CrossMark

Evaluation of simulated climatological diurnal temperature range in CMIP5 models from the perspective of planetary boundary layer turbulent mixing

Nan Wei[1,2] · Liming Zhou[2] · Yongjiu Dai[1]

Abstract This study examines the effects of modeled planetary boundary layer (PBL) mixing on the simulated temperature diurnal cycle climatology over land in 20 CMIP5 models with AMIP simulations. When compared with observations, the magnitude of diurnal temperature range (DTR) is systematically underestimated over almost all land areas due to a widespread warm bias of daily minimum temperature (T_{min}) and mostly a cold bias of daily maximum temperature (T_{max}). Analyses of the CMIP5 multi-model ensemble means suggest that the biases of the simulated PBL mixing could very likely contribute to the temperature biases. For the regions with the cold bias in T_{max}, the daytime PBL mixing is generally underestimated. The consequent more dry air entrainment from the free atmosphere could help maintain the surface humidity gradient, and thus produce more surface evaporation and potentially lower the T_{max}. The opposite situation holds true for the regions with the warm bias of T_{max}. This mechanism could be particularly applicable to the regions with moderate and wet climate conditions where surface evaporation depends more on the surface humidity gradient, but less on the available soil moisture. For the widespread warm bias of T_{min}, the widely-recognized overestimated PBL mixing at nighttime should play a dominant role by transferring more heat from the atmosphere to the near-surface to warm the T_{min}. Further analyses using the high resolution CFMIP2 output also support the CMIP5 results about the connections of the biases between the simulated turbulent mixing and the temperature diurnal cycle. The large inter-model variations of the simulated temperature diurnal cycle primarily appear over the arid and semi-arid regions and boreal arctic regions where the model differences in the PBL turbulence mixing could make equally significant contributions to the inter-model variations of DTR, T_{max} and T_{min} compared to the model differences in surface radiative processes. These results highlight the importance and need for accurate descriptions of the PBL processes with respect to the turbulent mixing in order to improve the temperature diurnal cycle simulations in climate models.

1 Introduction

The land surface air temperature (LSAT) has a characteristic diurnal cycle, which represents the most basic forms of climate patterns. It oscillates with the diurnal variation of solar radiation and other drivers between a minimum (T_{min}) before sunrise and a maximum (T_{max}) in the afternoon (e.g. Betts and Barr 1996), which thus generates an evident diurnal temperature range (DTR, DTR = $T_{max} - T_{min}$). T_{max}, T_{min} and DTR have been routinely considered as key meteorological variables in quantifying weather and climate change (IPCC 2007; Lobell et al. 2007). Realistically simulating the diurnal cycle of temperature and other

✉ Liming Zhou
lzhou@albany.edu

1 College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China

2 Department of Atmospheric and Environmental Sciences, University at Albany, State University of New York, 1400 Washington Avenue, Albany, NY 12222, USA

near-surface variables [e.g. evaporative fraction (EF), surface wind] are of great importance in weather and climate models as diurnal evolutions of these variables are tightly related to many hydrologic, biogeochemical and ecological processes (e.g. Zhang and Zheng 2004; Yi et al. 2010; Gentine et al. 2011; Shao et al. 2011; Hua et al. 2013, 2014).

The diurnal cycle of LSAT is primarily maintained by daytime solar heating and nighttime radiative cooling, along with the partitioning of net surface radiation between latent and sensible heat flux. Many factors related to these processes can determine the magnitude of DTR such as cloud cover, precipitation, land surface properties controlled by vegetation and soil moisture (SM), and radiative forcings due to increased greenhouse gases (GHGs) and aerosols (Gallo et al. 1996; Collatz et al. 2000; Bonan 2001; IPCC 2007; Zhou et al. 2009a, b, 2010). They all affect DTR through altering the land surface energy and hydrological balances. For example, increasing cloud cover greatly reduces T_{max}, and thus DTR, by reducing surface solar insolation, while it has a relatively small influence on T_{min} because its nighttime greenhouse warming effect tends to offset its daytime cooling effect on afternoon temperatures (Dai et al. 1999; Lobell et al. 2007; Zhou et al. 2008). Increasing SM limits the rise of T_{max} and T_{min} asymmetrically, and thus reduces DTR by lowering the Bowen ratio, increasing soil heat capacity (Dai et al. 1999; Zhang et al. 2009), and changing soil properties such as soil albedo and emissivity (Zhou et al. 2007, 2008).

The diurnal cycle of LSAT is also tightly connected to the extent of turbulent mixing in the atmosphere, which is described by the depth of planetary boundary layer (PBL) (Dirmeyer et al. 2013; Davy and Esau 2014a, b). As the lowest part of the atmosphere, the PBL necessarily impacts the diurnal evolution of near-surface thermodynamic variables through turbulent exchanges of momentum, heat and moisture, which are controlled by the nature of underlying land surface and its coupling with the PBL (Zhou et al. 2008, 2010; Pithan et al. 2015). One of the major interactive mechanisms that determine the magnitude of DTR could be derived from the effects of the diurnal evolution of PBL depth. During the daytime, the strong upward PBL mixing is buoyancy driven. The simultaneous dry air entrainment from the free atmosphere to the PBL can modify the humidity within the PBL and thus in turn regulate surface heat fluxes and temperatures (Van Heerwaarden et al. 2009; Svensson et al. 2011). In general, the air humidity decreases more rapidly with height at lower atmosphere. Thus, a relatively weaker turbulent mixing with a shallower convective PBL experiences a larger vertical humidity gradient at the top of PBL, leading to more dry air entrainment that reduces the PBL moisture content (Dirmeyer et al. 2013). As a result, the humidity gradient at the bottom of the well-mixed PBL is more effectively maintained, and

the surface latent heat flux is thus enhanced to adequately replenish the PBL moisture content, which eventually moistens the PBL and potentially reduces T_{max} (Dirmeyer et al. 2013). The opposite situation is true for a stronger turbulent mixing within a deeper convective PBL. During the nighttime, the land surface radiative cooling makes the LSAT drop to the minimum, along with a stably-stratified PBL. In this case, the turbulent mixing with the warm air above is not buoyancy-driven, but by wind shear and also modulated by surface roughness (Lindvall and Svensson 2015). A stronger downgradient turbulent diffusion is associated with a deeper PBL and a larger downward sensible heat flux, which makes the nighttime temperature decline slower. Hence T_{min} could potentially be increased by more heat transport from the above atmosphere, and vice versa (Holtslag et al. 2013; Lindvall et al. 2013; Bosveld et al. 2014). The close connections of T_{max} and T_{min} to the PBL depth indicate a tight dependence of DTR on the diurnal evolution of PBL mixing, which strongly suggests, together with other related effects, that the PBL mixing is not just a response to the temperature diurnal cycle as indicated in previous studies (e.g. Svensson et al. 2011; Bosveld et al. 2014; Lindvall and Svensson 2015), but could also be a feedback to the cycle through continuous interactions with the land surface (Holtslag et al. 2013).

In numerical models, any biases or deficiencies on processes related to the land surface energy budget (e.g. the PBL turbulent mixing) could manifest themselves as biases in the temperature diurnal cycle (Svensson and Lindvall 2015). In fact, the turbulent mixing in the PBL remains unresolved in all weather and climate models (Holtslag et al. 2013). Any set of equations that can be derived from first principles of physics to describe the PBL turbulence contains more unknowns than equations and thus cannot be solved, namely known as the turbulence closure problem. Thus, the PBL mixing associated with the vertical turbulent fluxes has to be parameterized for a given profile of temperature, moisture and wind (Holtslag et al. 2013; Pithan et al. 2015). The diurnal cycle of LSAT has been shown to be very sensitive to PBL parameterizations in mesoscale models (e.g. Zhang and Zheng 2004; Steeneveld et al. 2008; Kleczek et al. 2014). Shin and Hong (2011) compared five PBL schemes in the weather research and forecasting (WRF) model, and found that the non-local scheme with the entrainment flux proportional to the surface flux was more favorable in simulating temperatures and PBL structure under convective conditions. Sandu et al. (2013) examined the sensitivity of the European Centre for Medium-Range Weather Forecasts (ECMWF) model to the formulation of turbulent diffusion under stable conditions, and found that a reduced diffusive turbulent scheme could substantially change the global patterns of large-scale flow and LSAT. The global atmospheric boundary layer studies

(GABLS) model inter-comparisons have particularly isolated the PBL scheme's performance from other physical processes in clear-sky conditions, and showed that current turbulence parameterizations had great difficulties in representing real PBL mixings and the diurnal cycle of near-surface variables compared to in situ observations (Cuxart et al. 2006; Holtslag et al. 2013), with too little turbulence and a shallower PBL during the convective daytime and too much turbulence and a deeper PBL during the stably-stratified nighttime (Cuxart et al. 2006; Esau and Ziltinkevich 2010; Svensson et al. 2011; Holtslag et al. 2013). The percentile bias of the modeled PBL depth throughout the whole diurnal cycle could vary from 50 to 100 % (Svensson et al. 2011; Holtslag et al. 2013; Bosveld et al. 2014).

Modelers always hold different standpoints on the complexity of atmospheric turbulence and vertical mixing, and hence conduct different implementations of PBL schemes in practice (Svensson et al. 2011; Holtslag et al. 2013; Bosveld et al. 2014; Svensson and Lindvall 2015). Consequently the capabilities in simulating turbulent mixing and other PBL processes vary substantially among models and such uncertainties could lead to striking inter-model variations of the temperature diurnal cycle under some particular conditions (IPCC 2007, 2013; Holtslag et al. 2013). For example, Holtslag et al. (2013) and Svensson and Lindvall (2015) evaluated the diurnal cycle of LSAT in the state-of-the-art climate models participating in the Coupled Model Intercomparison Project phase 5 (CMIP5), and showed that the relatively larger inter-model variations appeared in the afternoon and early morning when T_{max} and T_{min} occurred, corresponding respectively to the dominantly strong convective and stable conditions that could differ notably among CMIP5 models. Sterk et al. (2013) presented a nonlinear behavior regarding the effects of the different degrees of the simulated turbulent mixing on LSAT in the Arctic stable PBL. Viterbo et al. (1999) and Holtslag et al. (2013) found that, even implementing slightly different stability functions for the PBL scheme of the ECMWF model in stable conditions, the LSAT still exhibited large differences even with unchanged forcing conditions. This suggests a strong sensitivity of LSAT to PBL schemes which could in practice differ substantially. GABLS2 examined the performances of 30 single column models with different PBL schemes, which produced various results in the strength of capping inversion, entrainment and the diurnal cycle of turbulent kinetic energy and PBL depth, although a prescribed surface temperature was used (Svensson et al. 2011; Holtslag et al. 2013). Hu et al. (2010) also compared three different PBL schemes in the WRF model, and pointed out that their differences concentrated on the vertical mixing strength and the entrainment of air from above the PBL. These uncertainties and difficulties in modeling the turbulent mixing and other PBL processes could

definitely contribute to the inter-model differences of the diurnal cycle of LSAT.

The diurnal cycle of LSAT has been examined in global climate models (GCMs) as useful diagnostic indices for model evaluations and climate projections (Kharin et al. 2007; Wild 2009). The majority of previous studies have mainly focused on the variability or long-term trends of T_{max}, T_{min} and DTR (e.g. IPCC 2007, 2013), while there are relatively few studies on analyzing their climatological biases and inter-model variations. The CMIP3 and CMIP5 models have shown to be capable of simulating general spatial patterns of climatological DTR as well as T_{max} and T_{min}, and qualitatively capturing their basic statistical relationships with various controlling factors (Zhou et al. 2007, 2008, 2009b, 2010; Dirmeyer et al. 2013; IPCC 2007, 2013; Lindvall and Svensson 2015). However, the modeled DTR was much weaker than observations in many regions by as much as 50 %, meaning that the simulated LSAT did not change enough between daytime and nighttime although the diurnal mean value matched observations much better (IPCC 2007; Cattiaux et al. 2013; Sillmann et al. 2013). In addition, most of the regions with large DTR biases also exhibited large inter-model variations (Holtslag et al. 2013; Lindvall et al. 2013; Svensson and Lindvall 2015).

The present paper focuses mostly on the global scale effects of the simulated PBL climatology on the diurnal cycle of LSAT in GCMs, while local and synoptic scale analyses on selected cases with mesoscale models have been performed in previous studies. Here we propose two hypotheses to explain the biases and uncertainties of the temperature diurnal cycle over land in CMIP5 models. First, the deficiencies of the modeled PBL mixing in CMIP5 models are at least partially responsible for the DTR biases over most of continental areas. To be specific, at daytime, CMIP5 models probably underestimate the PBL mixing over the regions with underestimated T_{max}, and thus overestimate dry air entrainment and surface latent heat, leading to a cold bias. The opposite situation could apply to the regions with overestimated T_{max}. At nighttime, the models likely overestimate the PBL mixing and the downward heat transfer from the atmosphere to the ground, leading to a warm bias of T_{min}, or vice versa. Second, the model uncertainties in PBL schemes with respect to turbulence mixing could make a significant contribution to the large inter-model variations of DTR over the regions with extreme diurnal PBL conditions (i.e., strong convective and stably-stratified conditions). This may be the case given the strong dependence of DTR simulations on the PBL mixing and distinct performances among PBL schemes in extreme PBL conditions whose turbulent mixing is poorly modeled. We will test these two hypotheses by comparing the CMIP5 metrics related to the PBL mixing with observations and reanalysis. As generating the global scale PBL climatology

from observations are extremely difficult (Lewis and Karoly 2013; Davy and Esau 2014b), we establish the global pseudo-observations for several key PBL variables from the ECMWF Interim (ERA-Interim) reanalysis.

Unlike most previous studies based on the fully-coupled models that cannot reproduce observed sea surface temperatures (SST) variations (e.g. Hoerling et al. 2010), we select 20 CMIP5 models with Atmospheric Model Intercomparison Project (AMIP) runs forced by observed SST and sea ice to minimize SST-related uncertainties in simulating the LSAT. Given the limits of availability and accuracy of the CMIP5 output and observations/reanalysis, we concentrate mainly on the multi-model ensemble mean biases and inter-model variations of the climatology in DTR, T_{max} and T_{min}, and their connections with the PBL depth and other relevant metrics representing the PBL mixing and land–atmosphere interactions. We further take advantage of the model output in experiments from the second phase of the Cloud Feedback Model Intercomparison Project (CFMIP2), which provide data with more vertical levels and can be used to validate the CMIP5 results. This work focuses more on the possible causality rather than

the formal detection-attribution framework, but the results could provide useful information and potential guidance on improving the modelling of the diurnal cycle of LSAT.

2 Data and methodology

2.1 CMIP5-AMIP and CFMIP2-AMIP simulations

Table 1 lists the 20 CMIP5 models with AMIP simulations analyzed in this study, along with the model centers and resolutions of atmospheric components. The model selection is mainly made based on the availability of data and the output frequency, which must be fully retrievable from the Program for Climate Model Diagnosis and Intercomparison (PCMDI) data server. In general, different versions of models from same centers produce different results due to their updated atmospheric and/or land components, thus more than one version of the models such as ACCESS series are used here. However, for those models sharing similar atmospheric and land components, only one model with the smallest global mean DTR bias is

Table 1 List of CMIP5 models analyzed and their horizontal and vertical resolutions

Model	Institute	Atm. resolution[a]
ACCESS1.0	Commonwealth Scientific and Industrial Research Organization and Bureau of Meteorology, Australia	192 × 145, L38
ACCESS1.3	Commonwealth Scientific and Industrial Research Organization and Bureau of Meteorology, Australia	192 × 145, L38
BNU-ESM	College of Global Change and Earth System Science, Beijing Normal University, China	128 × 64 (T42), L26
CanAM4	Canadian Climate Centre for Modelling and Analysis, Canada	128 × 64 (T42), L35
CCSM4	National Centre for Atmospheric Research, USA	288 × 192, L26
CESM1-CAM5	National Centre for Atmospheric Research, USA	288 × 192, L30
CMCC-CM	Centro Euro-Mediterraneo per i Cambiamenti, Italy	480 × 240 (T159), L31
CNRM-CM5	Centre National de Recherches Meteorologiques and Centre Europeen de Recherches et de Formation Avancee en Calcul Scientifique, France	256 × 128 (T127), L31
CSIRO-Mk3.6.0	Commonwealth Scientific and Industrial Research Organization and the Queensland Climate Change Centre of Excellence, Australia	192 × 96 (T63), L18
EC-EARTH	European Earth System Model, Europe	320 × 160 (T159), L62
FGOALS-s2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, China	128 × 108, L26
GFDL-CM3	NOAA Geophysical Fluid Dynamics Laboratory, USA	144 × 90, L48
GFDL-HIRAM-C360	NOAA Geophysical Fluid Dynamics Laboratory, USA	1152 × 720, L32
GISS-E2-R	NASA Goddard Institute for Space Studies, USA	144 × 90, L40
HadGEM2-A	Met Office Hadley Centre, UK	192 × 144, L38
INM-CM4	Institute for Numerical Mathematics, Russia	180 × 120, L21
MIROC5	Atmosphere and Ocean Research Institute, The University of Tokyo, National Institute for Environmental Studies, Japan Agency for Marine-Earth Science and Technology, Japan	256 × 128 (T85), L40
MPI-ESM-LR	Max Planck Institute for Meteorology, Germany	192 × 96 (T63), L47
MRI-AGCM3.2S	Meteorological Research Institute, Japan	1920 × 960, L64
MRI-CGCM3	Meteorological Research Institute, Japan	320 × 160, L48

[a] Model resolution is characterized by the size of a horizontal grid on which output is available from the model's atmospheric component and by the number of vertical levels. Spectral models are also characterized by their spectral truncations in parenthesis

selected to enhance sampling independence. For instance, CCSM4, GFDL-HIRAM-C360, MPI-ESM-LR and MRI-AGCM3.2S are selected in our analysis, while the corresponding similar models NorESM1-M, GFDL-HIRAM-C180, MPI-ESM-MR and MRI-AGCM3.2H are not used. For each model, only the first ensemble member denoted as "r1i1p1" (typically the first realization, initialization, and set of perturbed physics for each model) is chosen. Due to the huge volume of high temporal resolution data, we limit our analyses to monthly mean model output over a 30-year period (1979–2008), except the surface sensible heat flux which is stored as the three-hourly average to approximate the degrees of PBL mixing within a diurnal cycle. All variables are spatially remapped onto a $1° \times 1°$ grid box using the local area-conservative binning method. As GCMs are diagnosed to perform badly over steep orography no matter which parameterizations are implemented (IPCC 2007, 2013; Lindvall and Svensson 2015), the regions with an altitude >2 km are filtered out. We choose this 2-km threshold because it is assumed as the characteristic height for mountainous regions being exposed to the free atmosphere, and the global mean bias of DTR is insensitive to a 20 % variation in this threshold following the sensitivity test of Davy and Esau (2014a, b). In total, 14,466 grid boxes over land are considered in this study.

For some of the models (CanAM4, CNRM-CM5, HadGEM2-A, MPI-ESM-LR and MRI-CGCM3), the AMIP simulations for the CFMIP2 experiments are available. As the coarse vertical resolution of CMIP5 output prevents us from directly estimating the PBL depth, we use the CFMIP2 data with more model levels. Note that the CFMIP2 archive provides the quantities needed for PBL depth estimation only in the diurnal means. In addition, the CFMIP2 experiments also supply high-frequency (three-hourly) model output for certain locations. One of these locations, the Atmospheric Radiation Measurement Program Southern Great Plains (ARM-SGP) site, located in Lamont, Oklahoma (36.68°N, 97.58°W), has high-resolution observations and thus is chosen to validate the modeled entire diurnal cycle of LSAT and PBL depth.

2.2 Observations, satellite and reanalysis data

We use the global gridded surface air temperatures (T_{max}, T_{min}, and DTR) from the historical monthly dataset "CRU TS3.23" produced by the Climate Research Unit to evaluate the simulated temperatures in CMIP5 models. This dataset is calculated on a high-resolution ($0.5° \times 0.5°$) grid from over 4000 weather stations worldwide covering the period 1901–2014 (Mitchel and Jones 2005). To maintain the spatiotemporal consistency with the CMIP5 output, we extract the temperature time series over the period 1979–2008 from the original dataset and interpolate it to $1° \times 1°$

resolution as the models are remapped. The data over Antarctica and Greenland have large uncertainties due to the limited availability of metrological stations (Wang and Zeng 2013), and thus are excluded from our analyses. In addition, the CRU cloud cover data is similarly processed to examine the cloud effects on the modeled temperature biases.

The CFMIP2-simulated diurnal cycles of the LSAT and PBL mixing at the ARM-SGP site are compared against the corresponding observations—the ARM Best Estimate dataset (ARMBE). The ARMBE product provides a range of measured atmospheric and land surface variables for the period 1994–2012 (Xie et al. 2010). Although the period covered by the data is shorter than that of the CFMIP2, the differences in climatology due to the different temporal coverage are significantly smaller than the regional-scale model biases. To be consistent with the temporal resolution of the model output, we average the ARMBE from hourly to three-hourly data.

We use the Clouds and Earth's Radiant Energy System (CERES) energy balanced and filled (EBAF) surface products of downward shortwave and longwave radiation from 2000 to 2008 to evaluate the CMIP5-modeled surface radiation. The CERES-EBAF data is of high quality and has proved better than most reanalysis products (e.g. Wang and Dickinson 2013; Rutan et al. 2014; Zhang et al. 2016). The global gridded observations of precipitation from the Global Precipitation Climatology Project (GPCP) Version 2.2 dataset for the period 1979–2008 are also used to assess the modeled climatological precipitation bias.

The ERA-Interim reanalysis provides $1° \times 1°$ grid data for various land surface variables and atmospheric vertical profiles covering the period 1979–2008, which could be used for global comparison with the simulated PBL mixing in CMIP5/CFMIP2 models. The climatological diurnal means are analyzed for all the variables, except the surface sensible heat flux and the PBL depth whose three-hourly synoptic monthly means are used. For the CMIP5 models, the simulated PBL mixing is indirectly evaluated by analyzing its several proxies (see next section for more details). The vertical profiles of atmospheric temperature and humidity, near-surface air temperature and dew point temperature, and surface turbulent heat fluxes from ERA-Interim are used to calculate these proxies as the pseudo-truth. The ERA-Interim vertical profiles and near-surface temperatures have been assimilated with various observations, and proven to be of high quality (e.g. Simmons et al. 2010; Dee et al. 2011; Decker et al. 2012; Engeln and Teixeira 2013). Although not assimilated, the surface turbulent heat fluxes had the fairly accurate monthly averaged diurnal cycle compared to in situ observations, and the best variability of the six-hourly sensible heat flux among six commonly-used reanalysis products (Decker et al. 2012).

Therefore, it is reasonable to use these proxies estimated from the reanalysis as pseudo-truth even over Africa and boreal regions where observations are limited for model validation. For the CFMIP2 models, the PBL depth is directly estimated following the algorithm in Vogelezang and Holtslag (1996). The corresponding vertical gradients of atmospheric temperature, relative humidity, wind speed and geopotential height profiles in ERA-Interim, which are all assimilated with observations from a world-wide Radio-sonde network (Dee et al. 2011), are also used to calculate the pseudo-observed PBL depth with the same algorithm. Note that the PBL height in the ERA-Interim archive is a model-derived parameter (not assimilated) and thus not chosen as the pseudo-truth to validate the CFMIP2-estimated PBL height. Instead, we simply use it as the reference metric to define geospatial patterns of PBL mixing levels from daytime to nighttime to qualitatively analyze spatial characteristics of the inter-model variations of the temperature diurnal cycle.

2.3 Proxies for the PBL mixing

Among the variables analyzed in this study, those representing the PBL mixing need further clarification. The PBL depth should be the direct quantification of the degree of the turbulent mixing, but it is not a standard CMIP5 output variable. As the CMIP5 output with limited vertical levels (only five layers below 500 mb) can hardly depict any PBL process, we use some proxies for the PBL depth and other indices to represent the modeled PBL mixing.

During the daytime, T_{max} appears with the diurnal maximum PBL depth (PBL_{max}) (Figure SI1). Here we use the diurnal maximum of surface sensible heat flux (SH_{max}) as a proxy for the PBL_{max} because SH_{max} represents the maximum level of energy that induces PBL growth. Although SH_{max} may appear slightly earlier than PBL_{max} over some regions (Figure SI1) due to an earlier formed maximum of the near-surface temperature gradient at noon, their strong positive correlation (Figure SI2) ensures that SH_{max} could be a good proxy for PBL_{max}. Since the PBL growth enables the formation of clouds by mixing moist air upward to the height where temperature and pressure allow condensation to occur, the lifting condensation level (LCL) could be another proxy for the PBL depth (Betts 2004; Dirmeyer et al. 2014) which is defined as

$$LCL = (T_{2m} - T_{D2m})/(\Gamma_{Dry} - \Gamma_{Dew})$$
$$\approx 125(T_{2m} - T_{D2m}), \qquad (1)$$

where T_{2m} and T_{D2m} are 2-m air temperature and dewpoint temperature, Γ_{Dry} and Γ_{Dew} are dry adiabatic and dewpoint lapse rate. A deep PBL and a high LCL are both favored in hot and dry conditions while neither of them can be deeply developed in cool and moist conditions. As the

diurnal mean LCL is dominated by daytime values, which is an indicator of how much the PBL must grow to trigger convection (Dirmeyer et al. 2013; Svensson and Lindvall 2015), we use the diurnal mean LCL here as another proxy for PBL_{max}. In addition, we estimate the diurnal mean value of the Priestley–Taylor coefficient (PTC) based on the formulation of Dirmeyer et al. (2013):

$$PTC = \left(\frac{\lambda E}{SH + \lambda E}\right)\left(\frac{1 + \varepsilon}{\varepsilon}\right), \quad \varepsilon = \frac{\lambda}{C_P}\frac{dq}{dT}\bigg|_{T_{LCL}}, \qquad (2)$$

where SH and λE are the sensible heat and latent heat respectively, and the thermodynamic coefficient ε is related to the change of saturation mixing ratio with the temperature at the LCL (Betts 2004). PTC is a measure of the surface evaporation efficiency into the PBL, which is determined by dry-air entrainment at the top of the growing daytime PBL and surface/near-surface properties (e.g. aerodynamic resistance and stomatal conductance) (Dirmeyer et al. 2013). Although the surface/near-surface properties regulate PTC by changing the EF, Van Heerwaarden et al. (2009) and Dirmeyer et al. (2013) have both showed that PTC is a good indicator of the enhancement of evaporation by dry-air entrainment under all surface conditions.

During the nighttime, the PBL climatology is controlled by nearly constant stably-stratified conditions, which make the PBL state vary little during the whole night (e.g. Liu and Liang 2010; Svensson and Lindvall 2015). Thus, we use the mean PBL height averaged from midnight to early morning (PBL_{night}) to quantify the degree of the nighttime turbulent diffusion which affects the magnitude of T_{min}. As the sensible heat flux at nighttime (negative value) represents the amount of the downward heat transfer along with the downgradient diffusion in a stable PBL, we use the mean sensible heat flux also averaged from midnight to early morning (SH_{night}) as the proxy for PBL_{night}. Their general negative correlation further justifies our rationality for using this proxy (Figure SI2). Note that the nighttime surface turbulent momentum flux or friction velocity, which drive the nighttime turbulence, are not available in the CMIP5 archive, and hence cannot be analyzed.

The CFMIP2 data with more model levels allow us to estimate the PBL depth using the atmospheric vertical gradients following the algorithm in Vogelezang and Holtslag (1996) (referred to as the VH method hereafter). The VH method is a bulk-Richardson number method, which has been evaluated by Seidel et al. (2012) as the best method for climatological analysis due to its suitability for both the stable and convective PBL. The Richardson number is defined as

$$Ri(z) = \frac{(g/\theta_{vs})(\theta_{vz} - \theta_{vs})(z - z_s)}{(u_z - u_s)^2 + (v_z - v_s)^2 + (bu_*^2)}, \qquad (3)$$

where s denotes the surface, z is the height, g is the acceleration of gravity, θ_v is the virtual potential temperature, u and v are the wind speed components, b is a constant, and u_* is the surface friction velocity. Here we set b = 0 to ignore the surface frictional effect, and assume the surface values of u and v to be zero. The lowest level z at which the interpolated *Ri* crosses the critical value of 0.25, $z(Ri_{0.25})$, is the estimated PBL height for our analysis. Unfortunately, this method can only be used for the diurnal mean data of the CFMIP2, which represents primarily the daytime situation, because most of the necessary quantities on global scale are not available in the hourly CFMIP2 archive. Thus, the nighttime PBL depth remains underived. However, the 3-h CFMIP2 and ARMBE data over the ARM-SGP site can be used to estimate the entire diurnal cycle of the PBL depth using the VH method.

2.4 Methodology

Our analysis consists of two major components: (1) the multi-model ensemble mean biases and (2) the inter-model variations. For the first component, the biases of the simulated climatology of the temperatures (DTR, T_{max} and T_{min}) and PBL mixing from the CMIP5 multi-model ensemble means are examined to identify their connections. The modeled temperature time series are first assessed through an inter-annual correlative analysis at the grid level with their key controlling factors to assess the model's capacity in describing the well-known statistical relationships. Linear trends are removed from the time series before the correlations are calculated. The correlative coefficients are presented as the area-weighted averages over four large-scale climate zones defined later. Then the geospatial patterns of the modeled temperature climatology are compared with the CRU observations to quantify their biases and several simulated metrics representing the PBL mixing are validated against ERA-Interim. As the primary features of model biases with respect to PBL characteristics remain consistent across seasons, only annual mean results are presented in this paper. Note that not all models provide the variables that are studied, but of the total 20 models, at least 16 are included in any of the variable analysis. In addition, the PBL mixing estimated from the CFMIP2 data for the daytime situation and the whole diurnal cycle over the ARM-SGP site are presented to support our results from the CMIP5 models.

For the second component, the inter-model variations of the simulated DTR, T_{max} and T_{min} climatology are analyzed on the principle of the PBL mixing patterns. First, the regions with large inter-model variations of the simulated temperatures are sorted out based on the PBL_{max}, PBL_{night} and the diurnal PBL-depth range (DPR, DPR = $PBL_{max} - PBL_{night}$) climatology. Here the

PBL depth in the ERA-Interim archive is used because it includes the effects of turbulent parameters, and hence simulates the spatial variation much better than the data derived from the VH method (Davy and Esau 2014b). Then we attempt to plausibly explain the large inter-model variations of the simulated temperatures in terms of two dominating processes related to surface radiation and PBL mixing that affect land–atmosphere interactions, aiming at quantifying the contributions of the PBL mixing to model uncertainties in simulating the diurnal cycle of LSAT. With several dominant variables selected, an inter-model correlative analysis is used to determine which variables significantly impact the model differences of the simulated temperatures over certain climate regions, and a variance analysis method derived from the multiple-linear regression is implemented on these variables to quantify their relative contributions.

For the regional mean statistics of correlative analyses at grid levels, four typical climate zones are defined following Bonan et al. (2002): boreal arctic regions [Alaska and northwestern Canada (55°N–75°N, 170°W–90°W) + Siberia (55°N–75°N, 80°E–140°E)], tropical rainforests [Amazon (15°S–5°N, 75°W–45°W) + Congo (10°S–5°N, 10°E–25°E)], transitional zones between wet and dry climates [central US (30°N–50°N, 110°W–85°W) + central Asia (40°N–55°N, 70°E–110°E) + India (5°N–30°N, 70°E–90°E)], and arid/semi-arid regions [Sahara and Arabian Peninsula (10°N–30°N, 0°E–50°E) + southern Africa (35°S–15°S, 10°E–30°E)].

3 Results and discussion

3.1 Multi-model ensemble means

3.1.1 Inter-annual correlations between temperatures and key controlling variables

The CMIP5-modeled DTR, T_{max} and T_{min} time series are first evaluated by examining their basic statistical relationships with several key controlling factors over the four climate zones. Table 2 lists the areal averages of the grid-scale inter-annual correlations between the temperatures and 8 diagnostic variables. These eight variables are chosen because they are significantly ($p < 0.05$) correlated with at least one of the three temperatures. As expected, total cloud cover (TCC) negatively correlates with T_{max} (and thus DTR) over all climate zones while it has a negligible impact on T_{min}. The relatively weaker relationship between TCC and T_{max} over the boreal arctic regions could be attributed to less incident solar insolation than other climate zones. Total atmospheric water vapor content (WVC) reduces DTR by significantly increasing T_{min} due to water

Table 2 Inter-annual correlations between the temperatures (DTR, T_{max} and T_{min}) and eight state variables averaged by climate zone

Climate zone	Boreal arctic regions (Alaska + northwestern Canada + Siberia)			Tropical rainforests (Amazon + Congo)			Transitional zones (central US + central Asia + India)			Arid and semi-arid regions (Sahara + Arabian Peninsula + southern Africa)		
	DTR	T_{max}	T_{min}	DTR	T_{max}	T_{min}	DTR	T_{max}	T_{min}	DTR	T_{max}	T_{min}
TCC	**−0.39**	**−0.29**	0.11	**−0.8**	**−0.75**	0.05	**−0.74**	**−0.7**	0.1	**−0.61**	**−0.5**	0.15
WVC	−0.05	**0.77**	**0.76**	**−0.46**	0.19	**0.69**	**−0.57**	0.12	**0.49**	**−0.63**	0.04	**0.55**
EF	−0.08	−0.05	0.03	**−0.9**	**−0.69**	**−0.26**	**−0.57**	**−0.48**	−0.12	**−0.55**	**−0.39**	0.19
SM	−0.12	−0.07	0.03	**−0.8**	**−0.72**	−0.1	**−0.76**	**−0.56**	−0.19	**−0.5**	**−0.45**	0.03
VAS[a]	−0.15	**0.34**	**0.36**	−0.13	−0.16	−0.03	0.01	**0.32**	**0.33**	−0.11	**0.4**	**0.43**
SH_{max}	**0.34**	**0.37**	0.02	**0.8**	**0.72**	0.22	**0.7**	**0.58**	0.24	**0.66**	**0.55**	0.05
LCL	**0.3**	**0.39**	0.12	**0.74**	**0.79**	**0.4**	**0.55**	**0.7**	**0.4**	**0.48**	**0.46**	0.03
SH_{night}	**0.34**	0.02	**−0.33**	−0.19	−0.08	0.11	**0.38**	−0.10	**−0.39**	**0.35**	−0.10	**−0.38**

Values are areal mean correlation coefficients averaged by climate zone and those in boldface are considered statistically significant if at least half of the area of the climate zones have correlations passing the significance test at the 5 % level

The eight state variables include total cloud cover (TCC), vertically integrated atmospheric water vapor content (WVC), evaporative fraction (EF), uppermost 10-cm soil moisture (SM), 10-m meridional wind (VAS), diurnal maximum of sensible heat (SH_{max}), lifting condensation level (LCL) and nighttime mean of sensible heat (SH_{night}). All of these variables are the multi-model ensemble means of CMIP5 models

[a] Given the inverse direction of heat advective effects of VAS in the two hemispheres, the correlative coefficients are first calculated separately in the Northern (between +VAS and temperatures) and Southern (between −VAS and temperatures) hemisphere and then averaged by climate zone

vapor enhanced downward longwave radiation (Dai et al. 1999). Thus, the negative correlation between WVC and DTR is evident over all regions except the boreal arctic regions where WVC warms T_{max} and T_{min} comparably. EF measures the proportion of the latent heat in total turbulent fluxes which are partitioned from the available net surface radiation, and thus negatively correlates with T_{max} (and thus DTR) over most land areas. The similar negative correlation is also found between the uppermost 10-cm SM and T_{max} (and DTR), with the strongest correlations over the transitional areas and tropical rainforests, which are consistent with the findings of hot spots with strong SM-atmosphere coupling by Koster et al. (2004). Despite little correlation with DTR, the 10-m poleward meridional wind (VAS) has a comparable positive correlation with T_{max} and T_{min} over all the regions except the tropical rainforest where the meridional temperature gradients are weak. This corresponds to the fact that LSAT is sensitive to the horizontal heat advection in mid- and high latitudes (Dai et al. 1999; Wang and Zeng 2014). As for the PBL mixing effects, SH_{max} and LCL both show a positive correlation with DTR as T_{max} tends to be warmer with stronger convective mixing at daytime. During the nighttime, the downgradient diffusion warms T_{min} by increasing downward (negative) heat flux under stable-stratified conditions, and thus SH_{night} positively correlates with DTR. The relatively weak relationship between SH_{night} and T_{min}/DTR, particularly over the tropical forests, could be partly attributed to the non-linear effects of near-surface temperature gradient on the turbulence and sensible heat flux (Holtslag et al. 2007). In such cases, although the sensible heat flux

is proportional to the vertical temperature gradient, when the surface is extreme cold or the wind speed is too low, the stability effects could become so strong that the turbulence would die out and the heat flux would vanish. Nevertheless, the general positive correlations between SH_{night} and DTR still allow us to use SH_{night} to estimate the nighttime effects of the downgradient turbulent diffusion. Overall, these correlative relationships are all consistent with the basic physical principles and previous results, indicating that the CMIP5 models are able to describe general variations of the temperature diurnal cycle and their associations with key controlling factors.

3.1.2 Model biases in simulating the diurnal temperature climatology

Figure 1 shows the spatial patterns of the DTR, T_{max} and T_{min} climatology in the multi-model ensemble means and the corresponding CRU observations. Evidently, CMIP5 models are able to reproduce the general spatial characteristics of the observed temperatures, with a spatial correlation >0.9 ($p < 0.01$, n = 14,466) for all three variables. The regions with large diurnal variations of LSAT are well captured, such as the southwestern US, the Sahara desert, Central and Western Asia, southern Africa and most parts of Australia. However, quantitatively DTR, T_{max} and T_{min} exhibit biases similar to previous studies with fully-coupled models (e.g. IPCC 2007). The magnitude of DTR is underestimated over almost all land areas, with mostly a cold bias of T_{max} except central North America, the Amazon basin, eastern Siberia and parts of western and central Asia,
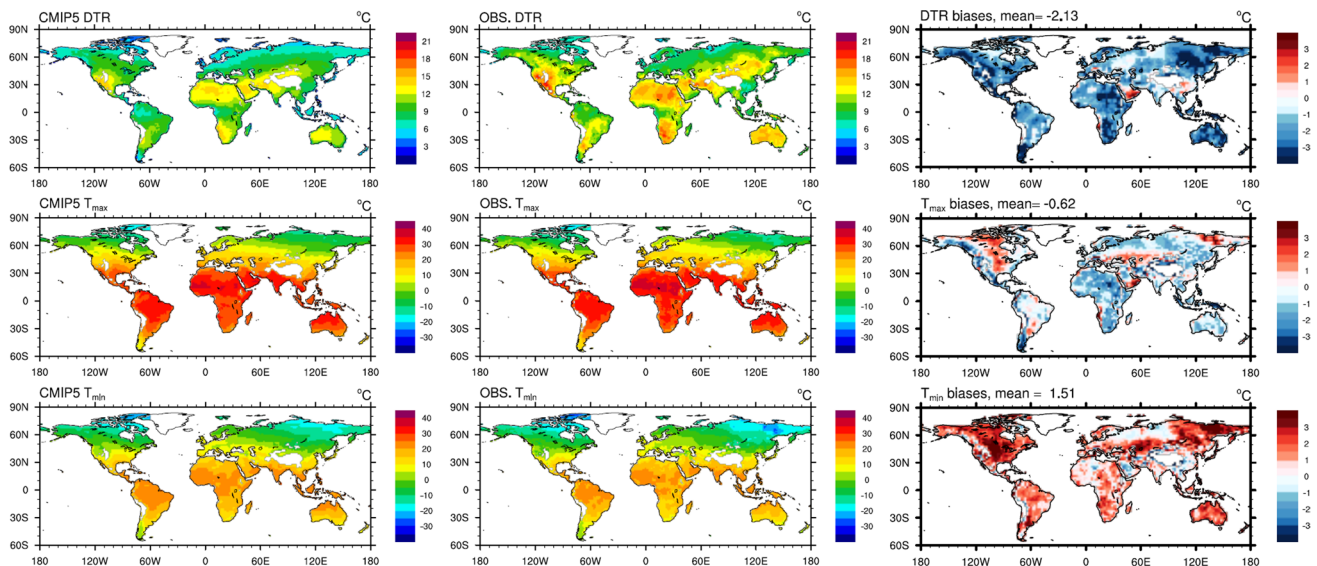
Fig. 1 Geographical distribution of the CMIP5-simulated and CRU-observed climatological DTR, T_{max} and T_{min} (°C) and the multi-model ensemble mean biases

Table 3 Global mean biases of DTR, T_{max} and T_{min} (°C) and global mean RMSE of DTR (°C) in multi-model ensemble mean and individual models along with the systematic RMSE of DTR ($RMSE_S$, °C) in parenthesis

	DTR bias	T_{max} bias	T_{min} bias	DTR RMSE ($RMSE_S$)
Ensemble mean	−2.13	−0.62	1.51	2.38 (1.48)
GFDL-HIRAM-C360	−1.45	−1.28	0.18	2.49 (1.86)
CMCC-CM	−2.17	−0.78	1.38	2.76 (1.7)
EC-EARTH	−2.24	−1.39	0.85	2.85 (1.52)
CNRM-CM5	0.52	0.36	−0.16	2.86 (1.71)
HadGEM2-A	−1.51	−1.13	0.38	2.89 (1.66)
MRI-AGCM3-2S	−2.17	−0.79	1.38	2.89 (1.77)
CSIRO-Mk3-6-0	−1.93	0.36	2.29	2.95 (1.72)
MRI-CGCM3	−2.29	−0.39	1.9	2.99 (1.81)
CCSM4	−2.1	−1.41	0.68	3.06 (2.11)
FGOALS-s2	−2.33	1.14	3.47	3.07 (1.87)
ACCESS1-0	−2.01	−1.14	0.86	3.14 (1.75)
CESM1-CAM5	−2.07	−0.99	1.08	3.15 (2.21)
MIROC5	−2.18	0.85	3.02	3.29 (1.78)
CanAM4	0.32	0.38	0.06	3.59 (1.78)
MPI-ESM-LR	−3.11	−0.54	2.56	3.63 (1.83)
GFDL-CM3	−3.05	−1.35	1.71	3.73 (1.98)
ACCESS1-3	−3.27	−1.3	1.97	3.79 (1.8)
BNU-ESM	−3.13	−1.77	1.36	3.84 (2.01)
GISS-E2-R	−3.32	−0.69	2.63	3.87 (1.89)
INM-CM4	1.79	−1.77	−3.55	4.58 (2.01)

and a widespread warm bias of T_{min} except northwestern Eurasia and parts of western and southern Asia (Fig. 1).

Table 3 lists the area-weighted global mean temperature biases for the multi-model ensemble means and individual models. The multi-model ensemble mean biases for DTR, T_{max} and T_{min} are −2.13, −0.62 and 1.51 °C respectively, which are consistent with the overall spatial features with respect to the characteristic sign of the biases (Fig. 1). Similarly, all individual models except CNRM-CM5, CanAM4 and INM-CM4 also systematically underestimate DTR due to underestimated T_{max} and overestimated T_{min}. CNRM-CM5 and CanAM4 slightly overestimate DTR because of overestimated T_{max} (and slightly underestimated T_{min} for CNRM-CM5), while INM-CM4 notably overestimates DTR because of severely underestimated T_{min}. To further quantify the fidelity in simulating the spatial pattern of the DTR climatology for each model, we also calculate the root mean square error (RMSE) of DTR and its systematic portion ($RMSE_S$) following Willmott (1984) using the grid-scale annual means of DTR as the samples of the modelled and observed values. All models have comparable $RMSE_S$, suggesting that they perform comparably well in reproducing the spatial gradient of the observed DTR, but the RMSE

shows distinct model variations, with the best results in GFDL-HIRAM-C360 and CMCC-CM (<2.8 °C) and the worst in INM-CM4 (4.58 °C). Hence, the inter-model differences in RMSE result largely from the unsystematic portion, which represents the model's fidelity in simulating localized observations. The multi-model ensemble mean exhibits the smallest RMSE value (2.38 °C), indicating its best accuracy in simulating DTR than any individual models. Nevertheless, both the multi-model ensemble means and individual models systematically underestimate the DTR, with possible causes investigated next mainly from the perspective of the PBL mixing.

3.1.3 Model biases in simulating PBL mixing

We first examine the model biases of the PBL mixing proxies derived from the CMIP5 output. Figure 2 shows the climatological SH_{max}, LCL and PTC from ERA-Interim and the corresponding biases in the multi-model ensemble means. SH_{max} and LCL represent the general characteristics of a convective daytime PBL. Their spatial patterns show an evident gradient from low values over cold and moist regions to high values over hot and dry regions, and their biases have a good spatial correspondence to the biases in T_{max}. A negative bias, which represents the weakened level of turbulent mixing, is evident over northern Eurasia, eastern Asia and the arid/semi-arid regions in mid- and low-latitudes such as western US, the Sahara desert, southern Africa and central Australia, while a positive bias, which represents the enhanced level of PBL mixing, mainly appears over the central US, parts of western and central Asia, India and tropical forest as also identified by Svensson and Lindvall (2015). For the regions with the weakened PBL mixing, a larger simulated humidity gradient at the entrainment zones with more dry air entrained into the PBL is expected, with a more persistent humidity gradient at the bottom of the PBL and a higher latent heat. The opposite situation is expected for the regions with the enhanced PBL mixing. The strong inverse spatial correspondence in the biases of PTC and SH_{max} and LCL is a good indication of such anticipation, showing the overestimated (underestimated) evaporation efficiency over the regions where SH_{max} and LCL are underestimated (overestimated). Since PTC is a good indicator of the enhanced surface evaporation due to the effect of dry-air entrainment (Van Heerwaarden et al. 2009; Dirmeyer et al. 2013), the overestimated evaporation efficiency and thus the underestimated T_{max} could be partly attributed to the enhanced dry-air entrainment induced by the weakened PBL turbulent mixing, and vice versa.

The modeled level of the nighttime PBL mixing is given in Fig. 3. The general negative SH_{night} indicates that the stably-stratified conditions dominate over almost all land
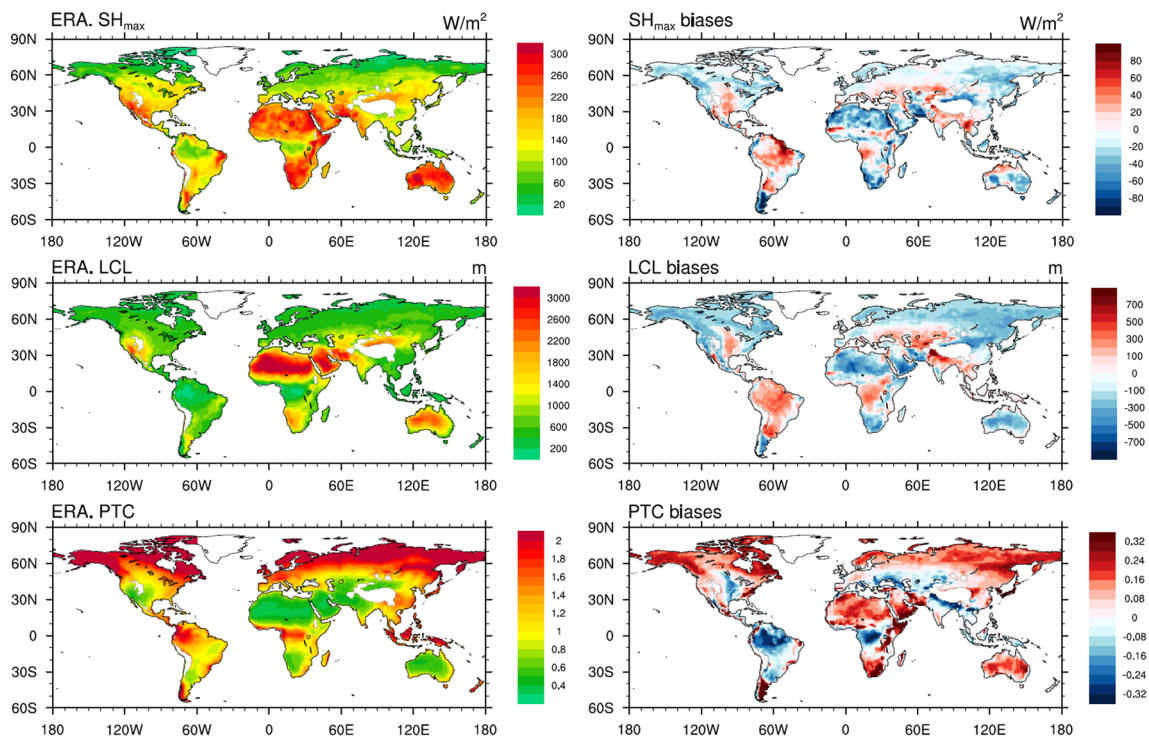
Fig. 2 Geographical distribution of the climatological SH_{max} (W/m^2), LCL (m) and PTC (unitless) in the ERA-Interim reanalysis (*left panels*) and the corresponding CMIP5 multi-model ensemble mean biases (*right panels*)
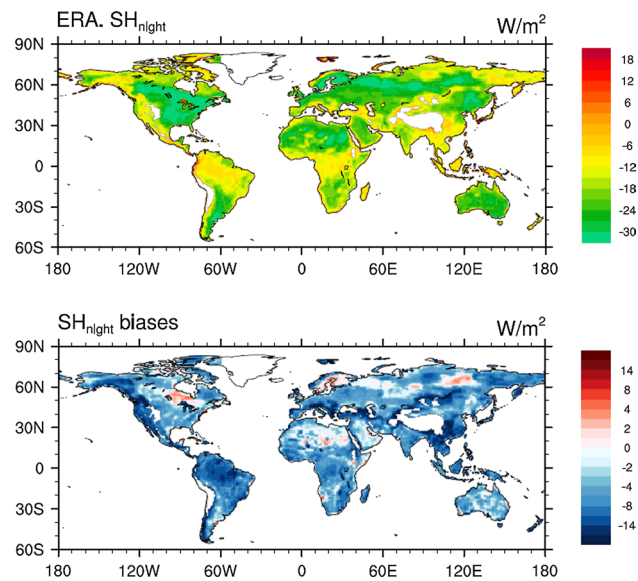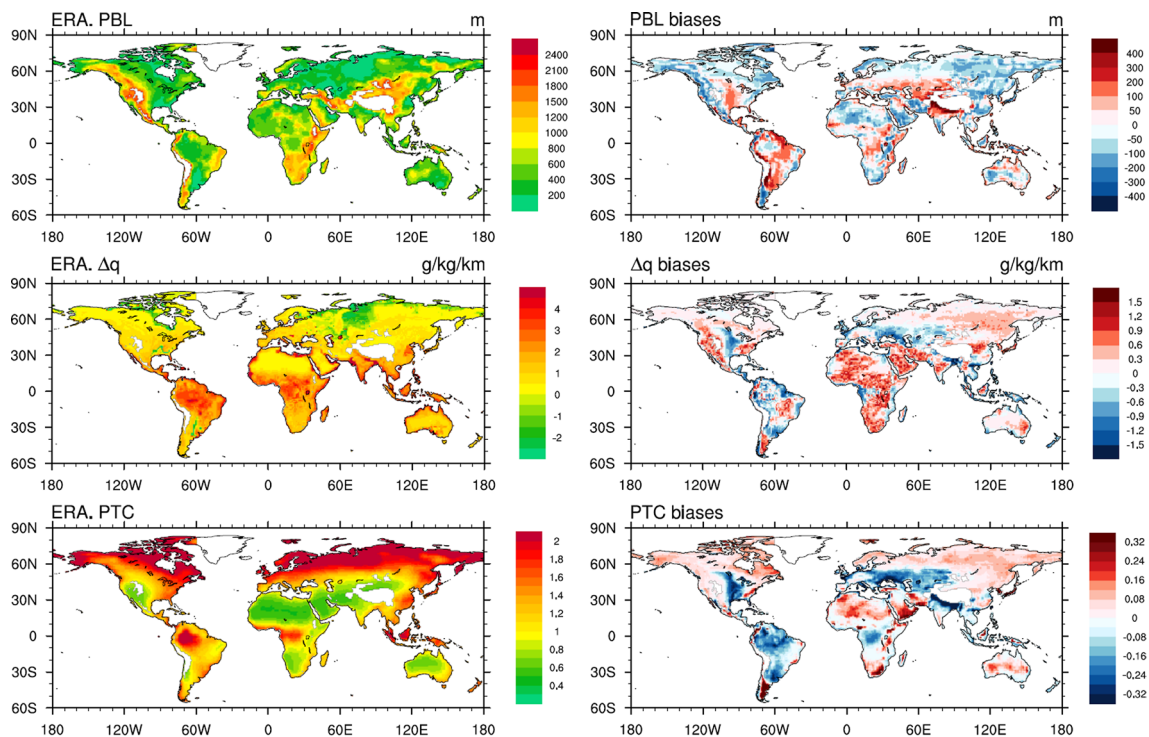

Fig. 3 Geographical distribution of the climatological SH_{night} (W/m^2) in the ERA-Interim reanalysis (*top panel*) and the corresponding CMIP5 multi-model ensemble mean biases (*bottom panel*)

areas. Hence, the larger downward sensible heat fluxes (i.e., negative biases in SH_{night}) suggest that the downgradient diffusion in the PBL is generally overestimated, and thus

transports more heat from the atmosphere to the near-surface, which creates the widespread warm bias of T_{min}. This enhanced mixing in a stably-stratified PBL has been widely recognized in previous studies on weather and climate models (e.g. Steeneveld et al. 2008; Shin and Hong 2011; Svensson et al. 2011; Holtslag et al. 2013), but addressing this long-standing issue has proven challenging because it effectively compensates for errors caused by other processes such as large-scale flow (Sandu et al. 2013).

To further confirm the above CMIP5 results, we next analyze the diurnal mean PBL mixing using the high vertical resolution CFMIP2 data which allow us to directly estimate the PBL depth with the VH method. Given that the multi-model ensemble mean temperature biases in CFMIP2 exhibit similar patterns as those in CMIP5 (Figure SI3), it is not surprising that the biases of the estimated PBL height against the ERA-Interim results (which are similarly derived using the VH method) are comparable to the biases of SH_{max} and LCL estimated from the CMIP5 output, with the overestimates over the warm T_{max} bias regions and the opposite over the cold T_{max} bias regions (Fig. 4). Here we use the jump of specific humidity at the PBL height (Δq), defined as the PBL value minus the free-atmosphere value, to quantify the potential effects of the dry air entrainment as done in Van Heerwaarden et al. (2009). The lower the PBL height is, the more the jump of the specific humidity

Fig. 4 Geographical distribution of the climatological PBL height (m), Δq (g/kg/km) and PTC (unitless) estimated from the ERA-Interim reanalysis (*left panels*) and the corresponding CFMIP2 multi-model ensemble mean biases (*right panels*)

and thus the dry air entrainment could be made. Figure 4 shows that Δq is generally overestimated over the regions with the weakened PBL depth such as central and eastern Asia, the Sahara desert and southern Africa, and basically underestimated in the regions with the enhanced PBL depth such as the central US and parts of western and central Asia. The biases of Δq correspond spatially well to these of PTC (Fig. 4), indicating that more dry air entrainment leads to higher surface evaporation and thus potentially lower T_{max}, and vice versa. These results agree with the patterns of the PBL mixing proxies in the CMIP5 analysis and consistently support our hypothesized cause for the daytime temperature biases. Due to lack of required hourly data in CFMIP2 archive, the biases of PBL depth at nighttime cannot be further validated at the global scale.

Finally we analyze the three-hourly CFMIP2 data together with the ARMBE observations to evaluate the entire diurnal cycle of the PBL mixing and its connection to temperature evolution at the ARM-SGP site. Figure 5 shows the climatological diurnal cycle of the LSAT, PBL height and sensible heat flux between the modeled and observed. The modeled T_{max} and T_{min} are both overestimated at the ARM-SGP site, which is consistent with the grid level results shown in Fig. 1. The underestimated DTR is due to the larger overestimation of T_{min} than T_{max}. Correspondingly, both the PBL_{max} and PBL_{night} are

overestimated, which is consistent with the results of Seidel et al. (2012) who used radiosonde observations to evaluate the simulated PBL climatology over the US. This overestimation could potentially warm T_{max} and T_{min} by less dry air entrainment at daytime and more downgradient turbulence diffusion at nighttime, respectively. The sensible heat shows a consistent bias pattern, with the overestimated SH_{max} representing a stronger PBL mixing and the underestimated SH_{night} (a larger downward absolute value) representing more downward heat transfer. This supports our rationality to use the sensible heat as a proxy for PBL depth in the CMIP5 analysis. In addition, a comparison of the vertical profile of specific humidity indicates that the overestimated PBL height at daytime generates a lower Δq, which represents a less amount of potential dry air entrainment into the PBL (Fig. 6). This further cements our hypothesis about the interactions among the PBL mixing, dry air entrainment and surface evaporation.
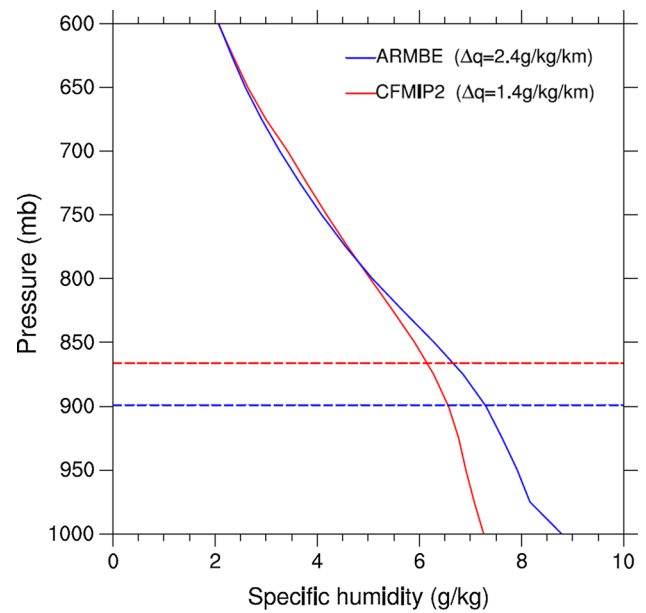
3.1.4 Temperature biases related to other contributing factors

We realize that the biases of the PBL mixing cannot fully explain the temperature biases and other factors may also have a contribution, especially at regional scales. For example, during the daytime, the overestimated solar radiation
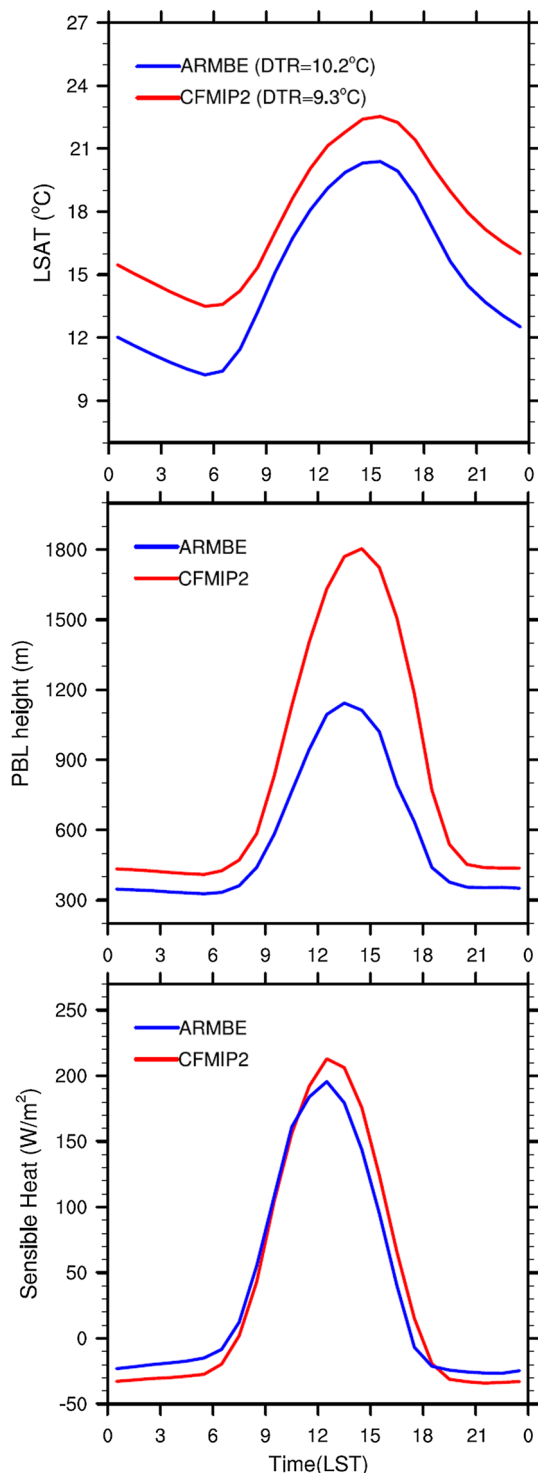
Fig. 5 The diurnal cycle of the LSAT (°C), PBL height (m) and sensible heat flux (W/m^2) as a function of local solar time (LST) from the CFMIP2 multi-model ensemble means and ARMBE data over the ARM-SGP site. The value of DTR (°C) is shown in the *top panel*

Fig. 6 Vertical profiles of specific humidity (g/kg) from the CFMIP2 multi-model ensemble means and ARMBE observations over the ARM-SGP site. The *horizontal dashed lines* indicate the PBL heights estimated from the two datasets. The jump of specific humidity at the PBL height, Δq (g/kg/km), defined as the PBL value minus the free-atmosphere value, is also shown

along with the underestimated TCC should play a role in explaining the warm bias of T_{max} over central North America and the Amazon basin (Figure SI4). The underestimated precipitation over these regions (Figure SI4) as also identified by Scheff and Frierson (2015) could have an effect by lowering the surface evaporation as well. However, most parts of central North America and the Amazon basin are dominated by wet climate conditions where the SM is plentiful. In this case, the surface evaporation and T_{max} depend much less on SM variations, but more on the near-surface humidity gradient which is effectively controlled by the dry air entrainment as indicated by Van Heerwaarden et al. (2009). Hence, the PBL mixing effects associated with the dry air entrainment could make a more important contribution to the T_{max} bias. In contrast, over the Sahara desert and southern Africa where dry climate conditions dominate, the cold bias in T_{max} (Fig. 1) could be largely attributed to the enhanced evaporation due to overestimated precipitation (Figure SI4). As these regions are extremely limited by the available SM, the PBL mixing effects on the humidity gradient may not be dominant. Nevertheless, the sensitivity test made by Van Heerwaarden et al. (2009) showed that the dry air entrainment could significantly enhance the surface evaporation under all conditions, even over dry land surfaces although it only plays a secondary role. For this reason, we propose this mechanism as a global scale contributor to the daytime T_{max} biases, including dry climate conditions.

During the nighttime, the overestimated PBL mixing always plays a major role in causing the warm biases of

T_{min}. Moreover, the surface downward longwave radiation, the main radiative driver for the temperature variation at night, is found to be generally underestimated (Figure SI4) as also identified by Steeneveld et al. (2008), which tends to induce a cold bias of T_{min}. Given the widespread warm biases of T_{min} (Fig. 1), the underestimated radiative forcing further highlights the contributions of the enhanced mixing to the T_{min} biases.

In summary, the biases of PBL mixing are consistent with our expectations about their roles in inducing the biases of the diurnal cycle of LSAT. It is reasonable to use the interaction between the dry air entrainment and surface evaporation at daytime and the enhanced mixing at nighttime as the possible explanation for the biases of T_{max} and T_{min}. Hence the systematic deficiencies of PBL schemes in simulating the turbulent mixing could very likely contribute to the biases of DTR although other factors also have an effect. The analyses on the multi-model ensemble mean biases mainly reflect the systematic errors in GCMs. However, in order to make significant improvement in GCMs, the regions with large inter-model disagreements of the simulated temperatures should be identified and highlighted because they are probably the locations where the representations of temperature-related physical processes have large uncertainties. Therefore, we will focus on the inter-model variability of the diurnal cycle of LSAT in the next section.

3.2 Inter-model variability

3.2.1 Inter-model RMSE in simulating DTR, T_{max} and T_{min} climatology

The multi-model ensemble means have shown above to some extent a spatially homogeneous sign of biases in the climatology of DTR, but may conceal some key information with respect to regions where the results among models vary substantially. For instance, assume that two models have an identical but opposite sign of the bias and their ensemble mean is zero, but in reality neither of the models generates the correct result. One way to avoid this problem is to examine the inter-model RMSE as the measurement of the model deviations from observations. Note that RMSE ($\sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - obs)^2}$) consists of two portions, namely ($\frac{1}{n} \sum_{i=1}^{n} (x_i - obs)^2) = (\bar{x} - obs)^2 + \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$, where n represents the total number of the models, x_i is the individual model result, and $\bar{x}$ is the multi-model ensemble mean. The first term represents the RMSE induced by the multi-model ensemble mean biases and the second term denotes the inter-model standard deviation (SD).

Despite a widespread negative DTR bias, the inter-model RMSE of DTR (Fig. 7) shows a strong locality in the
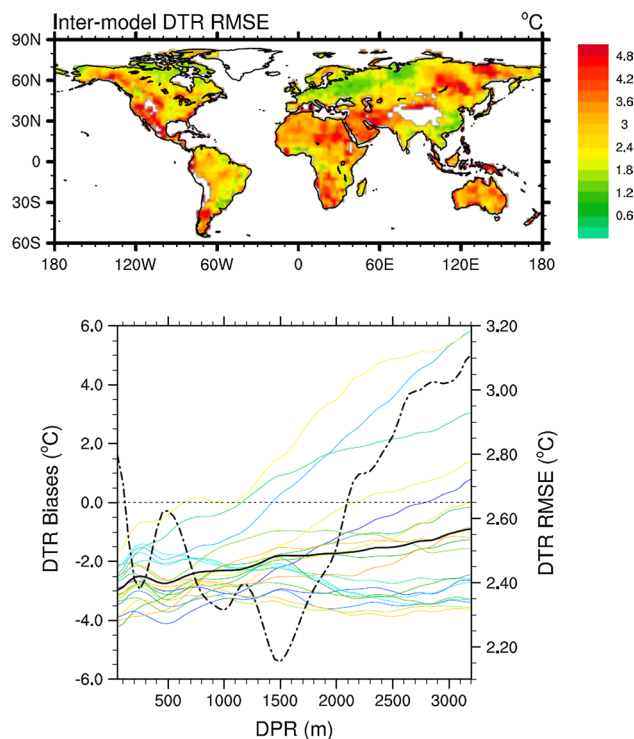
Fig. 7 Geographical distribution of inter-model RMSE of DTR (°C) climatology (*top panel*) and biases and RMSE of DTR climatology as a function of diurnal PBL-depth range (DPR, in m) provided by the ERA-Interim reanalysis (*bottom panel*). In the *bottom panel*, the *colored lines* represent 20 individual CMIP5-AMIP simulations, the *black solid line* is the multi-model ensemble mean, and the *black dashed line* is the inter-model RMSE of DTR climatology

spatial pattern. The large RMSE values mainly appear over the semi-arid and arid regions such as the Sahara desert and southern Africa and over the boreal cold and dry regions such as northeastern Russia and China. In high latitudes, the RMSE may result from the multi-model ensemble mean biases as they are spatially coupled. However, in mid- and low- latitudes, it seems that the inter-model SD plays an important role in determining RMSE, especially over the semi-arid and arid regions where the multi-model ensemble means strongly depend on the model selection. Here we take the Sahara desert as an example. The multi-model ensemble means of T_{max}, T_{min} and DTR biases derived from the top 5 models in Tables 3 are −1.22, 0.45 and −1.67 °C, respectively, whereas those from the bottom 5 models are −1.92, 0.78 and −2.7 °C, respectively. The notable differences between the two groups of models suggest that the ensemble mean biases could be largely affected by the choice of models, especially over the regions with the large inter-model variations.

We expect to see the large inter-model RMSE of the temperatures over the regions where CMIP5 models have substantial uncertainties in modelling the PBL mixing.

The DPR quantifies the amplitude of the PBL depth transition from convective daytime conditions to stably-stratified nighttime conditions. The regions with a large DPR always experience extreme PBL conditions which are poorly modeled by current PBL schemes (e.g. Holtslag et al. 2013; Sterk et al. 2013), and hence are expected to have large uncertainties in modeled DTR. Figure 7 shows the global mean DTR biases for each individual model as a function of DPR provided by the ERA-Interim reanalysis. All models demonstrate a negative DTR bias except several models with notably overestimated DTR. The inter-model spread is relatively small in low-DPR regions, but tends to increase gradually with the magnitude of DPR and maximizes with the largest DPR. This is consistent with the spatial pattern of the inter-model RMSE of DTR, suggesting to a large extent that the modeled PBL mixing induces large uncertainties in modeled DTR over the largest DPR regions such as the arid and semi-arid regions. In addition, the inter-model RMSE of DTR also has large values over the smallest DPR regions such as the boreal arctic regions, which could be caused by model's systematic biases as well.

The ensemble mean biases and RMSE are quantified based on the observations and thus also depend on the data quality of observations, which may vary in time and space. Due to lack of knowledge about the uncertainties of observations, next we will make a more specific quantitative analysis of the modeled DTR uncertainties by concentrating on the inter-model variations of the simulated temperatures rather than the model biases and RMSE which are affected by the uncertainties of observations. We will identify the regions with the largest inter-model variations of DTR as well as T_{max} and T_{min} based on the climatological PBL patterns, and then quantify the relative contributions to the inter-model temperatures variations from the modeled PBL turbulent mixing and other variables over these regions.

3.2.2 Spatial characteristics of inter-model variations of the simulated temperatures

Figure 8 shows the geospatial patterns of inter-model SD of the simulated DTR, T_{max} and T_{min} climatology. Compared to the inter-model RMSE (Fig. 7), the inter-model
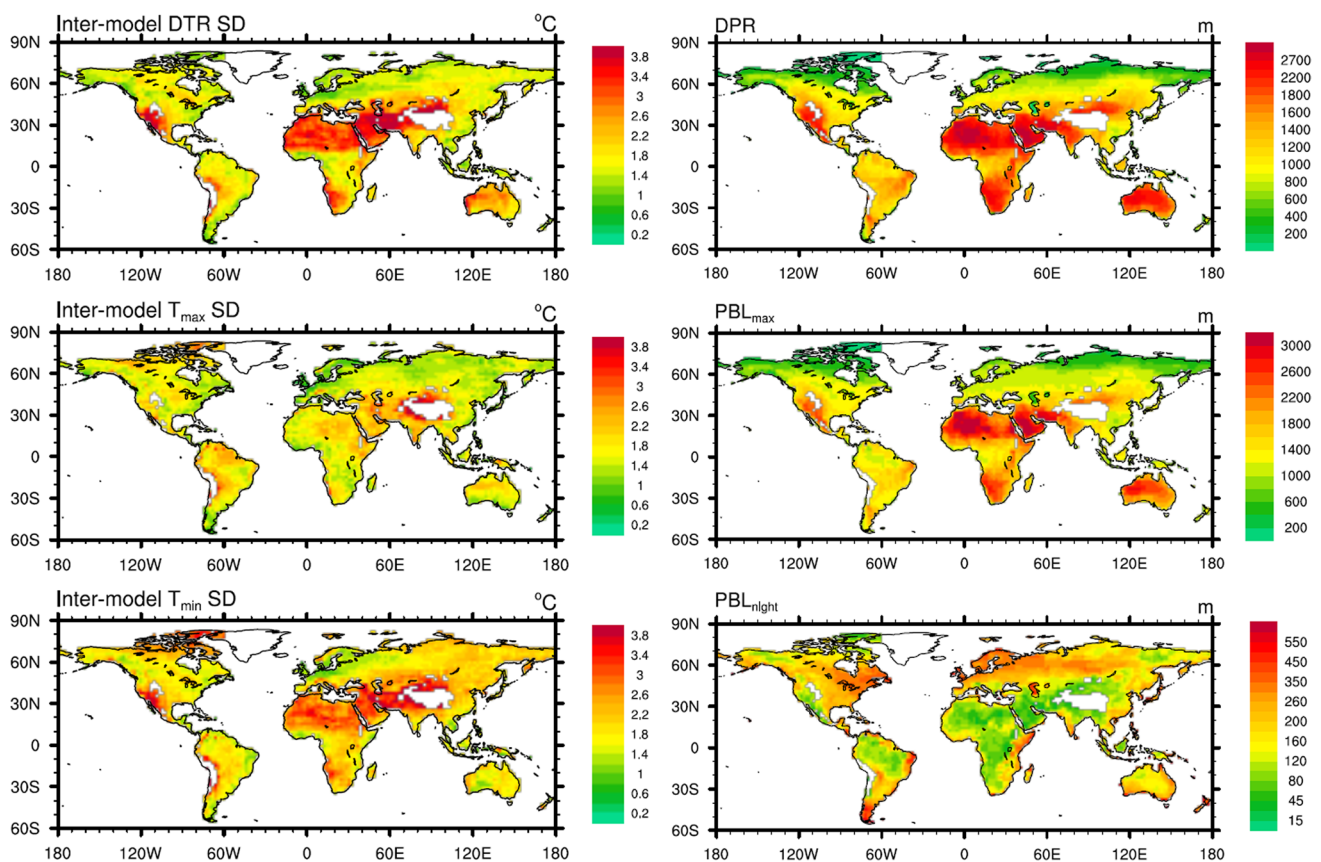
Fig. 8 Geographical distribution of inter-model standard deviation of T_{max}, T_{min} and DTR (°C) climatology (*left panels*) and the corresponding climatological diurnal maximum, nighttime mean and range of PBL depth (PBL$_{max}$, PBL$_{night}$ and DPR, in meters) provided by the ERA-Interim reanalysis (*right panels*)

SD of DTR shows a similar but much clearer locality pattern, with large values over the semi-arid and arid regions such as southwestern North America, the Sahara desert and Arabian Peninsula, southern Africa and Australia. The inter-model SD of T_{max} and T_{min} presents a consistent distribution with that of DTR, with more uncertainties in T_{min} because of the poor representation of stably-stratified conditions in PBL schemes (e.g. Steeneveld et al. 2008; Holtslag et al. 2013). Relatively large values of inter-model SD of T_{max} and T_{min} also appear over the boreal arctic regions such as northern Canada, which is more notable in winter (not shown). However, due to the similar magnitudes of T_{max} and T_{min} over these regions and thus fairly small DTR in each model, the inter-model SD of DTR is not as much as that of T_{max} and T_{min}.

The aforementioned spatial features of the inter-model SD of the temperatures are expected as they match very well with the climatological PBL patterns (PBL_{max}, PBL_{night} and DPR) provided by the ERA-Interim (Fig. 8). The arid and semi-arid regions experience both the strongest convective conditions with the largest PBL_{max} during the daytime, and the stable-stratified conditions with the smallest PBL_{night} during the nighttime. Hence, it is natural to see that the large inter-model SD of DTR appears over these regions with the greatest day-night contrast in the PBL depth (i.e., the largest DPR), and so do the large inter-model variations of T_{max} and T_{min}. The boreal arctic regions are dominated by long-lived stable conditions with a very shallow PBL during both daytime and nighttime compared to ambient conditions, and thus become another climate zone with the large inter-model SD of T_{max} and T_{min}. However, the long-lived stable conditions (with the smallest DPR) make the model differences of DTR insignificant. To further confirm these results, we plot the inter-model SD of the simulated DTR, T_{max} and T_{min} as a function of DPR in Fig. 9, which shows that the largest inter-model variations appear over the regions with the largest and smallest DPR, namely the arid and semi-arid regions and boreal arctic regions. The former controlled by alternate strong convective and stable-stratified conditions are identified as the primary regions with the largest inter-model variations of DTR, T_{max} and T_{min}, and the latter are also marked with the large inter-model variations in T_{max} and T_{min} but not in DTR.

3.2.3 Relative contributions of PBL mixing to inter-model variations

As the large inter-model variations of the simulated temperature diurnal cycle mainly appear over the regions which are always dominated by extreme PBL conditions, the uncertainties in the simulated PBL mixing are expected to play an important role. However, representations of other
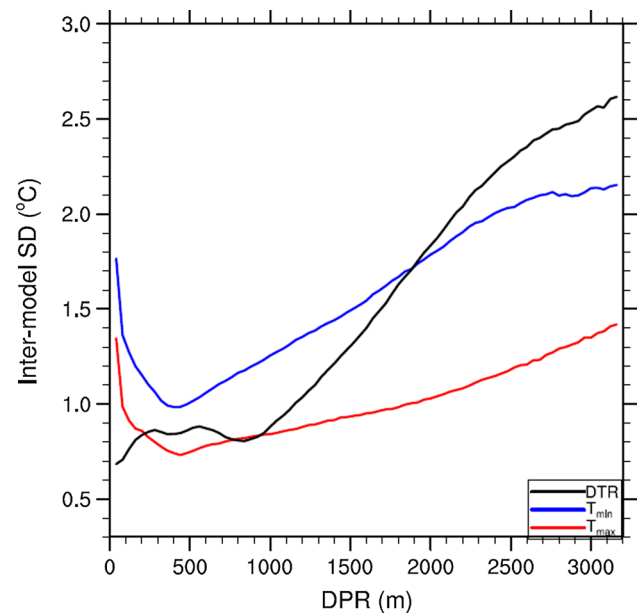
Fig. 9 Inter-model standard deviation of climatological DTR, T_{max} and T_{min} (°C) as a function of diurnal PBL-depth range (DPR, in m) provided by the ERA-Interim reanalysis

processes such as the surface radiation also affect the temperature variations. In this subsection, we will quantify the relative contributions to the large inter-model temperature variations from several key controlling variables related to surface radiative and PBL mixing processes. The key radiative variables considered here include surface downward shortwave and longwave radiation in clear sky conditions (DSRCS and DLRCS) and the corresponding cloud radiative effect (DSRCF and DLRCF) as well as the land surface albedo and emissivity which determine the amount of net radiation. The key variables related to the PBL mixing contain EF and top 10-cm SM which determine the partitioning of turbulent fluxes in terms of the evaporation and total available water, and another three variables (SH_{max}, LCL and SH_{night}) introduced in Sect. 2.3. Note that PTC is not chosen as it is highly correlated with EF and reciprocal SH_{max} or LCL. The relative contributions of these controlling variables to T_{max}, T_{min} and DTR are estimated based on a statistical method.

We first calculate the areal averages of the grid-scale inter-model correlations between the temperatures and selected controlling variables to sort out the significant influential factors by climate zone (Table 4). For the surface radiative processes, the shortwave solar radiation is largely modified by cloud cover during the daytime and thus DSRCF negatively correlates with T_{max} and DTR over all regions except the boreal arctic regions. With the effects of cloud cover removed, the correlation of DSRCS with the temperatures becomes insignificant except a positive

Table 4 Inter-model correlations between the temperatures (T_{max}, T_{min} and DTR) and key controlling variables of surface radiative and PBL mixing processes averaged by climate zone

Climate zone	Boreal arctic regions			Tropical rainforests			Transitional zones			Arid and semi-arid regions		
DPR (m)[a]	140			920			1702			2732		
	T_{max}	T_{min}	DTR	T_{max}	T_{min}	DTR	T_{max}	T_{min}	DTR	T_{max}	T_{min}	DTR
(a) Radiative variables												
DSRCF	0.05	0.23	−0.22	**−0.52**	0.04	**−0.56**	**−0.48**	0.06	**−0.56**	**−0.36**	0.18	**−0.46**
DSRCS	−0.02	0.16	−0.19	−0.11	0.09	−0.2	−0.08	0.18	−0.26	**0.33**	0.05	**0.3**
DLRCF	0.08	0.24	−0.20	**−0.33**	−0.13	−0.2	**−0.37**	−0.19	−0.15	−0.03	0.08	−0.1
DLRCS	**0.78**	**0.66**	0.1	**0.42**	**0.68**	**−0.3**	**0.56**	**0.7**	−0.2	**0.36**	**0.7**	**−0.4**
Albedo	**−0.46**	**−0.38**	−0.02	0.06	−0.1	0.17	**−0.36**	**−0.32**	0.04	−0.18	−0.13	−0.02
Emissivity	0.1	−0.07	0.16	0.1	−0.12	0.21	−0.11	0.1	−0.21	**−0.36**	−0.1	−0.17
(b) PBL mixing variables												
EF	0.1	**0.38**	**−0.36**	**−0.6**	−0.11	**−0.51**	**−0.48**	−0.06	**−0.39**	**−0.39**	−0.03	**−0.36**
SM	−0.06	0.11	−0.22	**−0.4**	0.01	**−0.38**	**−0.4**	−0.05	**−0.36**	**−0.42**	0.14	**−0.46**
SH$_{max}$	**0.45**	−0.14	**0.53**	**0.5**	−0.06	**0.63**	**0.59**	−0.1	**0.69**	**0.58**	−0.18	**0.68**
LCL	**0.33**	0.08	0.22	**0.75**	**0.54**	0.20	**0.69**	**0.64**	0.04	**0.44**	**0.48**	−0.05
SH$_{night}$	−0.08	**−0.40**	**0.33**	−0.15	**−0.46**	**0.33**	−0.02	**−0.33**	**0.32**	−0.07	**−0.36**	**0.32**

Values are areal mean correlation coefficients averaged by climate zone and those in boldface are considered statistically significant if at least half of the area of the climate zones have correlations passing the significance test at the 5 % level

[a] The climate zonal averaged DPR in meters is listed

correlation with T_{max} and DTR over the arid and semi-arid regions. DLRCF negatively correlates with T_{max} over the tropical rainforests and transitional zones, probably due to the greater effects of clouds on solar radiation. However, under clear-sky conditions, DLRCS shows a widespread positive correlation with T_{max} and T_{min}, which reflects a stronger dependence of the modeled temperatures on GHGs forced longwave radiation, especially for T_{min}. As for the land surface radiative parameters, the albedo negatively correlates with T_{max} and the subsequent T_{min} over the boreal cold areas and transitional zones, and a negative correlation between the emissivity and T_{max} exists over the arid and semi-arid regions. For the variables related to the PBL mixing, EF and SM, which act to hamper the levels of PBL mixing, negatively correlate with T_{max} and DTR over most regions. EF and T_{min} show a positive correlation over the boreal arctic areas, likely due to the enhanced warming effects of the water–vapor forced surface downward longwave radiation. As for the proxies for the PBL depth during the daytime, SH_{max} naturally increases T_{max} and thus DTR and the diurnal mean LCL both positively correlates with T_{max} and subsequent T_{min} over all regions. At night, the downgradient diffusion makes SH_{night} negatively correlate with T_{min} and hence positively correlate with DTR over most of the regions despite their nonlinear interactions as proposed by Holtslag et al. (2007). Overall, those significant correlative relationships in each climate zone depict the general dependence of the inter-model variations of the temperature diurnal cycle on surface radiative and PBL mixing processes.

As far as the two regions controlled by extreme PBL conditions are concerned, we could identify the major controlling factors as follows. Over the arid and semi-arid regions, the model differences in T_{max} are significantly controlled by DSRCF, DSRCS, DLRCS, land surface emissivity, EF, SM, SH_{max} and LCL, while those in T_{min} are mainly influenced by DLRCS, LCL and SH_{night}. Hence DSRCF, DSRCS, DLRCS, EF, SM, SH_{max} and SH_{night}, which significantly correlate with T_{max} and T_{min} asymmetrically, play an important role in determining the inter-model variations of DTR. Over the boreal arctic regions, DLRCS, the land surface albedo, SH_{max} and LCL significantly influence the inter-model variations of T_{max}, while DLRCS, the land surface albedo, EF and SH_{night} contribute largely to those of T_{min}. Although the DTR differs insignificantly among models over the boreal arctic regions, EF, SH_{max} and SH_{night}, which are significantly correlated with DTR, still have noticeable asymmetrical effects on T_{max} and T_{min}.

The relative contributions of these controlling factors to the inter-model variations of T_{max}, T_{min} and DTR are quantified using the variance analysis method based on the multiple linear regression. To reduce the collinearity among the variables, we only use EF and SH_{max} to represent the
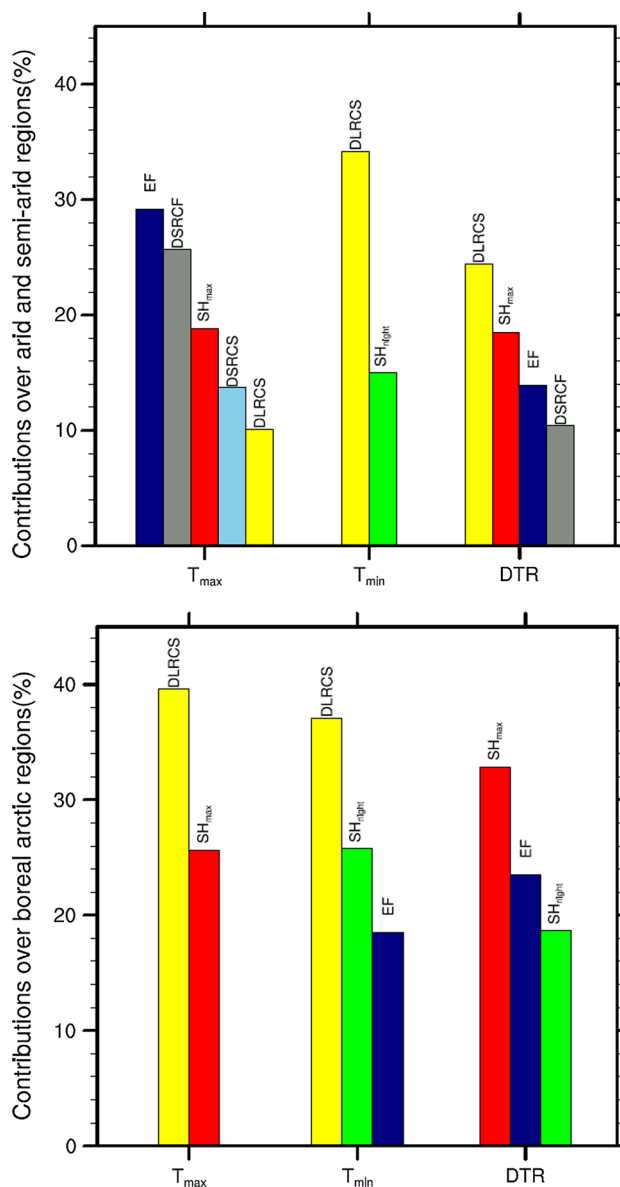
Fig. 10 Relative contributions of key controlling variables related to atmospheric radiative and PBL mixing processes to the inter-model variations of T_{max}, T_{min} and DTR over two climate zones with extreme PBL conditions: (*top panel*) arid and semi-arid climate regions and (*bottom panel*) boreal arctic regions

partitioning of the surface turbulent flux and the daytime PBL depth, while SM and LCL are excluded from the variance analysis. The percentage contributions of the controlling factors that pass the significance test over the two regions are shown in Fig. 10. For the arid and semi-arid regions, EF and DSRCF account for the largest amount (nearly 30 % each) of the inter-model SD of T_{max}, which is consistent with the notion that T_{max} is most affected by surface solar heating and the partitioning of turbulent fluxes for evaporation (e.g. Dai et al. 1999; Zhou et al. 2009b,

2010). SH_{max} explains up to 20 % of the inter-model SD of T_{max}. DSRCS and DLRCS are the other two significant factors for T_{max} variations that are associated with effects of atmospheric aerosols and GHGs. As for T_{min}, DLRCS associated with the warming effects of water vapor is the most important contributor and accounts for 34 % of the inter-model variations. SH_{night}, which represents the level of downward heat transport from the atmosphere, also considerably explains the T_{min} variations by 15 %. Consequently, the inter-model variations of DTR are explained by the variables with notably asymmetrical effects on T_{max} and T_{min}, namely DLRCS, SH_{max}, EF and DSRCF in a decreasing order. For the boreal arctic regions, DLRCS mostly explains both the inter-model variations of T_{max} and T_{min} by up to 40 %, followed by the PBL-mixing variables, namely SH_{max} and SH_{night}. EF also explains a comparable amount of the T_{min} variations, possibly due to the warming effects of the evaporated water vapor. The asymmetrical contributions of SH_{max}, EF and SH_{night} explain the inter-model variations of DTR although insignificant over the boreal arctic regions.

Here we emphasize the contributions of uncertainties in modeled PBL mixing to the inter-model variations of the temperature diurnal cycle. Under strong convective conditions such as the daytime PBL over the arid and semi-arid regions, the magnitude of T_{max} is very sensitive to changes in surface solar radiation and available SM for evaporation. Hence it is not surprising that the model differences in DSRCF and EF, which represent the uncertainties in modeled shortwave cloud radiative effects (Lewis and Karoly 2013) and turbulent moisture exchanges (Perkins et al. 2007), explain the most inter-model variations of T_{max}. If DSRCF and EF are fixed among the models, the next most notable contributor to the T_{max} variations is SH_{max}, representing the different degrees of PBL mixing resulting from various implementations of PBL schemes with different turbulent heat transfer coefficient, near-surface wind and entrainment zone elevation (e.g. Svensson et al. 2011). Such uncertainties could be even more significant than those originating from different inclusions of aerosol and GHG forcings among models as reflected by the smaller contributions of DSRCS and DLRCS to the T_{max} variations (Zhou et al. 2009b, 2010; Lindvall and Svensson 2015). In stable stratified conditions that often occur at nighttime in the arid and semi-arid regions and almost all-day in the boreal arctic regions, the PBL-mixing effects could be the dominant sources of the uncertainties in temperature modeling for three major reasons. First, more than half of the DLRCS originates within the lowest hundreds of meters of the atmosphere and so DLRCS is largely determined by temperature and moisture profiles controlled by the PBL mixing (Bosveld et al. 2014). Hence, the model differences in DLRCS, which explain the most of the

inter-model temperature variations, could be largely linked to the differences in PBL schemes. Second, the stable PBL conditions have been shown to be poorly simulated with overestimated turbulent mixing (e.g. Svensson et al. 2011; Holtslag et al. 2013; Bosveld et al. 2014). Different levels of this enhanced mixing among PBL schemes could significantly contribute to the uncertainties in temperature simulations, as reflected by the large inter-model variations of T_{min} accounted by SH_{night}, which is also diagnosed by Sandu et al. (2013) and Sterk et al. (2013). Last but not the least, the temperatures is extremely sensitive to external forcings in stable conditions because of the small atmospheric effective heat capacity (Esau et al. 2012; Davy and Esau 2014a, b; Davy et al. 2016). Given this small heat capacity, any slight differences in external forcings among models can lead to large inter-model temperature variations as indicated by Davy and Esau (2014b). This explains why the inter-model SD of T_{min} is overall larger than that of T_{max} and hence determines the inter-model SD of DTR (Fig. 8). Therefore, accurate and consistent representations of the diurnal PBL processes are needed, particularly in stably-stratified conditions, to reduce the uncertainties of the simulated temperature diurnal cycle.

Our above analyses indicate that the model uncertainties in PBL turbulent mixing could make significant contributions to large inter-model variations of the temperature diurnal cycle, especially over the regions with diurnal extreme (strong convective and stably-stratified) conditions, namely the arid and semi-arid regions and boreal arctic regions. Although the selected surface radiative and PBL mixing controlling factors cannot completely explain the inter-model temperature differences, the variance analysis indicates that the PBL-mixing processes could be as important as the surface radiative forcings which are direct drivers for temperature variations. Similar to Holtslag et al. (2013), our results highlight the need for a more reliable description of PBL processes associated with the turbulent mixing in GCMs to reduce the uncertainties in modelling the temperature diurnal evolution.

4 Conclusions and future prospects

In this research, we examine the effects of modeled PBL mixing on the simulated temperature diurnal cycle climatology over land in 20 CMIP5 models with AMIP simulations. When compared with observations, the magnitude of DTR is systematically underestimated over almost all land areas due to a widespread warm bias of T_{min} and mostly a cold bias of T_{max}. Analyses of the CMIP5 multi-model ensemble means suggest that the biases of the simulated PBL mixing could very likely contribute to the temperature biases. For the regions with the cold bias in T_{max}, the daytime PBL mixing

is generally underestimated. The consequent more dry air entrainment from the free atmosphere could help maintain the surface humidity gradient, and thus produce more surface evaporation and potentially lower the T_{max}. The opposite situation holds true for the regions with the warm bias of T_{max}. This mechanism could be particularly applicable to the regions with moderate and wet conditions where the surface evaporation depends more on the surface humidity gradient, but less on the available SM. For the widespread warm bias of T_{min}, the widely-recognized enhanced PBL mixing at nighttime should play a dominant role by transferring more heat from the atmosphere to the near-surface to warm the T_{min}. Our further analyses using the high resolution CFMIP2 output also support the CMIP5 results about the connections of the biases between the simulated turbulent mixing and the temperature diurnal cycle. The large inter-model variations of the simulated temperature diurnal cycle primarily appear over the arid and semi-arid regions and boreal arctic regions where the model differences in the PBL turbulence mixing could make equally significant contributions to the inter-model variations of DTR, T_{max} and T_{min} compared to the model differences in surface radiative processes. These results highlight the importance and need for accurate descriptions of the PBL processes with respect to the turbulent mixing in order to improve the temperature diurnal cycle simulations in GCMs.

We realize that the interactive mechanisms between the near-surface temperatures and PBL processes are more complex than what is analyzed here based on the statistical techniques. The large-scale circulation and many small-scale physical processes in the PBL such as radiation divergence, gravity waves and low-level jets can all modify the surface energy budgets and thus the temperature diurnal cycle (Holtslag et al. 2013; Lewis and Karoly 2013; Svensson and Lindvall 2015). However, the limitations of our understanding of complex PBL processes, the challenges in modeling the PBL processes, and the lack of high resolution model output and observations for turbulent parameters hinders our further analysis. Nevertheless, we establish a preliminary relationship between the effects of PBL mixing and the modeled temperature diurnal cycle in this work, and will carry out further attribution studies by conducting carefully designed sensitivity experiments with individual models. Our ultimate goal for the future work is to find a unified treatment of PBL processes that could reproduce the main climatic features on the diurnal cycle of near-surface variables in GCMs with unnecessary small-scale processes filtered out.

References

Betts AK (2004) Understanding hydrometeorology using global models. Bull Am Meteorol Soc 85:1673–1688. doi:10.1175/BAMS-85-11-1673

Betts AK, Barr AG (1996) First international satellite land surface climatology field experiment 1987 sonde budget revisited. J Geophys Res Atmos 101:23285–23288. doi:10.1029/96JD02247

Bonan GB (2001) Observational evidence for reduction of daily maximum temperature by croplands in the Midwest United States. J Clim 14:2430–2442. doi:10.1175/1520-0442(2001)

Bonan GB, Oleson KW, Vertenstein M, Levis S, Zeng XB, Dai YJ, Dickinson RE, Yang ZL (2002) The land surface climatology of the community land model coupled to the NCAR Community Climate Model. J Clim 15:3123–3149. doi:10.1175/1520-0442(2002)

Bosveld FC et al (2014) The third GABLS intercomparison case for evaluation studies of boundary-layer models. Part B: results and process understanding. Bound Layer Meteorol 152:157–187. doi:10.1007/s10546-014-9919-1

Cattiaux J, Douville H, Peings Y (2013) European temperatures in CMIP5: origins of present-day biases and future uncertainties. Clim Dyn 41:2889–2907. doi:10.1007/s00382-013-1731-y

Collatz GJ, Bounoua L, Los SO, Randall DA, Fung IY, Sellers PJ (2000) A mechanism for influence of vegetation on the response of the diurnal temperature range to changing climate. Geophys Res Lett 27:3381–3384. doi:10.1029/1999GL010947

Cuxart J et al (2006) Single-column model intercomparison for a stably stratified atmospheric boundary layer. Bound Layer Meteorol 118(2):273–303. doi:10.1007/s10546-005-3780-1

Dai AG, Trenberth KE, Karl TR (1999) Effects of clouds, soil moisture, precipitation, and water vapor on diurnal temperature range. J Clim 12:2451–2473. doi:10.1175/1520-0442(1999)

Davy R, Esau I (2014a) Surface air temperature variability in global climate models. Atmos Sci Lett 15:13–20. doi:10.1002/asl2.456

Davy R, Esau I (2014b) Global climate models' bias in surface temperature trends and variability. Environ Res Lett 9:114024. doi:10.1088/1748-9326/9/11/114024

Davy R, Esau I, Chernokulsky A, Outten S, Zilitinkevich S (2016) Diurnal asymmetry to the observed global warming. Int J Climatol. doi:10.1002/joc.4688

Decker M, Brunke MA, Wang Z, Sakaguchi K, Zeng X, Bosilovich MG (2012) Evaluation of the reanalysis products from GSFC, NCEP, and ECMWF using flux tower observations. J Clim 25:1916–1944. doi:10.1175/JCLI-D-11-00004.1

Dee DP et al (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q J R Meteorol Soc 137:553–597. doi:10.1002/qj.828

Dirmeyer PA, Jin Y, Singh B, Yan XQ (2013) Trends in land–atmosphere interactions from CMIP5 simulations. J Hydrometeorol 14:829–849. doi:10.1175/JHM-D-12-0107.1

Dirmeyer PA, Wang Z, Mbuh MJ, Norton HE (2014) Intensified land surface control on boundary layer growth in a changing climate. Geophys Res Lett 41:1290–1294. doi:10.1002/2013GL058826

Engeln AV, Teixeira J (2013) A planetary boundary layer height climatology derived from ECMWF reanalysis data. J Clim 26:6575–6590. doi:10.1175/JCLI-D-12-00385.1

Esau I, Ziltinkevich S (2010) On the role of the planetary boundary layer depth in the climate system. Adv Sci Res 4:63–69. doi:10.5194/asr-4-63-2010

Esau I, Davy R, Outten S (2012) Complementary explanation of temperature response in the lower atmosphere. Environ Res Lett 7:044026. doi:10.1088/1748-9326/7/4/044026

Gallo KP, Easterling DR, Peterson TC (1996) The influence of land use/land cover on climatological values of the diurnal temperature range. J Clim 9:2941–2944. doi:10.1175/1520-0442(1996)

Gentine P, Entekhabi D, Polcher J (2011) The diurnal behavior of evaporative fraction in the soil–vegetation–atmospheric boundary layer continuum. J Hydrometeorol 12:1530–1546. doi:10.1175/2011JHM1261.1

Hoerling M, Eischeid J, Perlwitz J (2010) Regional precipitation trends: distinguishing natural variability from anthropogenic forcing. J Clim 23:2131–2145. doi:10.1175/2009JCLI3420.1

Holtslag AAM, Steeneveld GJ, van deWiel BJH (2007) Role of land-surface temperature feedback on model performance for the stable boundary layer. Bound Layer Meteorol 125:361–376. doi:10.1007/s10546-007-9214-5

Holtslag AAM et al (2013) Stable atmospheric boundary layers and diurnal cycles: challenges for weather and climate models. Bull Am Meteorol Soc 94:1691–1706. doi:10.1175/BAMS-D-11-00187.1

Hu X, Nielsen-Gammon JW, Zhang F (2010) Evaluation of three planetary boundary layer schemes in the WRF model. J Appl Meteorol Climatol 49:1831–1844. doi:10.1175/2010JAMC2432.1

Hua WJ, Chen HS, Zhu SG, Sun SL, Yu M, Zhou L (2013) Hotspots of the sensitivity of the land surface hydrological cycle to climate change. Chin Sci Bull 58(30):3682–3688. doi:10.1007/s11434-013-5846-7

Hua WJ, Chen HS, Sun SL (2014) Uncertainty in land surface temperature simulation over China by CMIP3/CMIP5 models. Theor Appl Climatol 117(3–4):463–474. doi:10.1007/s00704-013-1020-z

IPCC (2007) Climate change 2007: the physical science basis, contribution of Working Group I to the Fourth Assessment Report of the IPCC. Cambridge University Press, Cambridge. ISBN:978- 0-521-88009-1

IPCC (2013) Climate change 2013: the physical science basis, the contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge. ISBN:978-1-107-05799-1

Kharin VV, Zwiers FW, Zhang XB, Hegerl GC (2007) Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. J Clim 20:1419–1444. doi:10.1175/JCLI4066.1

Kleczek MA, Steeneveld GJ, Holtslag AAM (2014) Evaluation of the weather research and forecasting mesoscale model for GABLS3: impact of boundary-layer schemes, boundary conditions and spin-up. Bound Layer Meteorol 152:213–243. doi:10.1007/s10546-014-9925-3

Koster RD et al (2004) Regions of coupling between soil moisture and precipitation. Science 305:1138–1140. doi:10.1126/science.1100217

Lewis SC, Karoly DJ (2013) Evaluation of historical diurnal temperature range trends in CMIP5 models. J Clim 26:9077–9089. doi:10.1175/JCLI-D-13-00032.1

Lindvall J, Svensson G (2015) The diurnal temperature range in the CMIP5 models. Clim Dyn 44:405–421. doi:10.1007/s00382-014-2144-2

Lindvall J, Svensson G, Hannay C (2013) Evaluation of near-surface parameters in the two versions of the atmospheric model in cesm1 using flux station observations. J Clim 26:26–44. doi:10.1175/JCLI-D-12-00020.1

Liu S, Liang XZ (2010) Observed diurnal cycle climatology of planetary boundary layer height. J Clim 23:5790–5809. doi:10.1175/2010JCLI3552.1

Lobell DB, Bonfils C, Duffy PB (2007) Climate change uncertainty for daily minimum and maximum temperatures: a model inter-comparison. Geophys Res Lett 34:L05715. doi:10.1029/2006GL028726

Mitchel TD, Jones PD (2005) An improved method of constructing a database of monthly climate observations and associated high-resolution grids. Int J Climatol 25:693–712. doi:10.1002/joc.1181

Perkins SE, Pitman AJ, Holbrook NJ, McAneney J (2007) Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. J Clim 20:4356–4376. doi:10.1175/JCLI4253.1

Pithan F, Angevine W, Mauritsen T (2015) Improving a global model from the boundary layer: total turbulent energy and the neutral limit Prandtl number. J Adv Model Earth Syst 7:791–805. doi:10.1002/2014MS000382

Rutan DA, Kato S, Doelling DR, Rose FG, Loeb N, Caldwell T, Nguyen LTC (2014) Validation of surface irradiance calculated in NASA's clouds and the earth radiant energy system (CERES) SYN and EBAF data products. In: Papers from 14th AMS conference on atmospheric radiation. https://ams.confex.com/ams/14CLOUD14ATRAD/webprogram/Paper249853.html

Sandu I, Beljaars A, Bechtold P, Mauritsen T, Balsamo G (2013) Why is it so difficult to represent stably stratified conditions in numerical weather prediction (NWP) models? J Adv Model Earth Syst 5:117–133. doi:10.1002/jame.20013

Scheff J, Frierson DMW (2015) Terrestrial aridity and its response to greenhouse warming across CMIP5 climate models. J Clim 28:5583–5600. doi:10.1175/JCLI-D-14-00480.1

Seidel DJ, Zhang Y, Beljaars A, Golaz J-C, Jacobson AR, Medeiros B (2012) Climatology of the planetary boundary layer over the continental United States and Europe. J Geophys Res 117:D17106. doi:10.1029/2012JD018143

Shao Y, Ishizuka M, Mikami M, Leys JF (2011) Parameterization of size-resolved dust emission and validation with measurements. J Geophys Res 116:D08203. doi:10.1029/2010JD014527

Shin HH, Hong S (2011) Intercomparison of planetary boundary-layer parametrizations in the WRF model for a single day from CASES-99. Bound Layer Meteorol 139:261–281. doi:10.1007/s10546-010-9583-z

Sillmann J, Kharin VV, Zhang X, Zwiers FW, Bronaugh D (2013) Climate extremes indices in the CMIP5 multi-model ensemble: part 1. Model evaluation in the present climate. J Geophys Res Atmos 118:1716–1733. doi:10.1002/jgrd.50203

Simmons AJ, Willett KM, Jones PD, Thorne PW, Dee DP (2010) Low-frequency variations in surface atmospheric humidity, temperature, and precipitation: inferences from reanalyses and monthly gridded observational data sets. J Geophys Res 115:D01110. doi:10.1029/2009JD012442

Steeneveld GJ, Mauritsen T, de Bruijn EIF, Vila-Guerau de Arellano J, Svensson G, Holtslag AAM (2008) Evaluation of limited area models for the representation of the diurnal cycle and contrasting nights in CASES99. J Appl Meteorol Climatol 47:869–887. doi:10.1175/2007JAMC1702.1

Sterk HAM, Steeneveld GJ, Holtslag AAM (2013) The role of snow-surface coupling, radiation, and turbulent mixing in modeling a stable boundary layer over Arctic sea ice. J Geophys Res Atmos 118:1199–1217. doi:10.1002/jgrd.50158

Svensson G, Lindvall J (2015) Evaluation of near-surface variables and the vertical structure of the boundary layer in CMIP5 models. J Clim 28:5233–5253. doi:10.1175/JCLI-D-14-00596.1

Svensson G et al (2011) Evaluation of the diurnal cycle in the atmospheric boundary layer over land as represented by a variety of single-column models: the second GABLS experiment. Bound Layer Meteorol 140:177–206. doi:10.1007/s10546-011-9611-7

Van Heerwaarden CC, Vilà-Guerau de Arellano J, Moene AF, Holtslag AAM (2009) Interactions between dry-air entrainment, surface evaporation and convective boundary-layer development. Q J R Meteorol Soc 135:1277–1291. doi:10.1002/qj.431

Viterbo P, Beljaars A, Mahfouf J-F, Teixeira J (1999) The representation of soil moisture freezing and its impact on the stable boundary layer. Q J R Meteorol Soc 125:2401–2426. doi:10.1002/qj.49712555904

Vogelezang DHP, Holtslag AAM (1996) Evaluation and model impacts of alternative boundary-layer height formulations. Bound Layer Meteorol 81:245–269. doi:10.1007/BF02430331

Wang K, Dickinson RE (2013) Global atmospheric downward longwave radiation at the surface from ground-based observations, satellite retrievals, and reanalyses. Rev Geophys 51:150–185. doi:10.1002/rog.20009

Wang AH, Zeng XB (2013) Development of global hourly 0.5° land surface air temperature datasets. J Clim 26:7676–7691. doi:10.1175/JCLI-D-12-00682.1

Wang AH, Zeng XB (2014) Range of monthly mean hourly land surface air temperature diurnal cycle over high northern latitudes. J Geophys Res Atmos 119:5836–5844. doi:10.1002/2014JD021602

Wild M (2009) How well do IPCC-AR4/CMIP3 climate models simulate global dimming/brightening and twentieth-century daytime and nighttime warming? J Geophys Res 114:D00D11. doi:10.1029/2008JD011372

Willmott CJ (1984) On the evaluation of model performance in physical geography. In: Gaile GL, Willmott CJ (eds) Spatial statistics and models. Springer, Dordrecht. ISBN:978-90-481-8385-2

Xie S et al (2010) Clouds and more: ARM climate modeling best estimate data. Bull Am Meteorol Soc 91:13–20. doi:10.1175/2009BAMS2891.1

Yi C et al (2010) Climate control of terrestrial carbon exchange across biomes and continents. Environ Res Lett 5:034007. doi:10.1088/1748-9326/5/3/034007

Zhang D, Zheng W (2004) Diurnal cycles of surface winds and temperatures as simulated by five boundary layer parameterizations. J Appl Meteorol 43:157–169. doi:10.1175/1520-0450

Zhang JY, Wang WC, Wu LY (2009) Land–atmosphere coupling and diurnal temperature range over the contiguous United States. Geophys Res Lett 36:L06706. doi:10.1029/2009GL037505

Zhang X, Liang S, Wang G, Yao Y, Jiang B, Cheng J (2016) Evaluation of the reanalysis surface incident shortwave radiation products from NCEP, ECMWF, GSFC, and JMA using satellite and surface observations. Remote Sens 8:225. doi:10.3390/rs8030225

Zhou L, Dickinson RE, Tian Y, Vose RS, Dai YJ (2007) Impact of vegetation removal and soil aridation on diurnal temperature range in a semiarid region: application to the Sahel. Proc Natl Acad Sci USA 104:17937–17942. doi:10.1073/pnas.0700290104

Zhou L, Dickinson RE, Dirmeyer P, Chen HS, Dai YJ, Tian Y (2008) Asymmetric response of maximum and minimum temperatures to soil emissivity change over the Northern African Sahel in a GCM. Geophys Res Lett 35:L05402. doi:10.1029/2007GL032953

Zhou L, Dai AG, Dai YJ, Vose RS, Zou CZ, Tian Y, Chen HS (2009a) Spatial dependence of diurnal temperature range trends on precipitation from 1950 to 2004. Clim Dyn 32:429–440. doi:10.1007/s00382-008-0387-5

Zhou L, Dickinson RE, Dirmeyer P, Dai A, Min SK (2009b) Spatiotemporal patterns of changes in maximum and minimum temperatures in multi-model simulations. Geophys Res Lett 36:L02702. doi:10.1029/2008GL036141

Zhou L, Dickinson RE, Dai A, Dirmeyer P (2010) Detection and attribution of anthropogenic forcing to diurnal temperature range changes from 1950 to 1999: comparing multi-model simulations with observations. Clim Dyn 35:1289–1307. doi:10.1007/s00382-009-0644-2