCEP-2 appears anomalous compared to the other three reanalyses (Figs. 2, 3, 6, and 9; Tables 2 and 3). Causes of the discrepancies among

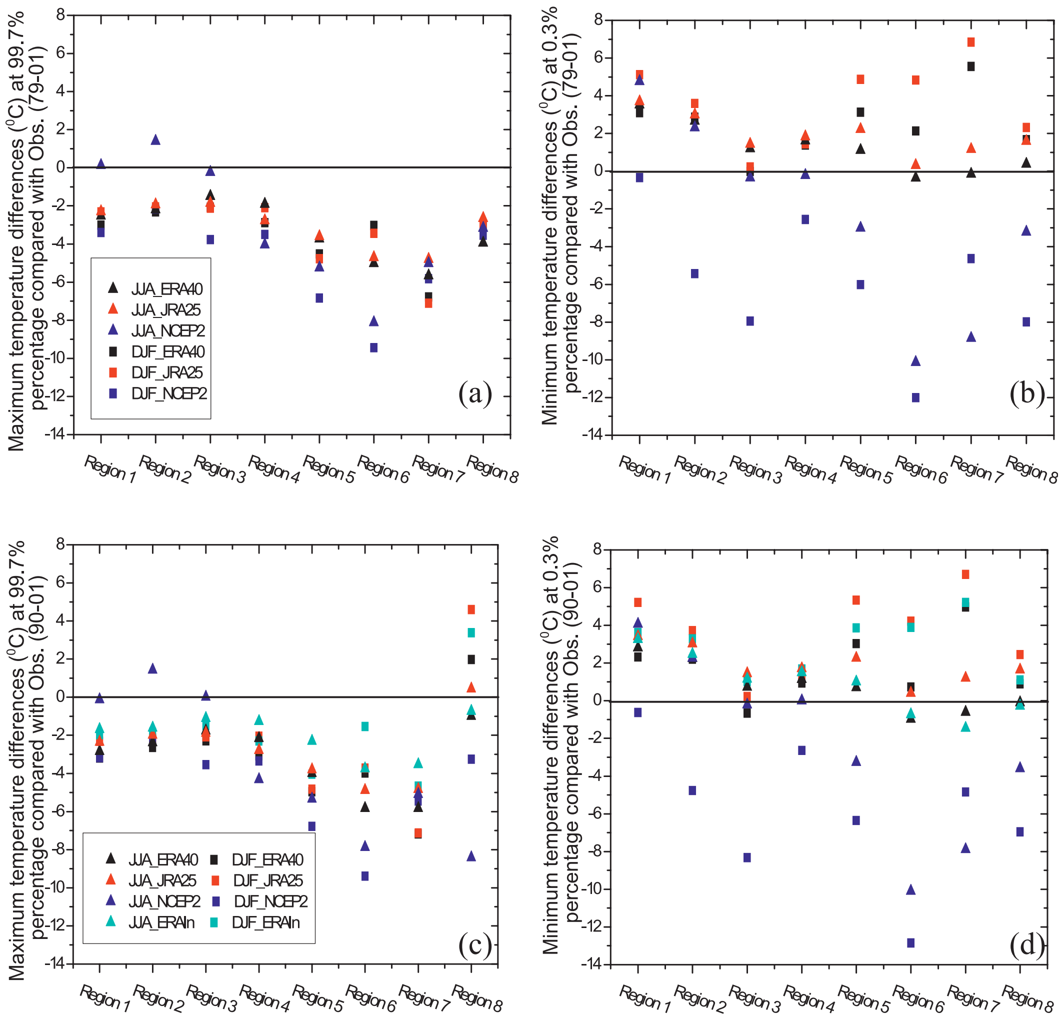

FIG. 10. The 99.7% percentile differences of maximum temperature and 0.3% percentile differences of minimum temperature compared to the observations for regions 1–8: (a) 1979–2001 T_{MAX}, (b) 1979–2001 T_{MIN}, (c) 1990–2001 T_{MAX}, and (d) 1990–2001 T_{MIN}.

these four reanalyses may include different assimilations, parameterization processes, the observations they adopt, and differing numbers of observational stations used. The performance of JRA-25 is especially impressive over the QTP (region 6), the region with the highest elevation and the longest snow-covered period in China (Figs. 2d, 3d, 6, and 9). Air temperatures from JRA-25 at the 2-m level are obtained by a two-dimensional optimum interpolation between the lowest model level and the surface, assimilated with ground-based temperatures (Onogi et al. 2007). The land surface temperatures, however, are quite sensitive to the existence of snow at high latitudes and in areas of high elevation; thus, it is important to incorporate consistent snow depth analyses for the assimilation system of the reanalysis. Compared to other reanalyses, JRA-25 first assimilates the digitized Chinese daily snow depth data over continental China (Onogi et al. 2007). As a result, this fusion contributed to improved snow depth analysis around the snow edges and, thus, improved forecasts of surface 2-m temperatures. NCEP-2 diagnoses the 2-m air temperature as a function of surface skin temperature, lowest atmospheric model temperature, vertical stability, and surface roughness (Pitman and Perkins 2009). Although it has more up-to-date physics and corrections for the known errors in NCEP1, NCEP- 2 in our study is still not as good over China as ERA-40, JRA-25, and ERAIn. Ma et al. (2008) argued that despite these improvements, compared with NCEP-1, the NCEP-2 system still has poor internal coherence. We hence point out here that the NCEP-2 reanalysis over China could be further improved in terms of the extreme temperatures. ERAIn produced relatively high skill scores over different regions during 1990–2001 for both JJA and DJF (Figs. 6c,d and 9c,d). In the case of the

TABLE 2. Ranking of reanalyses for T_{MAX} and T_{MIN} over all regions for 1979–2001.

	T_{MAX}						T_{MIN}					
	JJA	Rank	DJF	Rank	Total	Rank	JJA	Rank	DJF	Rank	Total	Rank
ERA-40	0.713	2	0.794	2	0.754	2	0.862	2	0.931	1	0.896	2
JRA-25	0.755	1	0.802	1	0.778	1	0.878	1	0.917	2	0.897	1
NCEP-2	0.651	3	0.713	3	0.682	3	0.791	3	0.836	3	0.814	3

percentiles, there is also considerable agreement between ERAIn and the observations for T_{MIN03} and T_{MAX99} (Figs. 10c and 10d). Built with an improved assimilation background model and additional observation data, the new ERAIn reanalysis is envisaged to eliminate or reduce several problems experienced in ERA-40. These objectives have been largely achieved as a result of a combination of factors, including many model improvements, the use of four-dimensional variational analysis, a revised humidity analysis, the use of variational bias correction for satellite data, and other improvements in data handling (Berrisford et al. 2009). Therefore, the notable description of extreme temperatures for ERAIn over China demonstrates great potential for future applications of this reanalysis on regional climatic change and assessment.

All four reanalyses tend to show a shift toward lower T_{MAX} both for JJA and DJF (Figs. 2 and 3, 4 and 5, and 10a and 10c). This cold bias of T_{MAX} is especially obvious over regions 5–7, which have a lower density of stations and more complex topography than in the other regions in China (Fig. 1). With regard to these underestimations of the reanalysis compared with meteorological observations, our findings are consistent with Ma et al. (2008). In their work, they assessed the correspondence of reanalysis air temperatures from ERA-40, NCEP1, and NCEP-2 with ground-based measurements from China. They illustrated that for China as a whole, climatologies for ERA-40, NCEP1, and NCEP-2 air temperatures are lower than the observations by $-1.13°$, $-2.34°$, and $-2.06°C$ for JJA, and by $-0.68°$, $-2.61°$, and $-2.11°C$ for DJF, respectively, during 1979–2001. Large negative differences for most of western China, where the terrain is complex, primarily contribute to this cool bias. In our results, however, we did not find the systematic underestimation of T_{MIN} for the whole of China. We also did not notice that the difference over region 6 for T_{MIN} is more evident than that for T_{MAX} over the same time period. China is a country with marked orographic gradients, and the topography is high in the west and generally low in the east. To explore the sources of these temperature differences, especially the errors for T_{MAX}, we corrected all the reanalyses for differences in the topography following the approach of Zhao et al. (2008), and we then selected the period 1990–2001 for analysis. Figure 11 shows the skill score differences in JJA and DJF T_{MAX} and T_{MIN} for each region with and without elevation correction. Generally, the elevation correction has a more variable influence over different regions for JJA T_{MAX} than DJF T_{MAX}. For JJA, higher skill scores were seen over regions 4–6, while lower scores were found over regions 7 and 8 after the correction. The improvement for high-elevation region 6 is most significant (Fig. 11a); while for DJF, only region 6 has a higher score and the differences for other regions are not evident (Fig. 11b). For T_{MIN}, three of the reanalyses (ERA-40, JRA-25, and NCEP-2) have lower scores over region 6, and all the reanalyses showed poorer results over region 7 for JJA (Fig. 11c). For DJF, there is a constant tendency toward poorer performance of T_{MIN} for all the reanalyses over most of the selected regions (Fig. 11d). Our results therefore suggest that errors, especially the underestimation of T_{MAX} between the reanalyzed and observed extremes, cannot be ascribed solely to the effects of elevation correction. The uniform use of elevation correction for the reanalyzed extreme temperatures over China should be carefully considered. Regional and seasonal differences for each extreme temperature should be fully taken into account when removing elevation-induced bias in the reanalysis extremes for their specific area of interest.

TABLE 3. Ranking of reanalyses for T_{MAX} and T_{MIN} over all regions for 1990–2001.

	T_{MAX}						T_{MIN}					
	JJA	Rank	DJF	Rank	Total	Rank	JJA	Rank	DJF	Rank	Total	Rank
ERA-40	0.686	3	0.762	3	0.724	3	0.873	2	0.921	1	0.897	2
JRA-25	0.755	1	0.801	2	0.778	2	0.879	1	0.917	2	0.898	1
NCEP-2	0.648	4	0.708	4	0.678	4	0.793	4	0.833	4	0.813	4
ERAIn	0.737	2	0.822	1	0.779	1	0.832	3	0.915	3	0.874	3

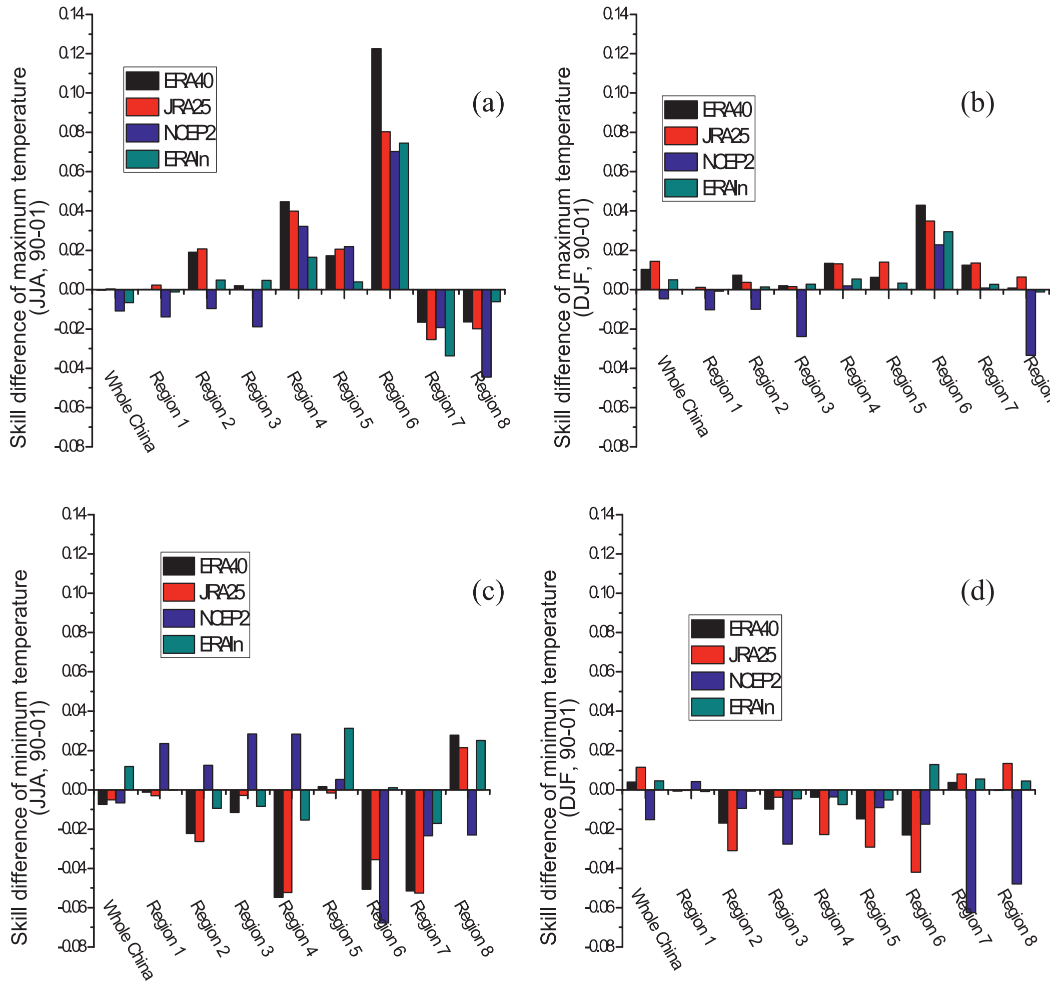

FIG. 11. Skill differences of (top) maximum and (bottom) minimum temperatures with and without elevation correction for regions 1–8: (a),(b) 1990–2001 T_{MAX} for JJA and DJF and (c),(d) 1990–2001 T_{MIN} for JJA and DJF.

We note that T_{MIN} tends to be better represented by the reanalyses than T_{MAX} in the sense that Tables 2 and 3 show more reanalyzed T_{MIN} with skill scores >0.8 both for JJA and DJF. The analysis and intercomparison of the extreme temperatures has the potential to disentangle the influence of surface solar and thermal radiation on global warming whereas an analysis of the mean temperature alone does not (Dai et al. 1999). The variables T_{MAX} and T_{MIN} are produced through quite different processes. Here, T_{MAX} is strongly affected by insolation, which is in turn affected by factors such as cloud cover, surface albedo, and water vapor (Dai et al. 1999; Pitman 2003). The key problem reproducing T_{MAX} over China is that most reanalysis products underestimate the probability of high values (Figs. 2–5). Since T_{MAX} is strongly affected by insolation, this gives us an indication that the reanalysis may for some reason underestimate the surface insolation during 1979–2001 over China. Another possible explanation for the discrepancies in T_{MAX} might be the approximation using the four times per day temperature to derive temperature extremes (section 2b). As T_{MAX} and T_{MIN} from one of the four reanalysis (ERA-40) were not available from the public server, our daily extreme temperatures were approximately obtained by the use of T_{MAX} and T_{MIN} of the four temperature values per day, which is similar to the approaches used in Pitman and Perkins (2009) and Kharin et al. (2005). As a result, such sampled extreme temperatures are likely to be less intense than the true T_{MAX} and T_{MIN} represented by the reanalysis, particularly for the T_{MAX} over the land (Kharin et al. 2005).

Compared to T_{MAX}, T_{MIN} is less affected by the solar flux, which is only present during daylight. On the other hand, T_{MIN} is affected by thermal radiative exchanges, which would seem easier to reproduce than the complicated interactions involved in the production of the daily T_{MAX}. Therefore, the consistency of reanalyzed T_{MIN} with the observations most likely relates to the well-simulated mechanism of the atmosphere to absorb and reemit thermal radiation toward the surface during the night. Other reasons for the differences in performance for extreme temperatures, especially T_{MAX} of the reanalysis, might arise from the climatic effects of the land-use–land-cover (LULC) change and the increase in anthropogenic aerosols over China during the last three decades (Zhou et al. 2004; Huang et al. 2006; Li et al. 2009a). Weak parameterization of the boundary layer, simulation of surface soil moisture, and coupling of the land to the atmosphere might also limit the reanalysis models' skill in producing the 2-m air temperature (Koster et al. 2004; Pitman and Perkins 2009). However, identifying and quantifying these potential mechanisms are beyond the scope of this paper; we suggest further evaluations and sensitivity tests at different time scales may help improve the performance of reanalysis products on continental scales.

5. Conclusions

The agreement of the reanalysis surface extreme temperatures from ERA-40, NCEP-2, JRA-25, and ERAIn with homogenized daily air temperatures from meteorological stations is evaluated across China using PDF skill scores. The results indicate that the selected reanalysis products can largely reproduce the changes in location and shape of the observed PDFs across most regions of China although there were seasonal and regional differences. Our evaluations of extreme 2-m temperatures, while limited to one continent, indicate that the main reanalysis products perform better than expected. They show considerable skill at subcontinental scales when assessed using daily data during summer and winter. This certainly builds further confidence in the use of these products for regional simulations and impact assessments. However, we also note that some reanalyses show apparent discrepancies that need to be addressed and explored. Further intercomparisons with more reanalysis data like the recently released Modern Era Retrospective-analysis for Research and Applications (MERRA; Bosilovich et al. 2008), will provide unique opportunities for fostering a deeper understanding of the strengths and weaknesses of the reanalyses on different temporal and spatial scales. In terms of the observations, uncertainties in sparse stations in the west of China and interpolation schemes from scattered stations into regular grids might contribute to the limits of our validations. Perkins et al. (2007) showed that the PDFs were actually very robust with respect to the observations, given the amount of daily data that formed the PDFs. The analysis presented here provides additional insights into the climatological means and standard deviations commonly used in assessments on time scales longer than days.

Acknowledgments. This research is supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research (BER) programs, and was performed at Oak Ridge National Laboratory (ORNL). ORNL is managed by UT-Battelle, LLC, for the U.S. Department of Energy under Contract DE-AC05-00OR22725. NCEP-2 reanalysis data were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, from its Web site (available at http://www.cdc.noaa.gov/). The JRA-25 data were provided from the cooperative research project of the JRA-25 long-term reanalysis by the Japan Meteorological Agency (JMA) and the Central Research Institute of Electric Power Industry (CRIEPI). We are very grateful to ECMWF for providing access to the ERA-40 and ERAIn data. Special thanks are also given to Prof. Andy Pitman and Dr. Sarah Perkins from the University of New South Wales, and to Wilfred M. Post at ORNL for their helpful discussions.

REFERENCES

Alexander, L. V., and J. M. Arblaster, 2009: Assessing trends in observed and modelled climate extremes over Australia in relation to future projections. *Int. J. Climatol.,* **29,** 417–435.

——, and Coauthors, 2006: Global observed changes in daily climate extremes of temperature and precipitation. *J. Geophys. Res.,* **111,** D05109, doi:10.1029/2005JD006290.

Berrisford, P., D. P. Dee, K. Fielding, M. Fuentes, P. Kållberg, S. Kobayashi, and S. M. Uppala, 2009: The ERA-Interim archive. ECMWF Rep. Series, No. 1, Reading, United Kingdom, 16 pp. [Available online at www.ecmwf.int/publications.]

Bosilovich, M. G., J. Chen, F. R. Robertson, and R. F. Adler, 2008: Evaluation of global precipitation in reanalyses. *J. Appl. Meteor. Climatol.,* **47,** 2279–2299.

Dai, A., K. E. Trenberth, and T. R. Karl, 1999: Effects of clouds, soil moisture, precipitation, and water vapor on diurnal temperature range. *J. Climate,* **12,** 2451–2473.

Ding, Y. H., 2004: Seasonal march of the East-Asian summer monsoon. *East Asian Monsoon,* C. P. Chang, Ed., World Scientific, 3–53.

Fu, C., Z. Jiang, Z. Guan, J. He, and Z. Xu, 2008: Climate extremes and related disasters in China. *Regional Climate Studies of China,* C. Fu et al., Eds., Springer, 313–344.

Gong, D. Y., Y. Z. Pan, and J. A. Wang, 2004: Changes in extreme daily mean temperatures in summer in eastern China during 1955–2000. *Theor. Appl. Climatol.,* **77,** 25–37.

——, D. Guo, and C. H. Ho, 2006: Weekend effect in diurnal temperature range in China: Opposite signals between winter and summer. *J. Geophys. Res.,* **111,** D18113, doi:10.1029/2006JD007068.

Griffiths, G. M., and Coauthors, 2005: Change in mean temperature as a predictor of extreme temperature change in the Asia–Pacific region. *Int. J. Climatol.,* **25,** 1301–1330.

Huang, Y., R. E. Dickinson, and W. L. Chameides, 2006: Impact of aerosol indirect effect on surface temperature over East Asia. *Proc. Natl. Acad. Sci. USA,* **103,** 4371–4376.

Jones, P. D., D. H. Lister, and Q. Li, 2008: Urbanization effects in large-scale temperature records, with an emphasis on China. *J. Geophys. Res.,* **113,** D16122, doi:10.1029/2008JD009916.

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

Kanamitsu, M., W. Ebisuzaki, J. Woolen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP–DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc.,* **83,** 1631–1643.

Kharin, V. V., F. W. Zwiers, and X. Zhang, 2005: Intercomparison of near-surface temperature and precipitation extremes in AMIP-2 simulations, reanalyses, and observations. *J. Climate,* **18,** 5201–5223.

Kistler, R., and Coauthors, 2001: The NCEP–NCAR 50-Year Reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.,* **82,** 247–267.

Koster, R. D., and Coauthors, 2004: Regions of coupling between soil moisture and precipitation. *Science,* **305,** 1138–1140.

Li, H. B., A. Robock, S. Liu, X. G. Mo, and P. Viterbo, 2005: Evaluation of reanalysis soil moisture simulations using updated Chinese soil moisture observations. *J. Hydrometeor.,* **6,** 180–193.

Li, Q., X. Liu, H. Zhang, T. C. Peterson, and D. R. Easterling, 2004: Detecting and adjusting temporal inhomogeneity in Chinese mean surface air temperature data. *Adv. Atmos. Sci.,* **21,** 260–268.

——, and Coauthors, 2006: China homogenized historical temperature datasets (1951–2004). Version 1.0. Climate Data Center, National Meteorological Information Center, Beijing, China, 65 pp.

——, W. Li, P. Si, G. Xiaorong, W. Dong, P. Jones, J. Huang, and L. Cao, 2009a: Assessment of surface air warming in northeast China, with emphasis on the impacts of urbanization. *Theor. Appl. Climatol.,* **99,** 469–478, doi:10.1007/s00704-009-0155-4.

——, H. Zhang, J. Chen, W. Li, X. Liu, and P. D. Jones, 2009b: A mainland China homogenized historical temperature dataset of 1951–2004. *Bull. Amer. Meteor. Soc.,* **90,** 1062–1065.

Liu, X., Z. Yin, X. Shao, and N. Qin, 2006: Temporal trends and variability of daily maximum and minimum, extreme temperature events, and growing season length over the eastern and central Tibetan Plateau during 1961–2003. *J. Geophys. Res.,* **111,** D19109, doi:10.1029/2005JD006915.

Ma, L., T. Zhang, Q. Li, O. W. Frauenfeld, and D. Qin, 2008: Evaluation of ERA-40, NCEP-1, and NCEP-2 reanalysis air temperatures with ground-based measurements in China. *J. Geophys. Res.,* **113,** D15115, doi:10.1029/2007JD009549.

——, ——, O. W. Frauenfeld, B. Ye, D. Yang, and D. Qin, 2009: Evaluation of precipitation from the ERA-40, NCEP-1, and NCEP-2 reanalyses and CMAP-1, CMAP-2, and GPCP-2 with ground-based measurements in China. *J. Geophys. Res.,* **114,** D09105, doi:10.1029/2008JD011178.

Maxino, C. C., B. J. McAvaney, A. J. Pitman, and S. E. Perkins, 2007: Ranking the AR4 climate models over the Murray Darling Basin using simulated maximum temperature, minimum temperature and precipitation. *Int. J. Climatol.,* **28,** 1097–1112.

Meehl, G. A., and Coauthors, 2000: Trends in extreme weather and climate events: Issues related to modeling extremes in projections of future climate change. *Bull. Amer. Meteor. Soc.,* **81,** 427–436.

Onogi, K., and Coauthors, 2007: The JRA-25 Reanalysis. *J. Meteor. Soc. Japan,* **85,** 369–432.

Parmesan, C., 2006: Ecological and evolutionary responses to recent climate change. *Ann. Rev. Ecol. Evol. Syst.,* **37,** 637–669.

Perkins, S. E., and A. J. Pitman, 2008: Do weak AR4 models bias projections of future climate changes over Australia? *Climatic Change,* **93,** 527–558.

——, ——, N. J. Holbrook, and J. McAneney, 2007: Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. *J. Climate,* **20,** 4356–4376.

Pitman, A. J., 2003: The evolution of, and revolution in, land surface schemes designed for climate models. *Int. J. Climatol.,* **23,** 479–510.

——, and S. E. Perkins, 2009: Global and regional comparison of daily 2-m and 1000-hPa maximum and minimum temperatures in three global reanalyses. *J. Climate,* **22,** 4667–4681.

Qian, W., and A. Qin, 2005: Spatial-temporal characteristics of temperature variation in China. *Meteor. Atmos. Phys.,* **93,** 1–16.

Roads, J., 2003: The NCEP-NCAR, NCEP-DOE, and TRMM tropical atmosphere hydrologic cycles. *J. Hydrometeor.,* **4,** 826–840.

Shi, X., and X. Xu, 2007: Regional characteristics of the interdecadal turning of winter/summer climate modes in Chinese mainland. *Chin. Sci. Bull.,* **52,** 101–112.

——, ——, and L. Xie, 2006: Reliability analyses of anomalies of NCEP/NCAR reanalysis wind speed and surface temperature in climate change research in China. *Acta Meteor. Sin.,* **64,** 709–722.

Simmons, A. J., S. M. Uppala, and D. P. Dee, 2007a: Update on ERA-Interim. *ECMWF Newsletter,* No. 111, ECMWF, Reading, United Kingdom, 5.

——, ——, ——, and S. Kobayashi, 2007b: ERA-Interim: New ECMWF reanalysis products from 1989 onwards. *ECMWF Newsletter,* No. 110, ECMWF, Reading, United Kingdom, 25–35.

Smith, S. R., D. M. Legler, and K. V. Verzone, 2001: Quantifying uncertainties in NCEP reanalyses using high-quality research vessel observations. *J. Climate,* **14,** 4062–4072.

Solomon, S., D. Qin, M. Manning, M. Marquis, K. Averyt, M. M. B. Tignor, H. L. Miller Jr., and Z. Chen, 2007: *Climate Change 2007: The Physical Sciences Basis*. Cambridge University Press, 996 pp.

Trenberth, K. E., and J. G. Olson, 1988: An evaluation and intercomparison of global analyses from NMC and ECMWF. *Bull. Amer. Meteor. Soc.,* **69,** 1047–1057.

——, and Coauthors, 2007: Observations: Surface and atmospheric climate change. *Climate Change 2007: The Physical Science Basis,* S. Solomon et al., Eds., Cambridge University Press, 235–336.

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

——, D. P. Dee, S. Kobayashi, P. Berrisford, and A. J. Simmons, 2008: Towards a climate data assimilation system: Status update of ERAInterim. *ECMWF Newsletter,* No. 115, ECMWF, Reading, United Kingdom, 12–18.

Vincent, L. A., X. Zhang, B. R. Bonsal, and W. D. Hogg, 2002: Homogenization of daily temperatures over Canada. *J. Climate,* **15,** 1322–1334.

Vose, R. S., D. R. Easterling, and B. Gleason, 2005: Maximum and minimum temperature trends for the globe: An update through 2004. *Geophys. Res. Lett.,* **32,** L23822, doi:10.1029/2005GL024379.

World Meteorological Organization, 1993: *Guide on the Global Data Processing System*. No. 305, World Meteorological Organization.

Xu, Y., Y. Ding, and Z. Zhao, 2001: Confidence analysis of NCEP/NCAR 50-year global reanalyzed data in climate change research in China (in Chinese). *J. Appl. Meter. Sci.,* **12,** 337–347.

Yan, Z., and Coauthors, 2002: Trends of extreme temperatures in Europe and China based on daily observations. *Climatic Change,* **53,** 355–392.

Zhai, P., and X. Pan, 2003: Trends in temperature extremes during 1951–1999 in China. *Geophys. Res. Lett.,* **30,** 1913, doi:10.1029/2003GL018004.

——, A. Sun, F. Ren, X. Liu, B. Gao, and Q. Zhang, 1999: Changes of climate extremes in China. *Climatic Change,* **42,** 203–218.

——, Q. Chao, and X. Zou, 2004: Progress in China's climate change study in the 20th century. *J. Geophys. Sci.,* **14,** 3–11.

Zhao, T., and C. Fu, 2006: Preliminary comparison and analysis between ERA-40, NCEP-2 reanalysis and observations over China (in Chinese). *Climatic Environ. Res.,* **11,** 14–32.

——, and ——, 2009: Applicability evaluation of surface air temperature from several reanalysis datasets in China (in Chinese). *Plateau Meteor.,* **28,** 594–606.

——, W. Guo, and C. Fu, 2008: Calibrating and evaluating reanalysis surface temperature error by topographic correction. *J. Climate,* **21,** 1440–1446.

Zhou, L., R. E. Dickinson, Y. Tian, J. Fang, Q. Li, R. K. Kaufmann, C. J. Tucker, and R. B. Myneni, 2004: Evidence for a significant urbanization effect on climate in China. *Proc. Natl. Acad. Sci. USA,* **101,** 9540–9544.