The Diurnal Temperature Range in CMIP6 Models: Climatology, Variability, and Evolution

KANG WANG

School of Geographic Sciences, East China Normal University, Shanghai, China

GARY D. CLOW

Institute of Arctic and Alpine Research, University of Colorado Boulder, Boulder, Colorado

(Manuscript received 2 December 2019, in final form 3 May 2020)

ABSTRACT

The diurnal temperature range (DTR) is an identifiable and sensitive indicator of the synchronicity of changes in diurnal temperature extrema, but capturing DTR dynamics is challenging for climate models. This study investigates the climatology, variability, and changes of DTR in recent models participating in phase 6 of the Coupled Model Intercomparison Project (CMIP6). The results show that the CMIP6 models underestimate DTR climatology relative to observations. Most individual models overestimate December–February variability, particularly at high latitudes of the Northern Hemisphere. The models show substantially different changes over land surfaces and do not fully capture the observed spatiotemporal evolution of DTR. Large intermodel differences seem to be controlled by daily minimum air temperature. In the Northern Hemisphere, precipitation and cloud longwave and shortwave radiative effects appear to make important contributions to the intermodel discrepancies. Evaporative fraction is an important factor contributing to the intermodel differences in DTR during the summer in the Northern Hemisphere. In general, CMIP6 models have not improved their ability to simulate temporal DTR changes in a consistent way over the entire analysis period (1901–2005) relative to their CMIP5 counterparts. For periods of rapid DTR decline (e.g., 1951–80) CMIP6 models are typically better than the CMIP5 versions at simulating DTR, whereas for other periods CMIP6 models underperform their CMIP5 counterparts.

1. Introduction

Air temperature is one of the primary variables used to judge the status and changes of the present climate (IPCC 2013). Indeed, mean annual, monthly, and daily air temperatures are commonly considered in climate diagnoses and climate model evaluations (IPCC 2013). However, many Earth system processes depend not only on mean values but also on extrema (Lindvall and Svensson 2015). The diurnal temperature range (DTR) captures the synchronicity of diurnal maximum and minimum temperatures (Karl et al. 1993; Easterling et al. 1997; Braganza et al. 2004) and thus has broad implications for the Earth system. Since daily minimum air temperatures are closely associated with the longwave radiative flux, whereas daily maximum air temperatures are basically determined by shortwave radiative forcing, changes in DTR could be a useful proxy of decadal surface radiative forcing changes caused by anthropogenic activities (Jackson and Forster 2013; Wang and Dickinson 2013). DTR also impacts crop growth (Lobell 2007; Battisti and Naylor 2009), carbon exchange in land–atmosphere interactions (Yi et al. 2010; Peng et al. 2013; Xia et al. 2014), and human health (Paaijmans et al. 2010). Thus, DTR has been listed as one of the core indices of climatic dynamics by the World Climate Research Program (WCRP; Peterson et al. 2001) and is a helpful diagnostic metric for evaluating climate models (Braganza et al. 2004).

DTR has declined, apparently nonlinearly, at the global scale since the early twentieth century (IPCC 2013; Thorne et al. 2016a,b; Wild 2009a) due to relatively stronger warming during the nighttime than during the day. DTR broadly increased prior to the 1950s (Wild 2009a; Thorne et al. 2016b; Sun et al. 2019), but recent studies show that DTR has decreased globally by about 0.044°–0.075°C decade⁻¹ since then (Thorne et al. 2016a,b; Sun et al. 2019). Since the end of the 1970s, slight decreases of DTR have been observed in several...
global datasets (Thorne et al. 2016b). Such DTR changes are essentially consistent with global dimming and brightening phases (Wild 2009a, 2016).

Capturing DTR variations is challenging for climate models (Stone and Weaver 2003) because DTR essentially depends on model performance at the daily scale. Cloud, precipitation, solar radiation, terrestrial vegetation, and local urbanization processes may play important roles in DTR dynamics (Dai et al. 1997; Zhou et al. 2004; Wild et al. 2005). Therefore, accurate simulation of DTR is more difficult than that of the mean temperature at monthly or annual scales. According to the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4), models participating in phase 3 of the Coupled Model Intercomparison Project (CMIP3) underestimated decadal DTR variations (Wild 2009b). More importantly, these models were unable to reproduce the shift from global dimming to brightening that occurred around the mid-1980s (Wild 2009b). Although the spread among the nine CMIP5 participant models was reduced compared to CMIP3 (Sillmann et al. 2013), their ensemble mean underestimated the DTR trend relative to observations since 1951 (Lewis and Karoly 2013). Interannual variations were also underestimated by CMIP5 models (Lewis and Karoly 2013; Lindvall and Svensson 2015), primarily due to a lack of decadal variations in surface solar radiation, cloud cover, and land surface processes (e.g., soil moisture) in these models (Lewis and Karoly 2013; Wild 2009b; Zhou et al. 2010). These systematic biases among CMIP5 models may still present a major challenge for current climate modeling (Stouffer et al. 2017).

CMIP6 retains the design of previous CMIPs, and is expected to provide modeling support for IPCC AR6 (Eyring et al. 2016). After a decade of model advancement, the models participating in CMIP6 are expected to show improvements because of our better understanding of Earth system physics, computational advances, and the participation of several new models. Thus, the questions, deservedly, are 1) how well do the CMIP6 models perform compared with observations and 2) how much have the CMIP6 models improved relative to their CMIP5 counterparts?

This study investigates DTR climatology, variability, and temporal changes in 15 CMIP6 participant models. The simulated DTR in the historical period (1901–2014) is compared with observational data to examine whether the models capture observed DTR changes. A correlative approach is employed to examine the causes of the large intermodel differences in historical simulations. Finally, earlier and current versions of the models are compared to evaluate if the latest versions have indeed been improved.

2. Data and methods

a. Observational datasets

This study uses the Climate Research Unit (CRU) datasets (v4.03, accessed 30 September 2019), developed and maintained by the CRU at the University of East Anglia, because 1) they cover the global land surface, except Antarctica, with a relatively high spatial resolution (0.5°), 2) they have been rigorously quality-controlled during data processing (Harris et al. 2014), and 3) they have been used in many evaluations of previous CMIPs (e.g., Wild 2009b; Lewis and Karoly 2013; Lindvall and Svensson 2015). The CRU datasets cover the period 1901–2018 (Harris et al. 2014), but this study focuses on the period 1901–2014 to match the timespan of CMIP6 historical simulations. In short, monthly mean DTR and daily maximum and minimum temperature data are used as benchmarks.

To confirm spatiotemporal features revealed by the CRU datasets, a new dataset created by Thorne et al. (2016b) (1901–2012) is included as an auxiliary data source. This dataset is a gridded product with 5° latitude–longitude spatial resolution. Thorne et al. (2016b) provided only the anomaly of DTR relative to the climatology of 1971–2000, which is slightly different from this study’s baseline period (1976–2005). Thus, HadGHCND (Caesar et al. 2006) is selected for providing a reference climatology (1976–2005) on 3.75° × 2.5° latitude–longitude grid. The HadGHCND dataset (Caesar et al. 2006) was compiled using daily observations from over 27 000 stations in the GHCN (Global Historical Climatology Network)–Daily database.

b. Model outputs

CMIP6 historical simulations cover the period 1850–2014. DTR is derived from the monthly mean of daily minimum and maximum air temperatures. To better diagnose and attribute model bias and intermodel divergence, model selection was based on the availability of the following nine variables in the historical ensemble member r1i1p1f1 (accessed 30 September 2019): monthly mean of daily minimum air temperature (TASMIN), monthly mean of daily maximum air temperature (TASMAX), precipitation (PR), surface downwelling longwave radiation (RLDS), surface downwelling longwave flux under clear sky (RLDSCS), surface downwelling shortwave radiation (RSDS), surface downwelling shortwave flux under clear sky (RSDSCS), surface upward latent-heat flux (HFLS), and surface upward sensible-heat flux (HFSS). The 15 models selected for analysis in this study are summarized in Table 1. Although the Nanjing University of Information Science and Technology (NUIST) Earth System Model version
Three derived variables, surface shortwave and longwave cloud radiative effects (SWCRE and LWCRE, respectively) and evaporative fraction (EF), were used as proxies for cloud radiative and evaporative effects following Ramanathan et al. (1989):

\[
\text{SWCRE} = \text{RSDS} - \text{RSDSCS}, \\
\text{LWCRE} = \text{RLDS} - \text{RLDSCS}, \\
\text{EF} = \frac{\text{HFLS}}{\text{HFLS} + \text{HFSS}}.
\]

Since the models have different spatial resolutions, they were linearly interpolated onto the coarsest grid (T42, 128 longitude grids × 64 latitude grids, roughly equivalent to 2.8° both latitudinally and longitudinally) and the ensemble mean calculated over the global land surface from the 15 models. Furthermore, corresponding earlier and current versions of the nine selected models that participated in both CMIP phases were compared to reveal the differences between CMIP6 and CMIP5 models (Table 1).
c. Statistical methods

Only land areas north of 60°S (i.e., excluding the Antarctic) were used in this study because of the coverage of the CRU dataset. Land masks were created for each model by remapping the raw mask (small-scale data, 1:110,000,000 scale) from Natural Earth. Detailed subregional features were investigated using the 21 subregional subdivisions of Giorgi and Francisco (2000) (Fig. 1).

Because observational data have become increasingly available in recent decades, and because CMIP5 ended in 2005, the period 1976–2005 was selected as the 30-yr baseline to calculate climatology and departures of specific variables. As regular latitude–longitude grid boxes are used in both the model outputs and CRU datasets, the area-weighted method (Jones and Hulme 1996) was used to calculate the averages of DTR and the other nine variables (listed in section 2b) in a subregion or latitudinal interval (T):

\[
T = \frac{\sum_{i=1}^{N} \cos\left(\text{lat}_i, \frac{\pi}{180}\right) \Delta T_i}{\sum_{i=1}^{N} \cos\left(\text{lat}_i, \frac{\pi}{180}\right)},
\]  

where \(N\) is the number of grid cells in the subregion or latitudinal interval, \(\text{lat}_i\) is the latitude of grid cell \(i\), and \(\Delta T_i\) is the observed anomaly (e.g., DTR) in grid cell \(i\). Annual, monthly, and seasonal anomalies were calculated relative to the 30-yr baseline period (1976–2005).

Climatic trends are one of the main indicators of climate change. Although they are generally estimated by linear regression, many studies have demonstrated climatic changes to be nonlinear (Wu et al. 2007; Ji et al. 2014; Franzke 2014). Thus, we used the ensemble empirical mode decomposition (EEMD) method (Wu and Huang 2009) to detect nonlinear trends and infer the spatiotemporal evolution of DTR. This method does not make any assumptions about the functional form (linear, cubic, etc.) of the target trend; that is, the form of the trend depends on the input data. EEMD decomposes and subtracts all oscillatory components from a climate signal, and the remainder constitutes the trend (Wu et al. 2007; Wu and Huang 2009). Therefore, EEMD avoids large uncertainties related to the selection of the onset and end of a linear trend analysis, or “eyeballing” the trend (Percival and Rothrock 2005).

The spatiotemporal evolution of DTR was investigated in latitudinal zones for both the models and observational dataset; this analysis was restricted to times since 1951 because of data availability. First, the 15 models and the CRU dataset were linearly interpolated onto the same 5° × 5° grid to make the models and CRU dataset comparable and eliminate noise from sporadic features. Second, latitudinally averaged anomalies were created for each 5° latitudinal interval using Eq. (4). Excluding two latitudinal zones without any data, a total of 28 latitudinal time series were obtained for each model and the CRU dataset. Third, the nonlinear trend was estimated for each latitudinal interval using EEMD. Finally, because of large differences in scale, the normalized root-mean-square deviation (NRMSTD, calculated in the “scikit-image” python package; van der Walt et al. 2014) was used to facilitate comparisons between the CRU dataset and the models. Furthermore, climatological changes were compared between 1951–80 and 1985–2014 for two observational datasets [the CRU dataset and the data from Thorne et al. (2016a)], individual models, and multimodel ensemble mean, respectively.
An intermodel correlative approach was used to examine which variables affected the differences in DTR temporal changes among the models. Eight variables were used in the correlative analysis: TASMIN, TASMAX, PR, EF, SWCRE, LWCRE, RSDS, and RLDS. Seasonal averages were used in the correlative analysis to avoid confusing warm and cold seasons in the Northern and Southern Hemispheres. The Spearman correlation is used in this study to avoid the inherent assumption of a normal distribution when using the Pearson method. The hypothesis test is two-sided at the 95% confidence level. Although the intermodel correlations do not represent causality, they indicate which parameters are likely to be important in explaining the differences among the models. A significant correlation indicates that intermodel differences in a specific variable may be important for explaining that seen in DTR.

**Fig. 2.** Climatology of the annually averaged diurnal temperature ranges (°C) during 1976–2005. The numbers in the center of each panel are the average plus and minus one standard deviation (°C) across available grids excluding the Antarctic region.
3. Results and discussion

a. Climatology

The annual mean DTR over land surfaces (excluding the Antarctic) was 11.4 ± 3.1°C during 1976–2005 according to the CRU dataset (Fig. 2). This observational dataset shows that the smallest observed annual mean overland DTRs are around 4–5°C, mainly over Arctic coastal regions, whereas northern Europe, northern Asia, and northern North America are in the range of 8–14°C. Large annual mean DTRs are generally found in dry (Australia, 13.8 ± 1.6°C; Sahara, 14.4 ± 1.9°C; western North America, 13.5 ± 2.8°C; southern Africa, 13.7 ± 2.8°C) and/or high-altitude regions (Tibet, 13.7 ± 1.4°C; Central America, 13.9 ± 3.1°C). In part of southeastern Asia (around 18°–30°N, 110°–120°E), the annual mean DTR is in the range of 6°–9°C, markedly lower than surrounding regions. The annual mean DTR from the HadGHCND dataset shows a similar spatial pattern over the available pixels (11.1 ± 2.4°C). One region that is significantly different in the HadGHCND and CRU datasets is Greenland, with DTR being much lower in the HadGHCND dataset (Figs. 2a,b).

CMIP6 models show large differences in annual mean DTRs, ranging from 8.5 ± 2.7°C (IPSL-CM6A-LR) to 10.4 ± 4.7°C (CanESM5) (Fig. 2). In addition, they generally underestimate DTR relative to the observations, consistent with previous intercomparisons of CMIP5 models (Sillmann et al. 2013; Lindvall and Svensson 2015). Obvious underestimates are found in Arctic coastal regions, including northern Eurasia and northern North America in most models (Fig. 2). GISS-E2.1-G, GISS-E2.1-G-CC, GISS-E2.1-H, IPSL-CM6A-LR, MIROC6, MRI-ESM2.0, and SAM0-UNICON also underestimate DTR by 2°–5°C in regions surrounding the Amazon Basin. Over Greenland and the Northern Territories, CanESM5 and IPSL-CM6A-LR simulate DTR relatively well (an underestimate of <3°C) whereas the remaining 13 models generally underestimate DTR by 3.6°–5.9°C. Globally, FGOALS-g3 (10.7 ± 2.9°C) represents DTR better than the other models, but still underestimates DTR by 3.6°C in Greenland and the Northern Territories relative to the CRU dataset. It is worth noting that EC-Earth3, EC-Earth3-Veg, GFDL models, GISS models, and MRI-ESM2.0 use different atmospheric modules, yet produce remarkably similar patterns, particularly at high latitudes in the Northern Hemisphere.

In each latitudinal zone, most models underestimate DTR relative to the CRU dataset, although there are a few notable large positive discrepancies as well (Fig. 3a). Most models have similar zonal patterns, with relatively low DTRs at high latitudes and in the tropics and high DTRs in midlatitude regions of both hemispheres. However, intermodel differences are fairly large in almost all latitudinal zones. An obvious overestimate is observed in...
midlatitude regions (mainly the Sahara), particularly in the Northern Hemisphere in the CanESM5, EC-Earth3, EC-Earth3-Veg, and MIROC6 models. In particular, MIROC6 greatly overestimates DTR around 20°N, primarily due to an overestimate of TASMAX (Figs. 3b,c).

b. Interannual variability

Model performance in simulating interannual variability was evaluated over the baseline period (1976–2005) before investigating DTR trends. Interannual variation is defined as the standard deviation of the detrended regional average anomalies (Lewis and Karoly 2013; Sillmann et al. 2013; Lindvall and Svensson 2015). These previous studies ignored nonlinear features because they deemed that assuming any polynomial form may introduce artificial biases (Sen Gupta et al. 2012). Here, the EEMD method is used to circumvent this problem (see section 2c).

Globally, the CRU dataset shows an interannual variability of 0.07°C over land surfaces during 1976–2005 (Fig. 4). Model outputs are fairly consistent with the annual mean values from the CRU dataset. Observed June–August and December–February seasonal variabilities are around 0.11°C. Most CMIP6 models underestimate June–August variability compared to the CRU dataset and overestimate December–February variability. In contrast, Lindvall and Svensson (2015) found that CMIP5 models slightly overestimate both June–August and December–February variabilities. Although individual models appear to be roughly consistent, the multimodel ensemble mean shows very low variability relative to the CRU dataset, likely because the interannual variabilities of different models are neutralized through averaging.

To further investigate details of interannual variability, the land surface is divided into five 30° latitudinal bands (Fig. 4), including high (60°–90°N), middle (30°–60°N), and low latitudes (0°–30°N) in the Northern Hemisphere and low (0°–30°S) and middle latitudes (30°–60°S) in the Southern Hemisphere. In a crude sense, the simulated DTR variabilities are similar to the observations; that is, variabilities in the winter are larger than those during the summer in both the Northern and Southern Hemispheres (with winter in the Southern Hemisphere corresponding to June–August). However, the models generally overestimate both the annual and seasonal DTR variabilities in these bands. Particularly, at high latitudes in the Northern Hemisphere, the models largely overestimate the annual and December–February variabilities.

c. Spatiotemporal evolution of DTR over land surfaces

The observational data show that DTR has declined about 0.25°C globally since 1901 (Fig. 5a). For comparison, the multimodel ensemble mean declined 0.30°C, overestimating the cumulative change in DTR over this period by 0.05°C or 20%. In previous studies of CMIP3 (Wild 2009b) and CMIP5 models (Lewis and Karoly 2013), the multimodel ensemble mean underestimated the DTR change relative to the observations. Among the models, MRI-ESM2.0 best reproduces the observed trend during 1901–2014 (−0.27°C) while EC-Earth3 and EC-Earth3-Veg overestimate DTR changes by 0.18°C and 0.21°C, respectively. The other models show a large spread of DTR changes, ranging from −0.38°C to −0.12°C.

A major decrease of DTR (−0.31°C) occurred in the observational data since the 1950s, implying a slight increase in DTR before the 1950s (Fig. 5a). Except for EC-Earth3, EC-Earth3-Veg, and IPSL-CM6A-LR, the CMIP6 models underestimate the changes in DTR since 1951 by 0.03°C–0.28°C. Two aspects may cause the reversal in behavior of the models before and after the 1950s in regard to DTR.

The first aspect concerns the uncertainty in the ground-based observational dataset (i.e., CRU DTR
FIG. 5. (a) Mean annual DTR anomalies during 1901–2014 for each model, the multimodel ensemble mean, the 5° × 5° gridded dataset from Thorne et al. (2016b), and the CRU dataset over the global land surface. See section 2c for details of the anomaly calculation. Anomalies are reported relative to values during 1976–2005. Changes in DTR since 1901 were calculated using the EEMD method. Cumulative changes during 1901–2014, 1951–2014, and 1981–2014 (except for Thorne et al. (2016b) because of its ending of 2012) are indicated to the right in colors corresponding to the legend. (b)–(d) As in (a), but for TASMAX, TASMIN, and RSDS, respectively.
data). The raw station data used in CRU datasets are relatively sparse prior to the 1950s, resulting in a large uncertainty in the early period in climate change diagnoses (Wang et al. 2017). However, this issue might be not too serious since the newly developed dataset (uninterpolated; Thorne et al. 2016a) also shows a similar change prior to the 1950s (Fig. 5a). Interannual variability in Thorne et al. (2016a) is larger than in CRU, mainly because of the coarse spatial resolution (5°). There are other DTR datasets available over the land surface (Berkeley Earth, HadEX2, etc.) that used different data interpolation approaches at various spatial resolutions and different spatial coverages. Focusing on their common covered area, these datasets are substantially coherent in their temporal dynamics (Thorne et al. 2016b), implying that the core data are essentially consistent. Detailed interproduct comparisons have been reported by Thorne et al. (2016a,b). We further examine a long-term global reanalysis product (ERA-20C) to see how reliable the DTR changes revealed by the observational datasets are prior to the 1950s. Changes in ERA-20C are roughly consistent with CRU and Thorne et al. (2016a) despite different interannual variability (Fig. 6). Thus, the ERA-20C product helps to confirm those changes revealed by the CRU dataset prior the 1950s. Detailed discussion about the differences between reanalysis products and observation-based products are beyond the scope of this study. Overall, the long-term change in DTR revealed by the CRU dataset should be relatively reliable. In the future, rescuing more historical data will confirm the DTR dynamics prior to the 1950s.

The second aspect relates to features missed by the models, which may be more important than the first aspect. According to the observations (Figs. 5b,c), the slight increase of DTR prior to the 1950s is mainly because there was a larger warming in TASMIN (0.095°C decade⁻¹) than in TASMAX (0.086°C decade⁻¹). However, this feature was not captured by the models (0.067°C vs 0.082°C decade⁻¹) during the same period (Figs. 5b,c). Therefore, the simulated solar radiation (RSDS) may be the most important factor causing the DTR discrepancy in the models because of its significant influence on TASMAX (Wild et al. 2007). Previous studies demonstrated that long-term RSDS records showed an increase prior to the 1950s (Wild 2009a) while the models showed a continuous decline over the entire twenty-first century (Fig. 5d). Other issues affecting DTR, including precipitation and cloud cover, might not be accurately simulated as well (Dai et al. 1997, 2006; Sun et al. 2019).

Based on the CRU observations, the rapid decline of DTR during the 1950s through 1980s (~0.05°C decade⁻¹; Fig. 5a) is mainly attributed to increasing TASMIN (0.043°C decade⁻¹; Fig. 5c) and declining TASMAX (~0.018°C decade⁻¹; Fig. 5b). The CMIP6 model ensemble shows a slower decline (~0.037°C decade⁻¹) during the same period. The primary reason is that, although the CMIP6 model ensemble showed a larger decline of TASMAX (~0.035°C decade⁻¹) than the observations, TASMIN did not show a significant change (0.002°C decade⁻¹). The large bias of TASMIN changes is probably caused by several factors, including aerosol (Huang et al. 2006), cloud (Dai et al. 1997), and urbanization effects (Wang et al. 2012, 2015), which may not be accurately simulated by the models.

Since the 1980s, no obvious changes in DTR are observed (Fig. 5a). This is because the observed changes in TASMIN and TASMAX are roughly equivalent (~0.76°C; Figs. 5b and 5c) during this period. These characteristics are not fully captured by the multimodel ensemble mean (Fig. 5a), which shows a continued decrease in DTR since the 1980s (0.08°C), although the trends are less than that since the 1950s. This continued decline in the model ensemble is mainly caused by a larger increase of TASMIN (1.30°C) relative to TASMAX (1.16°C). For the individual models, BCC-ESM1, GFDL-ESM4, MIROC6, and GISS capture the basic DTR changes since the 1980s. In contrast to the CRU observations, EC-Earth3, EC-Earth3-Veg, IPSL-CM6A-LR, and CanESM5 show large DTR declines (~0.29°C to ~0.16°C) since the 1980s.

Individual models and the ensemble mean display different spatial patterns of DTR changes compared to the observations during 1951–2014 (Fig. 7). This subperiod is selected because of the data availability in Thorne et al. (2016a). The CRU dataset indicates a change of ~0.22 ± 0.31°C over the land surface during this period. The data from Thorne et al. (2016a) show a smaller decline (~0.18°C) comparing to CRU while the spatial spread (0.36°C) is slightly larger than that in CRU (Fig. 7b). All individual models show smaller declines (up to ~0.19°C) relative to CRU. Among the individual models, the DTR changes
simulated by EC-Earth3 (−0.19° ± 0.21°C), IPSL-CM6A-LR (−0.18° ± 0.18°C), and EC-Earth3-Veg (−0.17° ± 0.17°C) are closest to that from CRU and Thorne et al. (2016a). Although EC-Earth3 shows similar statistics, the spatial pattern differs from the observed data. BCC-CSM2-MR, BCC-ESM1, and SAM0-UNICON show very low magnitudes of the DTR declines (0.00° to −0.06°C), largely because they predict unrealistic increases of DTR in the United States, Europe, and Australia (Figs. 7c,d,q). The spatial spread of DTR changes in the ensemble mean (±0.09°C) is less than one-third that of the observations, implying more uniform changes in the ensemble mean than observed over land surfaces.

Details of the spatiotemporal evolution (1951–2014) of DTR are more evident in zonally averaged changes (Fig. 8). Observations from CRU and the dataset by Thorne et al. (2016a) show major decreases in middle to high latitudes in the Northern Hemisphere since the 1950s. In contrast, the observed DTR has increased continuously since the 1950s in a zone around 10°S, and since the 1980s near 40°S.

Fig. 7. Changes of the annual mean DTR (°C) from the onset (1951–80) to the most recent period of analysis (1985–2014). Due to the time period available in the Thorne et al. (2016b) dataset, the recent period in (b) is 1985–2012.
Individual models show substantially different changes over the land surface and do not capture the signs or magnitudes of the observed changes (Figs. 8c–q). Overall, EC-Earth3-Veg (Fig. 8g) is the best at reproducing the observed spatiotemporal evolution (NRMSD = 0.73). GISS-E2.1-GCC (Fig. 8m) presents a similar pattern to the observations in most zones of the Northern Hemisphere, but does not simulate the observed increase in the Southern Hemisphere.
In the 1915-2000 period, intermodel correlation coefficients were calculated for changes in seasonal mean DTR and seasonal means of the other variables. Figure 9 shows the correlations for each climate subregion, with equatorial subregions shaded gray. Correlation coefficients significant at the 95% confidence level are indicated by large circles; values of 0.5 or -0.5 are statistically significant at the 95% confidence level based on Student's t-test.

**d. Factors influencing DTR trends in CMIP6 models**

To examine which variables potentially important for the diurnal cycle may explain the DTR changes over the last century, intermodel correlation coefficients were calculated from the different model responses for DTR and the other analyzed variables (section 2c) between 1915 (±14 years) and 2000 (±14 years). Figure 9 shows the correlations for each of the climate subregions on a seasonal basis.
Based on the correlations, TASMIN differences between the models contribute significantly to the intermodel differences in DTR in most Northern Hemisphere subregions during the autumn and winter (Figs. 9c,d). This implies that the models agree that DTR decreases with increasing TASMIN. In contrast, significant contributions of the TASMAX differences are mainly found in southern South America, southern Africa, and Australia (Fig. 9). The intermodel correlations between TASMAX and DTR changes are positive in these three subregions, meaning that models that warm the most show larger increases in DTR, whereas models with slight increases or decreases in TASMAX show smaller increases or marked decreases in DTR.

Unsurprisingly, RSDS and SWCRE are positively correlated with changes in DTR in most subregions across the Northern Hemisphere throughout the seasons (Fig. 9), similar to the observations (Wild 2009a; Wang and Dickinson 2013) because RSDS and SWCRE primarily increase TASMAX. Conversely, changes in RLDS and LWCRE are mainly negatively correlated with changes in DTR (Fig. 9) because RLDS and LWCRE mainly increase nighttime temperatures (i.e., TASMIN). Meanwhile, intermodel differences in precipitation changes are negatively correlated with DTR in several subregions; it is important to note, however, that precipitation is partly correlated with cloud and radiation components. Changes in the evaporative fraction (EF) mainly influence DTR changes during June–August (Fig. 9b). At middle to high latitudes of the Northern Hemisphere, EF is not correlated with DTR during December–February (Fig. 9d).

Furthermore, we find that the factors influencing intermodel TASMAX differences are much more complex than that for TASMIN (Figs. 10 and 11). In most seasons, downwelling longwave radiation (RLDS) is the most important factor explaining the intermodel TASMAX differences as the correlations between RLDS and TASMIN are significantly positive in most subregions (Fig. 11). This may explain why TASMIN contributes more significantly to the intermodel DTR differences than TASMAX. RLDS is important for intermodel differences in TASMAX as well (Fig. 10). However, the correlations between TASMIN and RLDS (Figs. 11c,d) are generally larger than that between TASMAX and RLDS (Figs. 10c,d), particularly in winter, which may be important for explaining the negative correlation between RLDS and DTR (Figs. 9c,d).

Several anthropogenic factors that the CMIP6 models may not simulate well may affect DTR, such as urbanization (Lewis and Karoly 2013), which also influences in situ measurements (Karl et al. 1993; Wang et al. 2014). A major impact of rapid urbanization may be changes to the surface energy balance due to anthropogenic aerosol changes (Lewis and Karoly 2013). With increased urbanization, urban boundaries have encroached on meteorological stations, resulting in substantial heat island impacts on surface temperature measurements (Wang et al. 2015). Factors contributing to heat-island effects include changes in vegetation, surface albedo, surface roughness, soil moisture, and the porosity of the surface, all of which affect daytime and nighttime temperatures over urban areas (Zhao et al. 2014; Bounoua et al. 2015; Chen and Frauenfeld 2016). A more complete understanding of natural and anthropogenic effects on DTR will therefore remain a challenge for climate models.

e. How much have the models improved from CMIP5 to CMIP6?

During the entire period of CMIP5–CMIP6 overlap (1901–2005), four CMIP6 models (CanESM5, GFDL-ESM4, GISS-E2.1-H, and MRI-ESM2.0) improved their ability to replicate the observed CRU DTR changes relative to their earlier CMIP5 counterparts (Fig. 12a). The most significant improvement was from CanESM2 to CanESM5, narrowing the gap between observed and simulated DTR changes by about 88% (0.17°C to −0.02°C). Three models (GFDL-ESM4, GISS-E2.1-H, and MRI-ESM2.0) improved 20%–40%. In contrast, the ability of the remaining models (BCC-CSM2-MR, FGOALS-g3, GFDLCM4, IPSL-CM6A-LR, and MIROC6) to simulate DTR changes that deteriorated between CMIP5 and CMIP6, at least when examining the entire overlap period 1901–2005 (gray areas, Fig. 12a). However, whether a model improved, or not, depends on the time period examined (Figs. 12b–d).

Over the first half of the twentieth century (1901–50), nearly all the CMIP6 models performed worse at simulating DTR changes than their CMIP5 counterparts (Fig. 12b). The only two models showing any improvement during this early period are MRI-ESM2.0 and BCC-CSM2-MR. The situation is reversed during the following three decades (1951–80) with seven of the CMIP6 models performing better than their CMIP5 counterparts (Fig. 12c). The two models showing a deterioration during this time are MRI-ESM2.0 and GFDL-CM4. Finally, during 1981–2005 two CMIP6 models (MRI-ESM2.0 and PSL-CM6A-LR) showed a 40%–50% improvement compared with their CMIP5 versions (Fig. 12d). CanESM5, GFDL-CM4, GFDL-ESM4, GISS-E2.1-H, and MIROC6 models did not show any significant differences (< ±0.02°C) with their CMIP5 counterparts during this time. In short, we confirm that these models do not keep improving, at least not in a consistent way, over the entire time period.
An interesting case is a comparison of the CMIP5 and CMIP6 versions of CanESM. Among the nine CMIP5 models, CanESM2 predicts the best DTR trend over the early period 1901–50 while CanESM5 overpredicts the DTR change during this time. However, small counteracting effects during the later period (1951–2005) leads CanESM5 to have the best DTR trend of the CMIP6 models over the entire period (1901–2005). The differences between the old and new versions of CanESM during 1901–50 might be attributed to several reasons, including 1) incorrect energy balance components or CMIP forcings of the climate system (Wild 2009a; Swart et al. 2019), 2) higher sensitivity in the model than the world (Zelinka et al. 2020; Swart et al. 2019), and/or 3) potential flaws or bugs in the model source codes (Swart et al. 2019). These issues probably exist in other models as well. A detailed investigation for each model counterpart may be necessary in future diagnoses.

Fig. 10. As in Fig. 9, but for TASMAX.

Unauthenticated | Downloaded 09/21/22 07:02 AM UTC
4. Conclusions

The ability of 15 CMIP6 models to simulate the climatology, interannual variability, and spatiotemporal evolution of DTR were comprehensively investigated over the global land surface during 1901–2014 using the CRU high-resolution datasets as a benchmark. HadGHCND and new data by Thorne et al. (2016a) were used as auxiliary benchmarks. The ensemble empirical mode decomposition (EEMD) method, instead of a traditional linear trend analysis, was used to investigate overland DTR changes and an intermodel correlative analysis was used to reveal potential causes of the large discrepancies among models. Based on this comprehensive comparison, the main findings are the following:

- CMIP6 participant models generally underestimate DTR climatology relative to the observations, consistent with previous findings from CMIP5 models. Most individual models overestimate December–February

Fig. 11. As in Fig. 9, but for TASMIN.
interannual variability, particularly at high latitudes of the Northern Hemisphere.

- Individual models show substantially different over-land DTR changes and do not fully capture the signs or magnitudes of the observations. Cumulative DTR changes (1901–2014) over the global land surface from the multimodel ensemble are nonetheless similar to the observations. However, the increasing DTR trend observed prior to the 1950s is not captured by the models. Furthermore, DTR simulated by the models continues to decline after the 1980s, whereas the observations reveal that DTR has been relatively stable during the past few decades.

- Large intermodel differences in DTR may be controlled by daily minimum air temperature. Downwelling longwave
radiation is a critical factor for explaining the intermodel discrepancy in daily minimum air temperature. In the Northern Hemisphere, precipitation and cloud radiative effects appear to have important contributions to the intermodel discrepancies. Evaporative fraction is an important factor contributing to the intermodel differences during the summer in the Northern Hemisphere.

- The models do not keep improving, at least not in a consistent way, over the entire historical time period. The differences between CMIP5 and CMIP6 models varies with the time period considered. The models have shown a major improvement during the period of rapid DTR decline (i.e., 1951–80).

Overall, DTR is highly sensitive to slight changes in both minimum and maximum temperatures. Changes in cloud cover, radiation, and land surface water balances caused by natural and/or anthropogenic factors play important roles in DTR dynamics. Further evaluations of CMIP6 models will prove useful for identifying model strengths and issues in simulating DTR-related processes. Finally, it is important to note that ground-based observations also contain uncertainties, including measurement biases, sparse station distributions, and station representativity issues. Rescuing more historical data is necessary to confirm the DTR dynamics prior to the 1950s.

Acknowledgments. We thank two anonymous reviewers and the editor, Dr. Xin-Zhong Liang, for their insightful comments, which helped to improve the manuscript. We appreciate all organizations and individuals listed in Table 1 for implementing and making available their models and the Program for Climate Model Diagnosis and Intercomparison (PCMDI), Lawrence Livermore National Laboratory (LLNL) for archiving the model outputs of CMIP5 (https://esgf-node.llnl.gov/search/cmip5/) and CMIP6 (https://esgf-node.llnl.gov/search/cmip6/). CRU TS v4.03 is available at https://crudata.uea.ac.uk/cru/data/ctv/cru_ts_4.03/cruTs.190501.1326.v4.03/. Data by Thorne et al. (2016a) are available at https://www.maynoothuniversity.ie/icarus/icarus-data. HadGHCND is available at https://www.metoffice.gov.uk/hadobs/hadghcnd/. EEMD decomposition was implemented using “libeemd” (https://bitbucket.org/luukko/libeemd/) and its Python interface “pyeemd” (https://bitbucket.org/luukko/pyeemd/). The raw land mask (small-scale data, 1:10000000 scale) was obtained from Natural Earth (https://www.naturalearthdata.com/http//www.naturalearthdata.com/download/110m/physical/ne_110m_land.zip). This study was supported by the National Natural Science Foundation of China (Award 41701064) and the U.S. National Science Foundation (Awards 1503559 and 1831623). Python analysis codes are available from the authors upon request. The authors declare that there are no conflicts of interest.

REFERENCES


