An Improved Dynamical Downscaling Method with GCM Bias Corrections and Its Validation with 30 Years of Climate Simulations

ZHONGFENG XU
Department of Geological Sciences, The Jackson School of Geosciences, The University of Texas at Austin, Austin, Texas, and RCE-TEA, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

ZONG-LIANG YANG
Department of Geological Sciences, The Jackson School of Geosciences, The University of Texas at Austin, Austin, Texas

(Manuscript received 22 December 2011, in final form 15 March 2012)

ABSTRACT
An improved dynamical downscaling method (IDD) with general circulation model (GCM) bias corrections is developed and assessed over North America. A set of regional climate simulations is performed with the Weather Research and Forecasting Model (WRF) version 3.3 embedded in the National Center for Atmospheric Research’s (NCAR’s) Community Atmosphere Model (CAM). The GCM climatological means and the amplitudes of interannual variations are adjusted based on the National Centers for Environmental Prediction (NCEP)–NCAR global reanalysis products (NNRP) before using them to drive WRF. In this study, the WRF downscaling experiments are identical except the initial and lateral boundary conditions derived from the NNRP, original GCM output, and bias-corrected GCM output, respectively. The analysis finds that the IDD greatly improves the downscaled climate in both climatological means and extreme events relative to the traditional dynamical downscaling approach (TDD). The errors of downscaled climatological mean air temperature, geopotential height, wind vector, moisture, and precipitation are greatly reduced when the GCM bias corrections are applied. In the meantime, IDD also improves the downscaled extreme events characterized by the reduced errors in 2-yr return levels of surface air temperature and precipitation. In comparison with TDD, IDD is also able to produce a more realistic probability distribution in summer daily maximum temperature over the central U.S.–Canada region as well as in summer and winter daily precipitation over the middle and eastern United States.

1. Introduction
A relatively accurate regional projection of future climate and its impacts on society and environment has become crucial for public policy and decision making. Regional climate projection has been a major topic in the Intergovernmental Panel on Climate Change (IPCC) assessment report (Solomon et al. 2007). High-resolution regional climate model (RCM) simulation (“dynamical downscaling”) or statistical methods (“statistical downscaling”) are the common approaches to obtain fine spatial-scale information from long-term global circulation model (GCM) simulations. RCMs are formulated in terms of physical principles and therefore have the potential for capturing fine spatial-scale nonlinear effect, which increases confidence in their abilities to downscale future climate. Additionally, RCMs can better resolve orographic effects than the coarse-resolution GCMs. RCMs have been widely used to provide a better projection of future climate at the regional scale and have proven to be a powerful tool in the dynamical downscaling of regional climate since the 1980s (e.g., Dickinson et al. 1989; Giorgi et al. 1994; Wang et al. 2004; Lo et al. 2008; Heikkila et al. 2010).

The common approach to future climate dynamical downscaling employs a continuous integration of RCMs where GCM data are used to provide initial conditions (ICs), lateral boundary conditions (LBCs), sea surface temperatures (SSTs), and initial land surface conditions. These GCM-driven simulations are herein called the traditional dynamical downscaling (TDD) approach in which GCM outputs are directly used as ICs and LBCs for
RCMs (e.g., Giorgi and Hewitson 2001; Seth et al. 2007; Bukovsky and Karoly 2011). This approach has been employed by numerous regional climate modeling and assessment projects such as the Regional Climate Model Intercomparison Project for Asia (RMIP; Fu et al. 2005), Ensembles-Based Predictions of Climate Changes and Their Impacts (ENSEMBLES; van der Linden and Mitchell 2009), and the North American Regional Climate Change Assessment Program (NARCCAP) (Mears et al. 2009). It is known that GCMs are not perfect and all simulations suffer from systematic biases to a certain extent. The downscaled simulation is strongly influenced by the skill of the GCM. The TDD certainly brings GCM bias into RCMs through the LBC of the RCMs and degrades the downscaled simulation (Kim and Miller 2000; Rojas and Seth 2003; Wu et al. 2005; Seth et al. 2007; Cook and Vizy 2008; Caldwell et al. 2009).

Some bias correction methods have been used to improve the regional climate downscaling simulation. Wu and Lynch (2000) investigated the response of seasonal carbon cycle in Alaska to future climate change through a dynamical downscaling approach in which they constructed the LBC of an RCM by adding projected changes of temperature and specific humidity in a GCM simulation to reanalysis climate. A similar method was employed by Sato et al. (2007) to investigate the influence of global warming on regional precipitation over Mongolia. The bias correction was applied to wind speed, temperature, geopotential height, specific humidity, and sea surface temperature in Sato et al. (2007). Their results suggested that the rainfall intensity simulated with the new method was closer to observations than the traditional method. Using a method similar to that of Sato et al. (2007), Cook and Vizy (2008) investigated the effect of climate change on the Amazon rain forest and Patricola and Cook (2010) projected the Northern African climate at the end of the twenty-first century. The climatological LBCs in the aforementioned studies maintain variations on the seasonal time scale but eliminate the diurnal and synoptic effects. The climatological mean boundary conditions are not suitable to investigate transient climate variabilities (e.g., interannual variability) or climate extremes.

Holland et al. (2010) developed a more complex bias correction method for hurricane simulation. Their bias correction retained the diurnal and synoptic effects and the interannual variations in the LBC by correcting GCM climatological mean bias with 6-hourly National Centers for Environmental Prediction (NCEP)–National Center for Atmospheric Research (NCAR) reanalysis data and GCM outputs. Their results suggested that the dynamical downscaling simulation with GCM bias correction can produce reasonable tropical cyclone frequency because the bias correction reduced the unrealistic high vertical wind shear over the tropical North Atlantic. Jin et al. (2011) developed a statistical regression model between GCM and reanalysis data to eliminate the GCM climatological bias, and then the bias-corrected GCM data were used to force an RCM to simulate winter precipitation over the western United States. This approach also produced a precipitation simulation closer to observations than the TDD approach.

The previous downscaling studies have not performed in-depth comparisons between the TDD and the GCM bias-corrected dynamical downscaling methods in which the GCM bias correction has been employed. It is still not quite clear to what extent the GCM bias correction is able to improve the downscaled regional climate projection. Additionally, the previous bias corrections only corrected the GCM climatological mean bias. No effort has been made to correct GCM variance bias yet. The GCM variance may impact extreme events simulated by an RCM through the LBC of the RCM. In this study we develop an improved dynamical downscaling method (IDD) for regional climate projection, which incorporates both the GCM climatological mean and variance bias corrections. In-depth analyses are performed to assess the performance of IDD relative to TDD. Many previous validations of RCM performance were based on reanalysis-driven RCM simulations (e.g., Leung et al. 2003; Seth and Rojas 2003; Lo et al. 2008; Kanamitsu et al. 2010, Ainslie and Jackson 2010; Heikkila et al. 2010). The difference between reanalysis-driven RCM simulations and observations can be regarded as RCM errors if we take the reanalysis data as the “perfect” IC and LBC. In comparison with the reanalysis-driven simulations, the GCM-driven simulations usually show larger bias due to the GCM bias, which is an additional error source in RCM simulations. Therefore, differences between GCM-driven simulations and observations result from both the GCM and RCM biases. Two biases could appear with the same or opposite sign. The opposite sign biases would offset each other. In this case, the GCM-driven simulation may be closer to observation than the reanalysis-driven simulation, which does not mean that the GCM-driven simulation is actually better than the reanalysis-driven simulation. In this study, to examine the influence of GCM bias correction on the performance of an RCM downscaling simulation, the GCM-driven simulations will be compared with reanalysis-driven simulations rather than with observations, which is different from previous validations (e.g., Sato et al. 2007; Jin et al. 2011). In addition, long-term simulations are necessary to compare the climatological means between different dynamical downscaling experiments.

We describe the GCM bias correction method in section 2. Section 3 introduces the numerical models and
experiment design. Section 4 presents the dynamical downscaling simulations by comparing the differences between the IDD and TDD simulations. Conclusions and discussion are given in section 5.

2. Methodology for dynamical downscaling with GCM bias corrections

The global model used in this study is the Community Atmosphere Model (CAM), which is the atmospheric component of the Community Earth System Model (CESM) developed at NCAR. CAM biases were corrected by using NCEP–NCAR global reanalysis products (NNRP) data over the period 1950–79. To facilitate the validation of RCM simulations on future climate, we broke down 61 years (1950–2010) into two periods: the past (1950–79) and the future (1980–2010). No future NNRP data were used when we produced the bias-corrected CAM data over the future period. The future CAM-driven RCM simulations were then compared with the future reanalysis-driven RCM simulation to assess the performance of CAM-driven RCM in simulating the future climate. Six-hourly NNRP data and CAM outputs were broken down into a climatological mean plus a perturbation term:

\[ \text{CAM} = \overline{\text{CAM}} + \text{CAM}' \]  \hfill (1)

\[ \text{NNRP} = \overline{\text{NNRP}} + \text{NNRP}' \]  \hfill (2)

The future 6-hourly mean CAM data, CAM\(_F\), were written as follows:

\[ \text{CAM}_F = \overline{\text{CAM}_F} + \text{CAM}'_F \]

\[ = \overline{\text{NNRP}}_P + (\overline{\text{CAM}}_P - \overline{\text{NNRP}}_P) \]

\[ + (\overline{\text{CAM}_F} - \overline{\text{CAM}}_P) + \text{CAM}'_F. \]  \hfill (3)

The subscripts \(P\) and \(F\) represent the past (1950–79) and the future (1980–2010), respectively. Four terms on the right-hand side of Eq. (3) are the NNRP climatological mean over the past period, the CAM mean value bias relative to NNRP over the past period, the CAM simulated climate change between the future and the past periods, and the CAM anomaly in the future. The climate mean bias-corrected CAM data were constructed by removing the CAM bias \(\overline{\text{CAM}_F} - \overline{\text{NNRP}}_P\) in Eq. (3):

\[ \text{CAM}^*_F = \overline{\text{NNRP}}_P + (\overline{\text{CAM}_F} - \overline{\text{CAM}}_P) + \text{CAM}'_F \]

\[ = \overline{\text{NNRP}}_P + \text{CAM}_F - \overline{\text{CAM}}_P. \]  \hfill (4)

The bias-corrected 6-hourly CAM data, CAM\(_F^*\), over the future period (1980–2010) have a base climate provided by the NNRP data from the period of 1950–79, mean climate change between the future (1980–2010) and the past (1950–79), and future weather and climate variabilities simulated by the CAM. However, the CAM simulations can have also biases in temporal variations. We therefore improved the bias correction method further by correcting both the climatological mean and variance biases in the CAM outputs over the future period:

\[ \text{CAM}^*_F = \overline{\text{NNRP}}_P + (\overline{\text{CAM}_F} - \overline{\text{CAM}}_P) \]

\[ + \text{CAM}_F^* \frac{\text{S}_{\text{NNRP}}}{\text{S}_{\text{CAM}}}, \]  \hfill (5)

where \(\text{S}_{\text{NNRP}}\) and \(\text{S}_{\text{CAM}}\) represent the standard deviation of NNRP data and CAM simulations, respectively. Clearly Eq. (4) only adjusted the climatological mean states of CAM simulations, which is the same as the bias correction method used by Holland et al. (2010). In contrast, Eq. (5) adjusted both the climatological mean states and the variances of CAM simulations. In the meantime, the bias correction method allowed retaining the CAM simulated climatic change in the mean seasonal state, diurnal cycle, and variance of interannual variation. The bias correction to CAM variance described herein has not been attempted in previous GCM-driven dynamical downscaling simulations to the authors’ knowledge.

The bias corrections were applied to the air temperature, zonal wind, meridional wind, geopotential height, and relative humidity at each model grid vertical level for every 6-hourly CAM output file. The surface variables are not bias corrected in the experiments reported in this paper. As an example, Fig. 1 shows the original and the bias-corrected 200-hPa July temperature in a single CAM model grid (55°N, 140°W) versus the NNRP data during 1981–2010. The 30-yr averaged value and standard deviation in each data are showed on the top of the figure. In comparison with the NNRP, the original CAM simulation underestimates 200-hPa temperature by 9.7 K and its standard deviation by 1.1 K. In Fig. 1, CAMbc_ave and CAMbc_std represent the CAM\(_F^*\) and CAM\(_F^{**}\) in Eqs. (4) and (5), respectively. The mean temperature bias is remarkably reduced from 9.7 K in CAM to 0.3 K in CAMbc_ave, while the standard deviation remains the same as that in the original CAM output. In contrast, both the CAM mean value and variance biases are largely reduced in the CAMbc_std experiment. To be brief, the bias correction method removes CAM biases of mean value and variance through shifting and scaling the original CAM simulation based on the NNRP data. However the bias correction methods do not change the climatic trend and phase of interannual variability simulated by the CAM.
3. Model description and experimental setup

The Community Atmosphere Model (Neale et al. 2010) was coupled to an active land model (CLM) (Oleson et al. 2010), a thermodynamic only sea ice model (CICE), and a data ocean model (DOCN). In this study, a 63-yr simulation was performed by using the CAM at a resolution of T42 (approximately 2.8° × 2.8°) and 30 vertical layers with observed monthly SSTs from 1948 to 2010. The first two years were discarded for spinup. Model outputs were saved in 6-h interval.

The regional model used in this study is the Weather Research and Forecasting Model (WRF) with the Advanced Research WRF (ARW) dynamic core version 3.3 (Skamarock et al. 2008). This model has been developed and maintained by NCAR. It is a non-hydrostatic model with 28 vertical levels designed to serve both atmospheric research and operational forecasting needs. The WRF domain is centered at 40°N, 97°W with dimensions of 106 × 76 horizontal grid points (Fig. 2). Horizontal resolution of 60 km was used with 28 vertical levels, and the time step was 360 s. The main physical options we used included the new Kain–Fritsch convective parameterization (Kain 2004), CAM short-wave and longwave radiation schemes (Collins et al. 2004), the Noah land surface model (Chen and Dudhia 2001), and the Yonsei University planetary boundary layer scheme (Hong et al. 2006). The ICs, LBCs, and SST were given by the CAM 6-hourly outputs or NNRP 6-hourly data.

There is a 6-h mismatch of the diurnal cycles between the NNRP data and CAM simulations because each 6-hourly CAM output saved the past 6-hourly mean data and each NNRP saved the future 6-hourly mean data. A 6-h shift to the CAM output was performed to have the diurnal cycle in the CAM output match with that in the NNRP. As summarized in Table 1, four dynamical downscaling simulations were carried out to assess the performance of the IDD relative to the TDD in downscaling the future climate. The WRF configurations were kept the same in all experiments except that the ICs and LBCs were derived from the NNRP data (WRF_NNRP), the original CAM output (WRF_CAM), CAM output with mean value bias correction (WRF_CAMbc_ave), and CAM output with both mean value and variance bias corrections (WRF_CAMbc_std). All experiments employed the same SST, sea ice, vegetation distributions, and background albedo. Each WRF simulation was integrated over 31 years from 1980 to 2010. The first year was discarded for spinup. Model outputs were saved in 3-h intervals.

4. Dynamical downscaling simulations

Since the differences between RCM experiments and observations result from both the GCM and RCM
biases, we therefore compare the WRF_CAM, WRF_CAMbc_ave, and WRF_CAMbc_std experiments with the WRF_NNRP experiment to isolate the GCM bias correction influence. Given the identical model setup, the differences between various WRF simulations result from the differences in IC and LBCs. We take NNRP as the “perfect” IC and LBC and all bias corrections are based on the NNRP data. We therefore expect that the WRF_CAMbc_std simulation will be closer to the WRF_NNRP simulation than the WRF_CAM simulation is.

The WRF dynamical downscaling experiments are assessed by some statistical verification techniques. Root-mean-square error (RMSE) is one of the commonly used error statistics, and it is defined as

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (M_i - O_i)^2},$$

where $N$ is the total number of grid points and $M$ and $O$ represent the CAM-driven WRF simulations and NNRP-driven simulations, respectively. Following Shukla and Saha (1974) with a slight modification, the wind vector RMSE is defined by

$$\text{RMSE}_V = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \sqrt{(M_{u,i} - O_{u,i})^2 + (M_{v,i} - O_{v,i})^2}},$$

where $u$ and $v$ subscripts indicate the zonal and meridional wind, respectively. All statistics presented in this paper are based on the 30-yr WRF simulations from 1981 to 2010. The thick outline in Fig. 2 indicates the verification region where the performance of dynamical downscaling simulations will be examined. The verification region excludes all buffer zones of the WRF.

### a. Upper air variables

Figure 3 shows the annual mean profile of RMSEs of air temperature, geopotential height, wind vector, and water vapor mixing ratio. The annual mean RMSEs are able to represent the overall performance of each simulation in one year since RMSE is always equal to or larger than zero. The RMSEs of air temperature and geopotential height are greatly reduced, especially over the upper troposphere, when the GCM bias corrections are applied (Figs. 4a,b). We also computed the difference of area- and seasonally-averaged air temperature and geopotential height between the WRF simulations with and without GCM bias corrections (not shown). Generally the WRF_CAM experiment shows a cold bias in the troposphere relative to the WRF_NNRP experiment throughout the year. The maximum cold bias appears in winter. As a result the simulated geopotential height is systematically lower over the mid to upper troposphere in the WRF_CAM experiment than

### Table 1. Brief summary of downscaling simulations.

<table>
<thead>
<tr>
<th>Experiment identifier</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>WRF_NNRP</td>
<td>WRF experiment with the NNRP data as the initial and lateral boundary conditions</td>
</tr>
<tr>
<td>WRF_CAM</td>
<td>Traditional dynamical downscaling approach. WRF experiment with the original CAM output as the initial and lateral boundary conditions</td>
</tr>
<tr>
<td>WRF_CAMbc_ave</td>
<td>Same as WRF_CAM except the climatological mean biases in CAM output are corrected.</td>
</tr>
<tr>
<td>WRF_CAMbc_std</td>
<td>Improved dynamical downscaling approach. Same as WRF_CAM except that both the climatological mean biases and the variance biases in CAM output are corrected.</td>
</tr>
</tbody>
</table>
that in the WRF_NNRP experiment. The cold bias in WRF can be largely removed in spring, summer, and autumn when the GCM bias corrections are applied. In contrast, only part of the cold bias is removed in lower troposphere in winter. Similar to the air temperature and geopotential height, the wind vector simulation is also improved in the IDD simulation relative to that in the TDD. The improvement appears to be more remarkable at the upper troposphere than the lower troposphere for all variables except the moisture. One possible reason is that the CAM shows larger bias in the upper troposphere. This feature has also been found in an earlier version of CAM (e.g., Khairoutdinov et al. 2005). As a result, the bias correction influence is larger in the upper troposphere and its influence can be more easily transported into the RCM domain through advection effects than in the lower troposphere, and the lower troposphere is strongly impacted by the land surface condition (Gold 1967). The change of low-level air temperature is largely dominated by the land surface process or by the sea surface condition where no SST bias correction is applied yet. Lower tropospheric moisture is also improved, which indicates that the moisture transfer simulated by the IDD experiment is more reasonable than that in the TDD experiment.

Generally the WRF_CAMbc_ave and WRF_CAMbc_std experiments show very similar performance in simulating climatological mean variables, which means that the bias correction to variance only plays a minor role in modulating the climatological mean states of the downscaled climate (Fig. 3). The errors of 500-hPa geopotential height for the simulations with and without GCM bias corrections in winter [December–February (DJF)] and summer [June–August (JJA)] give a spatial overview of the model performance. A very significant improvement to 500-hPa geopotential height simulation has been found in winter when the GCM bias corrections

---

**Fig. 3.** Annual mean RMSEs of (a) air temperature, (b) geopotential height, (c) wind vector, and (d) water vapor mixing ratio. The RMSEs are computed over the verification region by monthly mean data.
are applied (Figs. 4a,b) and very significant improvements can also be found in spring and autumn (not shown), whereas the improvement in summer is not as remarkable as in other seasons although the simulation is also been improved (Figs. 4c,d). The reason is that the CAM bias is generally smaller in summer than in other seasons. There is no room for large improvement in summer season in the IDD experiment. Similar results are found in 850-hPa winds. The winter wind vector bias is much smaller over the North American continental region in the WRF_CAMbc_std than in the WRF_CAM experiment (Figs. 5a,b), while the improvement in summer is not as significant as that in winter (Figs. 5c,d). Given the identical model setup, these differences between the WRF_CAMbc_std and WRF_CAM experiments are clearly due to the differences in their ICs and LBCs. The GCM bias corrections significantly improve the air temperature, wind, geopotential height, and relative humidity in the CAM output, which in turn leads to the better WRF dynamical downscaling simulations.

b. Surface air temperature

Surface air temperature is one of the principal climatic elements influencing human life and is also an important indicator in measuring the performance of climate simulation. Given the importance of surface air temperature in climate simulation, Fig. 6 shows the monthly RMSEs of climatological mean 2-m air temperature (T2m) for three CAM-driven WRF downscaling experiments. All experiments show relatively large bias in winter and small bias in summer. The IDD simulation shows smaller RMSEs than the TDD simulation throughout the year except in June. Generally the new method performs better in spring [March–May (MAM)] and autumn [September–November (SON)] than in winter (DJF) and summer (JJA). The WRF_CAMbc_ave and WRF_CAMbc_std experiments show similar performance in the simulation of T2m, which means that the GCM variance bias correction only improves the downscaled mean climate slightly.

Figure 7 illustrates the spatial pattern of errors of climatological mean T2m in winter and summer. In winter, the WRF_CAM shows a pronounced cold bias within the model domain. Similar cold biases were also found in previous studies (Caldwell et al. 2009; Jin et al. 2011). The cold bias is slightly reduced in the WRF_CAMbc_std experiment as compared to the WRF_CAM experiment (Figs. 7a,b). The cold bias may be partially attributed to the fact that the CAM underestimates the observed air temperature. In this study
The CAM model was forced by the observed monthly SST. The greenhouse gases were fixed at year 2000 level (Solomon et al. 2007). However, the lower-tropospheric (below 500 hPa) air temperature simulated by the CAM still shows a 1.1\degree C cold bias relative to the NNRP during the period of 1980–2010, which in turn leads to a lower temperature in the WRF_CAM experiment than in the WRF_NNRP experiment. The bias corrections can remove CAM mean value bias \( \overline{\text{CAM}}_P - \overline{\text{NNRP}}_P \) in Eq. (3) while the climate change term \( \text{CAM}_F - \text{CAM}_P \) all depends on the GCM simulation. There is still a 0.3\degree C cold bias in the bias-corrected CAM data relative to the NNRP data, which may partly account for the winter cold bias in the downscaled climate. In addition to the cold bias in the LBC of WRF, the cold temperature–snow–albedo positive feedback may also account for the cold bias in

![Figure 5](image1)

**FIG. 5.** As in Fig. 4, but for 850-hPa winds (m s\(^{-1}\)). Shaded areas indicate the overestimated or underestimated wind speed reach 99% confidence.

![Figure 6](image2)

**FIG. 6.** Root-mean-square errors (RMSEs) of 2-m air temperature over the verification region. RMSEs are computed based on climatological mean monthly 2-m air temperature.

![Additional figure](image3)
the WRF_CAM and WRF_CAMbc_std experiments. We compared the albedo and snow cover simulated by the WRF_CAM, WRF_CAMbc_std, and WRF_NNRP experiments. Generally the WRF_CAM and WRF_CAMbc_std experiments overestimate snow cover, which leads to a higher surface albedo. The cold bias pattern in Fig. 7a well corresponds to the region with higher land surface albedo. The WRF_CAM also underestimates the summer temperature by 2°C over the central U.S.–Canada region (Fig. 7c). The cold bias is reduced when the GCM bias corrections are applied in the WRF_CAMbc_std experiment (Fig. 7d). The summer cold bias is closely related to the wet bias in the WRF_CAM experiment, which will be elucidated in the following section.

c. Precipitation

Figure 8 illustrates the RMSEs of precipitation from the WRF_CAM, WRF_CAMbc_ave, and WRF_CAMbc_std experiments. The comparison of the WRF_CAM with the WRF_CAMbc_ave and WRF_CAMbc_std experiments indicates that the errors of downscaled precipitation are smaller throughout the year when the GCM bias corrections are applied. The most significant improvement appears in August, September, and October. To show spatial difference between different dynamical downscaling experiments, Fig. 9 illustrates the differences of precipitation between the CAM-driven WRF simulation and the NNRP-driven WRF simulation in winter and summer. The shaded area indicates the difference reaches the 99% confidence level. In winter the WRF_CAM experiment overestimates precipitation over Mexico and the western Atlantic Ocean and underestimates the precipitation over the middle United States (Fig. 9a). These errors are greatly removed when the GCM bias corrections are applied (Fig. 9b). The WRF_CAM experiment significantly overestimates summer precipitation by 0.5–1.5 mm day⁻¹ over the central U.S.–Canada region and 2–12 mm day⁻¹ over the western Atlantic Ocean and Gulf of Mexico (Fig. 9c). The WRF_CAM overestimates low troposphere moisture by 3%–6% over the central U.S.–Canada region, which likely plays a substantial role in the higher-than-average precipitation simulated by the WRF_CAM (not shown). The overestimated precipitation over the central U.S.–Canada region in the WRF_CAM experiment leads to a higher soil moisture content and enhanced evaporation, which in turn leads to a lower surface air temperature (Figs. 7c and 9c). These significant errors of precipitation in the WRF_CAM experiment are largely removed in the WRF_CAMbc_std experiment because of the GCM bias corrections (Fig. 9d). The new dynamical downscaling method shows significant improvement in precipitation simulation over the
North American domain although the new method may also degrade the dynamical downscaling performance somewhere, such as the winter precipitation over the boundary of the United States and Canada and the summer precipitation over western Canada.

d. Extreme events

To examine the influence of GCM bias corrections on RCM simulation of extreme events, the difference of 2-yr return levels of daily minimum and maximum temperature in winter and summer are respectively presented in Fig. 10. An $N$-yr return level is defined as the event occurring on average once every $N$ years (i.e., $N$ is the return period). For example, the 2-yr return level of daily maximum temperature in a region is 40°C, which indicates that the extreme event with daily maximum temperature higher than 40°C occurs on average every 2 years. Return level has been widely used in climate extreme studies (e.g., Wigley 2009; Cooley 2009; Pausader et al. 2011). The 2-yr return level shown in Fig. 10 is equivalent to the 99.5th percentile value of maximum or minimum temperature approximately. We also computed the 1-yr and 5-yr return levels. It turns out that the conclusions presented in this study are the same no

(a) WRF_CAM – WRF_NNRP in winter

(b) WRF_CAMbc_std – WRF_NNRP in winter

(c) WRF_CAM – WRF_NNRP in summer

(d) WRF_CAMbc_std – WRF_NNRP in summer

Fig. 8. As in Fig. 6, but for precipitation.

Fig. 9. As in Fig. 7, but for precipitation. Contour interval is 0.5 mm day$^{-1}$ for the contours between –2 and 2 mm day$^{-1}$, and 2 mm day$^{-1}$ for the remaining contours.
matter which return level is selected. The daily minimum
temperature in the WRF_CAM suffers from problems
similar to climatological mean temperature in the
WRF_CAM simulations: both the climatological mean
and extreme daily minimum temperature are under-
estimated, and the CAM bias corrections only slightly re-
move the cold bias over the northern part of the
verification domain (Figs. 7a,b and 10a–c). In contrast, the
CAM bias corrections do significantly improve daily
maximum T2m simulation in summer. The TDD simu-
lation, WRF_CAM, underestimates the extreme daily
maximum temperature by 2–6°C to the east of Great
Lakes in summer (Fig. 10d). The WRF_CAMbc_ave
experiment greatly reduced the cold bias through
the CAM mean value bias correction (Fig. 10e). The
WRF_CAMbc_std experiment further improves the per-
formance of dynamical downscaling simulation by re-
moving almost all biases to the south of 50°N by means of
the CAM bias corrections in both mean value and variance.

Similarly, the WRF_CAMbc_std also improves the
extreme precipitation simulation as illustrated in Fig. 11
for the 2-yr return levels of daily precipitation in winter
and summer. All CAM-driven simulations underestimate
the extreme precipitation by 3–15 mm day\textsuperscript{-1} over the
central to eastern United States and overestimate ex-
treme precipitation by 3–18 mm day\textsuperscript{-1} over Mexico in
winter. The comparison of Fig. 11a with Figs. 11b and 11c
indicates that the GCM bias corrections are able to im-
prove winter extreme precipitation simulation charac-
terized by the smaller errors of 2-yr return level of

![Figure 10](https://example.com/fig10.png)

**FIG. 10.** Errors of 2-yr return levels of (a)–(c) daily minimum temperature in winter and (d)–(f) maximum temperature in summer. The
shaded areas denote the difference of daily minimum temperature is smaller than \(-2°C\) in winter and the difference of daily maximum
temperature is larger than \(1°C\) or smaller than \(-1°C\) in summer.
precipitation. For the downscaled summer extreme precipitation, the GCM bias corrections moderately improve the simulation in general (Figs. 11d–f). The WRF_CAMbc_ave produces a more realistic extreme precipitation over most parts of the United States and Canada than the WRF_CAM does except for western Canada. The WRF_CAMbc_std produces an overall better simulation of extreme precipitation over the verification domain relative to the WRF_CAM experiment except over the southwestern United States, where the WRF_CAMbc_std slightly overestimates the extreme precipitation.

IDD improves not only the extreme events simulation but also the possibility distributions of temperature and precipitation. Figure 12a shows the daily maximum T2m distribution computed over the central United States and Canada (40°–50°N, 100°–85°W) where TDD shows a large bias in daily maximum temperature. The maximum probability appears in 26°–28°C in the WRF_CAM simulation and 28°–30°C in the WRF_CAMbc_ave, WRF_CAMbc_std, and WRF_NNRP simulations. The WRF_CAM underestimates the frequency of extreme high temperature events. Clearly there are errors in both the mean and variance of the WRF_CAM relative to the WRF_NNRP experiment. The WRF_CAMbc_ave experiment is able to remove the error of mean and partly remove the error of variance. The WRF_CAMbc_std experiment is able to produce a distribution of daily maximum temperature very close to that in the WRF_NNRP experiment. Both errors in mean and variance are removed in the WRF_CAMbc_std experiment. The improvement is likely related to the better simulation on summer precipitation in the WRF_CAMbc_std experiment characterized by reduced wet
bias in precipitation and soil moisture over the central U.S.–Canada region, which in turn increases extreme high temperature events due to the reduced evaporation in association with the decreased soil moisture. However, the GCM bias corrections do not help improve the distribution of winter minimum temperature. The winter minimum temperature at a given probability is generally lower in the CAM-driven simulations than in the reanalysis-driven simulation (Fig. 12b). The daily averaged temperature also shows a distribution similar to the daily minimum temperature in winter (not shown). As a result, the climatological mean temperature in the CAM-driven simulations is lower than the reanalysis-driven simulation (Figs. 4a,b). IDD failed to improve the distribution of winter minimum temperature, which might be for the same reason that the IDD failed to improve climatological mean temperature in winter.

Similarly, the possibility distributions for summer and winter daily precipitation are shown in Figs. 12c and 12d, respectively. TDD overestimates the heavy precipitation in summer and underestimates heavy precipitation in winter. IDD improves the distribution of precipitation characterized by a more realistic tail of precipitation (Figs. 12c,d). The comparison of various
dynamical downscaling experiments shows that the WRF_CAMbc_std also improve the rainfall days simulation in summer. The error of summer rainfall days is less than $\pm 5\%$ over the central U.S.–Canada region in the WRF_CAMbc_std as opposed to 10%–15% in the WRF_CAM experiment (not shown).

5. Conclusions and discussion

A new dynamical downscaling method with GCM bias corrections for the regional projection of further climate was developed and validated by comparing the GCM-driven WRF simulations to the NNRP-driven WRF simulation. The GCM bias corrections involved shifting and scaling to adjust the GCM mean and variance so they matched with those in the NNRP data. The comparison of various dynamical downscaling simulations shows that the improved dynamical downscaling method with GCM bias corrections is superior to the traditional dynamical downscaling method characterized by the significant improvement in downscaled air temperature, geopotential height, wind vector, and water vapor. The most significant improvement appears in the middle-upper troposphere for the air temperature, geopotential height, and wind vectors, and in the lower troposphere for the water vapor. As a result, the downscaled surface air temperature and precipitation are also improved. The most significant improvement of summer precipitation appears in the central U.S.–Canada region where the TDD overestimates precipitation by 0.5–1.5 mm day$^{-1}$. The overestimated precipitation over the central U.S.–Canada region in TDD leads to a higher moisture content and enhanced evaporation, which in turn leads to a cold bias of surface air temperature. These significant errors in precipitation and surface temperature are largely removed in the IDD because of the GCM bias corrections.

Additionally, IDD is also able to improve the downscaled extreme events and the possibility distribution of temperature and precipitation. The 2-yr return level of summer daily maximum temperature simulated by the TDD is underestimated by 2°–6°C over the central U.S.–Canada region. In contrast, the bias is generally less than $\pm 1°C$ in the IDD experiment. IDD simulates a more realistic probability density distribution of daily maximum temperature over the central U.S.–Canada region. Similarly conclusions are found in the downscaled precipitation. TDD overestimates summer extreme precipitation and underestimates winter extreme precipitation. In contrast, IDD is able to produce better extreme precipitation in both seasons.

Although the new dynamical downscaling method is able to improve regional climate simulation of both climatological means and extreme events through removing GCM systemic bias, the performance of the dynamical downscaling simulation for regional climate is still sensitive to the GCM, especially to its projected climate change. The GCM bias corrections reported in this paper do not correct GCM bias in climate change simulation. Additionally, the performance of dynamical downscaling is also strongly impacted by RCM bias, which is not considered in this study. We selected the CAM radiation scheme in the WRF configuration, which is especially suited for regional climate simulation (Skamarock et al. 2008). We did not find a significant climate drift during the WRF simulation although both the reanalysis-driven and CAM-driven long-term (31 yr) WRF simulations show a similar warming trend of $0.1°–0.2°C$ decade$^{-1}$ in the surface air temperature. The warming trend should result from the CAM radiation scheme (Collins et al. 2004) employed in WRF that prescribes the CO$_2$ mixing ratio from the IPCC Special Report on Emissions Scenarios A2 scenario.

This study has demonstrated that the IDD is superior to the TDD when it is applied to North America, and it is expected that the method can be implemented for regional climate projection elsewhere. In addition to the future climate projection, IDD can also be applied to downscale GCM sensitivity simulations. In doing so, the GCM control (sensitivity) simulation is corresponding to the past (future) GCM simulation in the paper. The difference between the GCM sensitivity and control simulations is corresponding to that between the GCM future and past simulations as described in the present paper. Thus the difference between the GCM sensitivity and control simulations can be retained and passed onto the RCM, thereby further impacting the downscaled simulations. It is known the LBC becomes a stronger constraint to the regional climate simulation when the regional model domain size is smaller. We can therefore expect that the IDD will perform better when the regional model domain is smaller. Alternatively, spectral nudging may also help improve IDD simulation further, which will be presented in another paper. Although the influence of RCM systematic biases on the performance dynamical downscaling was not investigated in this study, the IDD would be expected to continue to produce more accurate downscaled climate when the RCM biases are reduced with the improvement of RCM in the future. We only examined the influences of the LBC bias corrections, while the surface condition influences (e.g., SSTs) were not considered. It is known that the SSTs simulated by a fully coupled climate system model also show significant biases. The bias corrections of temperature, zonal wind, meridional wind, geopotential height, and relative humidity only impact the ICs and LBCs of RCM. In contrast, the SST bias correction would continuously impact low boundary conditions.
inside the RCM domain during the model integration, which also plays a very important role in regulating the RCM simulations. The relative importance of bias corrections applied to LBCs versus SSTs is worth exploring in the future study.

Acknowledgments. The authors thank the Texas Advanced Computing Center for providing computer resources and technical support. NCEP–NCAR reanalysis data were obtained from the Research Data Archive (http://dss.ucar.edu/datasets/ds090.0/), which is maintained by the Computational and Information Systems Laboratory (CISL) at NCAR. Financial support was provided by NASA (Grant NNX11AE42G), “National Basic Research Program of China” Project 2011CB952004, National Natural Science Foundation of China General Program (Grant 40905042), and KAUST.

REFERENCES


