Changes in the ground surface temperature in permafrost regions along the Qinghai−Tibet engineering corridor from 1900 to 2014: A modified assessment of CMIP6

Zan-Pin XING [a,b], Lin ZHAO [b,c,*], Lei FAN [d], Guo-Jie HU [a], De-Fu ZOU [a], Chong WANG [c], Shu-Ci LIU [e], Er-Ji DU [a], Yao XIAO [a], Ren LI [a], Guang-Yue LIU [a], Yong-Ping QIAO [a], Jian-Zong SHI [a]

[a] *Cryosphere Research Station on the Qinghai−Tibet Plateau, State Key Laboratory of Cryospheric Science, Northwest Institute of Eco-Environment and Resource, Chinese Academy of Sciences, Lanzhou 730000, China*
[b] *University of Chinese Academy of Sciences, Beijing 100049, China*
[c] *School of Geographical Sciences, Nanjing University of Information Science & Technology, Nanjing 210044, China*
[d] *Chongqing Jinfo Mountain Karst Ecosystem National Observation and Research Station, School of Geographical Sciences, Southwest University, Chongqing 400715, China*
[e] *Department of Environment and Science, Queensland Government, Brisbane, QLD 4102, Australia*

Abstract

Numerous studies were published in the last two decades to evaluate and project the permafrost changes in its thermal state, mainly based on the soil temperature datasets from the Coupled Model Intercomparison Project (CMIP), and discuss the impacts of permafrost changes on regional hydrological, ecological and climatic systems and even carbon cycles. However, limited monitored soil temperature data are available to validate the CMIP outputs, resulting in the over-projection of future permafrost changes in CMIP3 and CMIP5. Moreover, future permafrost changes in CMIP6, particularly over the Qinghai−Tibet Plateau (QTP), where permafrost covers more than 40% of its territory, are still un-known. To address this gap, we evaluated and calibrated the monthly ground surface temperature (GST; 5 cm below the ground surface), which was often used as the upper boundary to simulate and project permafrost changes derived from 19 CMIP6 Earth System Models (ESMs) against *in situ* measurements over the QTP. We generated the monthly GST series from 1900 to 2014 for five sites based on the linear calibration models and validated them through the three other sites using the same calibration methods. Results showed that all of the ESMs could capture the dynamics of *in situ* GST with high correlations ($r > 0.90$). However, large errors were detected with a broad range of centred root-mean-square errors (1.14−4.98 °C). The Top 5 model ensembles (MME5) performed better than most individual ESMs and averaged multi-model ensembles (MME19). The calibrated GST performed better than the GST obtained from MME5. Both annual and seasonal GSTs exhibited warming trends with an average annual rate of 0.04 °C per decade in the annual GST. The average seasonal warming rate was highest in winter and spring and lowest in summer. This reconstructed GST data series could be used to simulate the long-term permafrost temperature over the QTP.

Keywords: Ground surface temperature; CMIP6; Cold bias; Permafrost; Qinghai−Tibet engineering corridor

* Corresponding author. Cryosphere Research Station on the Qinghai−Tibet Plateau, State Key Laboratory of Cryospheric Science, Northwest Institute of Eco-Environment and Resource, Chinese Academy of Sciences, Lanzhou 730000, China.

E-mail address: lzhao@nuist.edu.cn (ZHAO L.).

1. Introduction

The Qinghai−Tibet Plateau (QTP) occupies the largest permafrost area (approximately 1.06 million km^2) in the middle-latitude and low-latitude regions (Zou et al., 2017).

Permafrost reflects energy exchange (Guo et al., 2011), hydrological processes (Walvoord and Kurylyk, 2016), biogeochemical processes (Mu et al., 2017) and engineering activities (Wu et al., 2015). Ground surface temperature (GST) is generally defined as the ground temperature at a certain depth between 0 cm and 10 cm below the ground surface. GST is one of the critical parameters in investigating permafrost changes because it is an essential indicator of climate change and an upper boundary in permafrost modelling (Guo et al., 2018). GST variation can reflect active layer changes (Zhao et al., 2011), and its seasonal variation is directly related to the heat exchange between the surface and the atmosphere, thus affecting the regional climate and East Asia's atmospheric circulation (Zhang et al., 2008). Permafrost is the long-term climatic consequence of the land−atmosphere interaction under specific climatic conditions, which respond slowly to climate change. Therefore, the investigation of long-term GST variability enables us to better understand the climate changes in permafrost regions and their ecosystem stabilities.

Because of the harsh natural environment and inconvenient accessibility over the permafrost regions of the QTP, research is often impeded by sparse observations (Fang et al., 2018; Liu et al., 2015; Zhu et al., 2017). However, alternative and reliable long-term datasets over the large spatial scale of GST are needed. Reanalysis and satellite products have been widely used to cope with this deficiency. Most previous studies have indicated that reanalysis products had large uncertainties over the permafrost regions of the QTP compared with *in situ* observations because of assimilation, quality control and forecast modelling (Hu et al., 2019a; Huang et al., 2022; Qin et al., 2017, 2020; Yang et al., 2020). Moreover, most of the reanalysis and satellite products can only be available from 1979 or later, which fails to extend GST changes to a more extended period (Ding et al., 2018; Guo, 2017). Nevertheless, with the growth and development of numerical modelling simulation, high spatiotemporal GST can be simulated.

Earth System Models (ESMs), developed by various modelling groups across the globe under the aegis of the Coupled Model Intercomparison Project (CMIP) since the 1990s, has been widely used to investigate the impacts of past, present and future climate changes at the global scale (Meehl et al., 1997; Taylor et al., 2007, 2012). Nevertheless, the soil temperature simulated by the CMIP3 and CMIP5 models tended to over-project future permafrost changes. Koven et al. (2013) and Lawrence et al. (2012) indicated that the earlier intercomparison phases (e.g. CMIP3 and CMIP5) could not simulate GST well over the permafrost regions. For instance, Slater and Lawrence (2013) detected permafrost over various models. They reported that CMIP5 models overestimated the mean soil temperature from 1986 to 2005 by more than 7 °C in the present-day permafrost region, resulting in rapid permafrost degradation (Burke et al., 2020; Guo et al., 2012). The permafrost projections also showed that the permafrost distribution would decrease by 60%−80% or even disappear on the QTP by 2100 under the highest Representative Concentration Pathway (RCP8.5) scenario (Anisimov and Nelson, 1996; Guo and Wang, 2016; Lu et al., 2017; Zhang et al.,

2022). Recently, the slow thermal response of permafrost to climate warming was noted at the two investigated sites over the permafrost regions of the QTP by Sun et al. (2019), who reported that shallow permafrost might disappear, but deep permafrost would persist by 2100 even under the most radical warming scenario (RCP8.5). Thus, the ESM-based CMIP6 GST needs to be evaluated and optimised before its applications over the permafrost regions of the QTP.

Compared with the earlier intercomparison phases, CMIP6 has addressed the long-standing problems of model bias and poor quantification of radiative forcing in CMIP5 (Eyring et al., 2016; Stouffer et al., 2017). Moreover, the CMIP6 models were considered to better represent the quasi-biennial oscillation variability and the dynamics of the Indian summer monsoon (Gusain et al., 2020; Li et al., 2020). Therefore, the GST dataset of CMIP6 ESMs over the permafrost regions of the QTP needs to be evaluated such that it can be used to assess the permafrost thermal state and changes. However, previous research on the evaluation of CMIP6 ESMs over the QTP mainly focused on meteorological variables (e.g. precipitation and surface air temperature (TAS); (Gao et al., 2022; Lun et al., 2021; Shang et al., 2021; Wu et al., 2020; Zhu and Yang, 2020), surface energy flux (Ma et al., 2022) and surface soil moisture (SSM) (Li et al., 2017). Because of scarce GST observations over the permafrost regions of the QTP, few works were conducted in this region to evaluate the performance of the CMIP6 ESM-based GST.

In this study, we focused on permafrost regions of the QTP to 1) evaluate the performance of 19 ESMs archived by the state-of-the-art CMIP6 in reproducing GST relative to *in situ* measurements, 2) establish the linear calibration models to obtain accurate long-term monthly GST and 3) analyse the annual and seasonal GST variations over the past 100 years to better understand the thermal response of permafrost to climate warming.

2. Materials and methods

2.1. In situ measurements

The active layer's soil temperature and moisture monitoring network has been deployed on the QTP by the Cryosphere Research Station, Chinese Academy of Sciences (Zhao et al., 2021). Five sites on this network, from the northern boundary to the southern boundary over the permafrost regions of the QTP, were selected to evaluate the GST (Fig. 1 and Table 1). The flat terrain condition of all five sites enabled the observed soil temperatures to better represent a large region (Fig. A1). The surface vegetation conditions are dominated by grassland, specifically alpine wet meadow (CN04), alpine meadow (QT04 and QT09), alpine steppe (CN03) and alpine desert steppe (QT08). Soil temperatures were monitored by 105T thermocouple probes (Campbell, USA) with an accuracy of ±0.5 °C (Zhao et al., 2021). At each site, soil temperature sensors are horizontally inserted into 1, 2 or 5 cm topsoil, and the measuring time interval is set to 30 min. We also collected daily SSM (0−10 cm) for QT08 and QT09 from 2013 to 2014
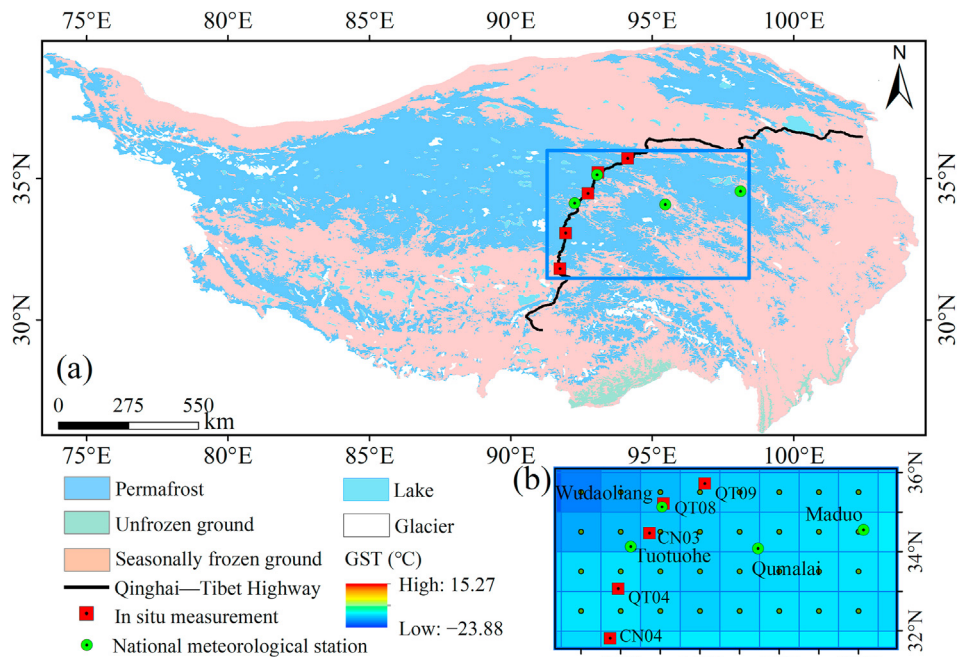
Fig. 1. Location of the monitoring sites and study area, (a) *in situ* sites, and (b) grid cells of the study area with a resolution of $1° \times 1°$ (The frozen ground map was derived from Zou et al. (2017)).

and TAS at 2 m height for QT04, QT08 and QT09 from 2010 to 2014 to explore the uncertainties of ESMs. The SSM was measured using the Hydra soil moisture sensor (Campbell, USA) with an accuracy of ±2.5%, and the TAS was measured using the HMP45C temperature/relative humidity sensor with an accuracy of ±0.5 °C. The SSM from July to August in the thawed soil at each site was used as the modelled SSM considering both the liquid and solid water contents. However, the *in situ* sensor is only sensitive to the liquid water content.

Furthermore, to validate the established linear calibration models, the monthly GSTs at the Tuotuohe (alpine steppe), Qumalai (alpine meadow) and Maduo (alpine wet meadow) sites from 1961 to 2014 were collected from the China Meteorological Administration (http://data.cma.cn/; Fig. 1). The GLDAS-Noah reanalysis GST product from 1961 to 2014 was also collected to compare with the reconstructed GST in the subsequent analyses. Only high-quality *in situ* measurements after quality controls were used. Their locations, active layer thickness and available periods are listed in Table 1 (see Zhao et al. (2021) for more details on the quality control procedures).

2.2. CMIP6 ESMs

We focused on the CMIP historical simulations (1850−2014) of the CMIP6 project and selected 19 ESMs (Table 2) as the variant label of r1i1p1f1 had been completed. 'ripf' (i.e. r: realization; i: initialization; p: physics; and f: forcing) distinguishes each member of an ensemble (see https://pcmdi.llnl.gov/CMIP6/Guide/dataUsers.html for more details). CMIP is an all-forcing simulation of past climate and is primarily based on observations (Pascoe et al., 2019). We only analysed the monthly ESMs because of insufficient daily or sub-daily ESM soil temperature datasets. The following monthly output variables from the CMIP historical simulations were collected: 1) tsl: temperature of soil; 2) tas: 2-m near-surface air temperature; and 3) mrsos: moisture in the upper portion of the soil column (0−10 cm). Information, including modelling groups, countries, abbreviations, resolution, land surface model, soil layer and depth of each experiment, is shown in Table 2. Further information on their configurations or features can be obtained from the CMIP6 website at https://esgf-node.llnl.gov/search/cmip6/.

Table 1
In situ sites used.

Site	Location	Longitude (°E)	Latitude (°N)	Altitude (m)	Active layer thickness (m)	Vegetation type	Period	Variable
CN03	Wuli	92.73	34.47	4625	2.8	Alpine steppe	2002−2014	GST
CN04	Liangdaohe	91.74	31.82	4808	1.2	Alpine wet meadow	2002−2014	GST
QT04	Tanggula	91.94	33.07	5100	3.5	Alpine meadow	2007−2014	GST and TAS
QT08	Wudaoliang	93.08	35.22	4783	2.4	Alpine desert steppe	2010−2014	GST, TAS and SSM
QT09	Xidatan	94.13	35.72	4538	1.3	Alpine meadow	2008−2014	GST, TAS and SSM

Table 2
Summary of the Coupled Model Intercomparison Project Phase 6 (CMIP6) ESMs used.

Model	Institute	Country	Grid number (longitude × latitude)	Land surface model	Soil layers	Soil depth (m)	Reference
ACCESS-ESM1-5	CSIRO-ARCCSS	Australia	192 × 145	CABLE2.4	6	2.9	Ziehn et al. (2020)
AWI-ESM-1-1-LR	AWI	Germany	192 × 96	JSBACH3.2	5	7	Danek et al. (2020)
BCC-CSM2-MR	BCC	China	320 × 160	BCC_AVIM2	10	2.9	Li et al. (2019)
CanESM5	CCCma	Canada	128 × 64	CLASS3.6.2/CTEM1.2	3	4.1	Swart et al. (2019)
CAS-ESM2-0	CAS	China	256 × 128	CoLM/IPA_DGVM/ CoLM-LBCM	15	35.2	Zhang et al. (2020)
FGOALS-f3-L			288 × 192	CLM4	15	35.2	He et al. (2019)
CESM2	NCAR	USA	288 × 192	CLM5	25	42	Danabasoglu et al. (2020)
EC-Earth3	EC-Earth-Consortium	Europe	512 × 256	HTESSEL	4	2	Döscher et al. (2020)
E3SM-1-0	E3SM	USA	360 × 180	CLM4.5	15	35.2	Golaz et al. (2019)
GFDL-CM4	NOAA-GFDL	USA	288 × 180	LM4.0.1	20	8.8	Held et al. (2019)
GISS-E2-1-G	NASA-GISS	USA	144 × 90	GISS LSM6	6	2.8	Miller et al. (2021)
IPSL-CM6A-LR	IPSL	France	144 × 143	ORCHIDEE2.0	18	65.6	Boucher et al. (2020)
MPI-ESM1-2-LR	MPI-M	Germany	192 × 96	JSBACH3.2	5	7	Mauritsen et al. (2019)
MPI-ESM1-2-HR			384 × 192	JSBACH3.2	5	7	
MIROC6	MIROC	Japan	256 × 128	MATSIRO6.0	6	9	Tatebe et al. (2019)
MRI-ESM2-0	MRI	Japan	320 × 160	HAL 1.0 14	14	8.5	Yukimoto et al. (2019)
NorESM2-LM	NCC	Norway	144 × 96	CLM	25	42	Tjiputra et al. (2020)
NorESM2-MM			288 × 192	CLM	25	42	
TaiESM1	AS-RCEC	China	288 × 192	CLM	15	35.2	Lee and Liang (2020)

2.3. Method

2.3.1. Preprocessing

Because of the different temporal resolutions of the *in situ* measurements and CMIP6 ESM outputs, the *in situ* measurements were processed into monthly average GST, SSM and TAS. The topsoil depths of the 19 ESM soil temperature datasets range from 0.048 to 10 cm. To better match surface soil depth between the CMIP6 ESM outputs and *in situ* measurements, the tsl and mrsos datasets were vertically linear interpolated into 5 cm to obtain GST and SSM because it is a convenient depth for all ESMs (Cao et al., 2020). Given that the model outputs were in different spatial resolutions, all ESM datasets were re-gridded to a common 1.0° × 1.0° resolution using the nearest interpolation following Hu et al. (2019b). Differences among the grid cells of the ESMs and *in situ* sites could be detected. Thus, we use centred root−mean−square error (cRMSE) and unbiased root−mean−square error (ubRMSE) as the evaluation metrics of performance because they are less impacted (i.e. less than bias) by the spatial mismatch between ESMs and *in situ* measurements. Given that the SSM unit of CMIP6 ESMs is kg m^{-2}, for comparison, it was converted into volumetric water content (m^3 m^{-3}) as instructed by Cui et al. (2018).

2.3.2. Evaluation metrics

The Taylor diagrams (Taylor, 2001), which provide a way to graphically summarise how closely a pattern matches the observations, were used to evaluate the performance of ESM-based GST (GST$_{ESM}$) against *in situ* GST (GST$_{REF}$) from 2010 to 2014. To ensure fair intercomparisons, the selection of the evaluation period is a compromise as the record for the

QT08 site is available for a shorter period (2010−2014) than the other sites. The normalized standard deviation (SDV; Eq. 1) indicates the ratio of the SDV of GST$_{ESM}$ to the SDV of GST$_{REF}$. r (Eq. 2) and cRMSE (Eq. 3) are the Pearson correlation coefficient and the cRMSE between GST$_{ESM}$ and GST$_{REF}$, respectively.

$$SDV = \frac{\sigma_{GST_{ESM}}}{\sigma_{GST_{REF}}} \tag{1}$$

$$r = \frac{1}{N-1} \sum_{i=1}^{N} \left(\frac{GST_{ESMi} - \overline{GST_{ESM}}}{\sigma_{GST_{ESM}}} \right) \left(\frac{GST_{REFi} - \overline{GST_{REF}}}{\sigma_{GST_{REF}}} \right) \tag{2}$$

$$cRMSE^2 = \frac{RMSE^2 - Bias^2}{\sigma_{GST_{REF}}^2} \tag{3}$$

ubRMSE (°C) is defined as follows:

$$ubRMSE = \sqrt{RMSE^2 - Bias^2} \tag{4}$$

where N is the total number of *in situ* measurements. In the Taylor diagrams, SDV and r are shown as the radial distance and an angle in the polar plot. Thus, cRMSE is the distance to the point 'Obs' on the Taylor diagrams. The shorter this distance, the better the agreement between GST$_{ESM}$ and GST$_{REF}$ (Al-Yaari et al., 2016; Albergel et al., 2013). cRMSE and ubRMSE are recommended for the evaluation of performance because the bias can be effectively eliminated (Al-Yaari et al., 2016).

The comprehensive model ranking (MR; Chen et al., 2011) measures the consistency of the simulations of each ESM, which is defined as follows:

$$MR = 1 - \frac{1}{1 \times m \times n} \sum_{i=1}^{n} rank_i \qquad (5)$$

where m and n are the numbers of ESMs and indices, and $rank_i$ is based on the ESM's order of performance on each index. The MR of the best-performing ESM is close to 1, indicating high skill (You et al., 2017).

MME19 was obtained by averaging the GST of the available ESMs. Based on the MR, the Top 5 ESMs from 19 ESMs of each site were selected, and the Top 8 ESMs from 19 ESMs of all 5 sites were selected. Then, the optimal five model ensemble (hereafter MME5), optimal eight model ensemble (hereafter MMEsup) and MME19 GST simulations were compared with the *in situ* measurements based on the error metrics of ubRMSE (Eq. 4), slope and coefficient of determination (R^2). The slope is used to evaluate the consistency between GST_{ESM} and GST_{REF} at each site (the closer to 1, the better the consistency).

2.3.3. Bias calibration

The least squares regression method is used to adjust the GST_{ESM} with the hypothesis that the systematic bias characteristics between GST_{REF} and GST_{ESM} during the training period still apply in the correction period. For the evaluation, the observation periods at the five sites were divided into two periods. Specifically, 2002−2013 for CN03 and CN04, 2007−2013 for QT04, 2010−2013 for QT08 and 2008−2013 for QT09 were selected as the training periods to establish the linear calibration models, and 2014 was chosen for all sites as the evaluation period. Moreover, ubRMSE (Eq. 4) and bias) were used to evaluate the calibration results. Afterwards,

GST_{ESM} from 1900 to 2014 was revised based on the linear calibration models.

3. Results and discussion

3.1. GST evaluation and calibration

3.1.1. GST evaluation

The performance criteria (i.e. SDV, r and cRMSE) estimated between the 19 CMIP6 GST_{ESM} and GST_{REF} at the five sites from 2010 to 2014 are presented in the Taylor diagrams shown in Fig. 2. Regarding correlation, all ESMs captured the temporal dynamics of GST well over the study region with high temporal correlation coefficients ($r > 0.9$). However, large errors with a broad range of cRMSE (1.14−4.98 °C) were detected among different ESMs and sites, indicating the inconsistent performance of different ESMs over different sites. IPSL-CM6A-LR obtained the highest cRMSE at QT08 (cRMSE = 4.98 °C). Furthermore, for a given site, most of the ESMs were frequently out of the arc of one normalized SDV, except for CN03, indicating a large variability of modelling GST at the other sites. The GSTs of IPSL-CM6A-LR, CAS-ESM2-0 and GFDL-CM4 at CN04 exhibited significant variability. Most of the GST_{ESM} were colder than the GST_{REF}, as indicated by a negative bias (GST_{ESM} minus GST_{ESM}), among which IPSL-CM6A-LR has the largest biases with an average value of −7.70 °C for all sites (Fig. A2).

The large errors of GST_{ESM} could be partly attributed to the uncertainties of the estimates simulated by the climate models (e.g. TAS) and the structural weakness in the soil hydrology of the underlying LSMs. Fig. 3 shows the bias of TAS between
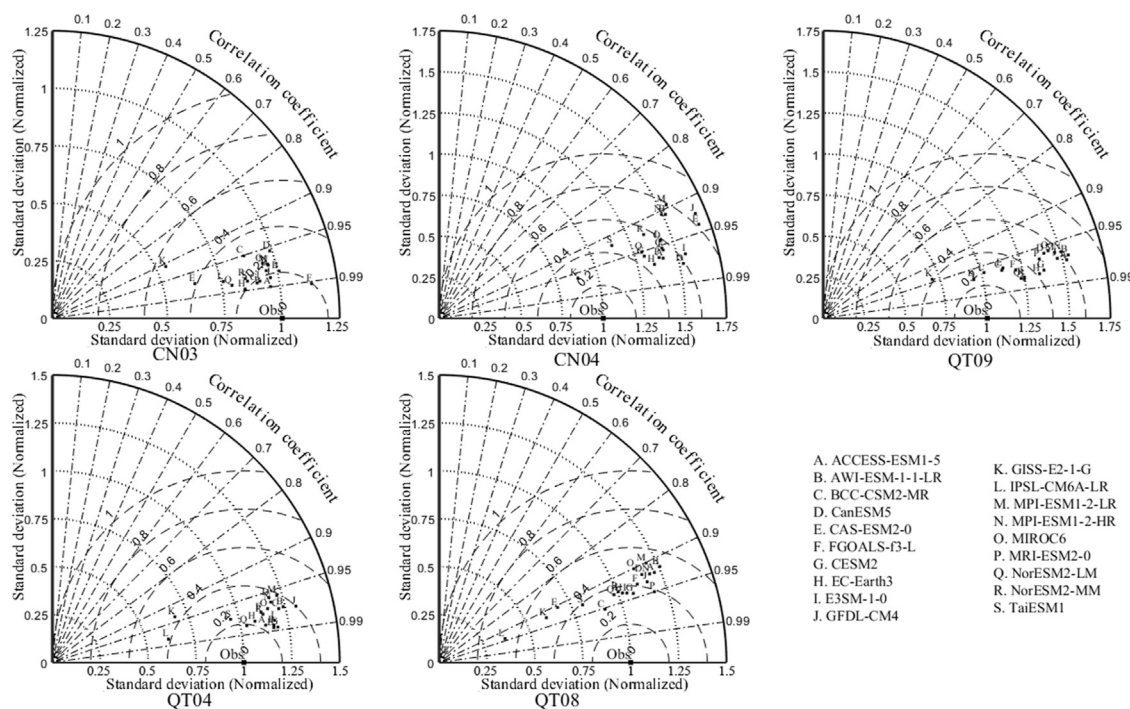

Fig. 2. Standardised Taylor diagrams for comparing the monthly ESM-based GST (GST_{ESM}) against *in situ* GST (GST_{REF}) at the five sites from 2010 to 2014 (cRMSE is the distance to the point 'Obs').
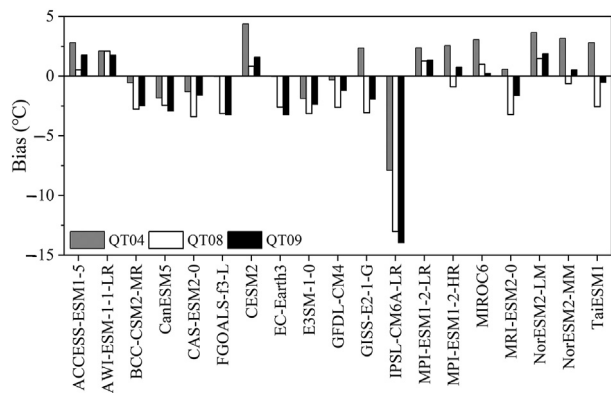
Fig. 3. Average annual TAS bias between the 19 CMIP6 ESMs and *in situ* measurements at QT04, QT08 and QT09 on the QTP from 2010 to 2014.

climate models and *in situ* measurements. Notably, the TAS derived from the climate models exhibited poor performance in terms of a broad range of biases (ESM TAS minus *in situ* TAS = −14 to 5 °C) at three sites (i.e. QT04, QT08 and QT09). Most ESMs presented cold bias, which could partly explain the underestimation of GST_{ESM} (Fig. A2). For instance, TAS and GST of IPSL-CM6A-LR were observed to have the most significant cold bias (average TAS bias = −11.62 °C and average GST bias = −8.08 °C), indicating that the TAS simulated by the climate models do influence the performance of GST simulated by subsequent LSMs. This finding is consistent with those of Hu et al. (2019b) and Cheon et al. (2013), who reported a close relationship between air and soil temperature in LSMs. Moreover, the structural weakness in the soil hydrology of different LSMs is highly related to the performance of GST_{ESM} over the permafrost regions (Slater and Lawrence, 2013; Zhou and Du, 2015). Large wet biases of the SSM derived from LSMs are observed in Fig. 4, which further illustrate their influence on the cold bias of GST_{ESM}. The more water retained in the soil, the more heat is consumed by water evaporation, leading to a decrease in GST (Li et al., 2012). All LSMs equipped in the ESMs were not initially developed for permafrost research, and many parameterisation schemes were unsuitable for
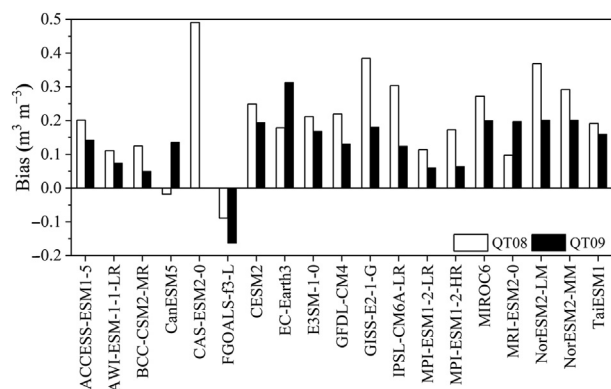
permafrost (Hu et al., 2019a; Yang et al., 2020). Previous studies have noted that soil thermal conductivity and unfrozen water content parameterisation schemes used in the LSMs were unsuitable for the permafrost regions of the QTP (Hu et al., 2017, 2020). Furthermore, the accuracy of the simulation of soil thermal properties (e.g. soil thermal conductivity) is dependent on several parameters, such as soil moisture, soil texture (e.g. FAO soil dataset) and bulk density, resulting in high complexity in representing the mechanism of the soil hydrothermal process in the LSMs (Luo et al., 2009; Wang et al., 2021). Moreover, many previous studies have attempted to enhance the site 'representation' by averaging the *in situ* measurements from multiple sites within one grid to represent the ground truth. However, this study was conducted based on individual sites because of insufficient observations over the permafrost regions of the QTP. It is difficult to tell how representative the observations are for the ESM data characterised by grid resolutions of 1.0° × 1.0°, which may introduce some uncertainties to the analysis.

The portrait diagram shows the overall ranking (i.e. MR) of all ESMs for the monthly GST at all sites (Fig. 5). Each ESM was ranked from 1 (best) to 19 (worst) for each site to illustrate their overall performance. Notably, all of the ESMs have large differences in simulating GST at different sites, and none of the best-performing ESMs is uniformly superior at all sites. ACCESS-ESM1-5, CESM2, E3SM-1-0, EC-Earth3, FGOALS-f3-L, NorESM2-LM, NorESM2-MM and TaiESM1 obtained better scores than the other ESMs. Meanwhile, AWI-ESM-1-1-LR, CanESM5, CAS-ESM2-0, GFDL-CM4, GISS-E2-1-G, IPSL-CM6A-LR, MPI-ESM1-2-LR and MPI-ESM1-2-HR exhibited poor performances. Because of the GST_{ESM} performance discrepancy at these sites, the Top 5 ESMs with the sequence number in the portrait diagram at each site were defined as optimal ESMs.

The performance of two ensemble simulations was evaluated based on three statistical metrics (i.e. slope, R^2 and ubRMSE) against GST_{ESM} at the five sites over the permafrost regions from 2010 to 2014 (Fig. 6). One ensemble simulation only selected the Top 5 ESMs (MME5), whereas the other ensemble simulation used all 19 ESMs (MME19). As expected, the two multi-model ensemble simulations had better R^2 and ubRMSE than most of the individual ESMs, with R^2 higher than 0.91 and ubRMSE lower than 2.30 °C. Moreover, MME5 (slope ranged from 0.91 to 1.13) performed better than MME19 (slope ranged from 0.87 to 1.33) at all sites because the slope of MME5 is closer to 1. Notably, MME5 is more consistent with GST_{REF} than MME19. Thus, MME5 of GST_{ESM} was obtained.

3.1.2. GST calibration

Based on the results presented in Section 3.1.1, we calibrated the monthly GST_{ESM} obtained from the MME5 using the least squares regression method. Table 3 shows the calibration models and their correlation coefficients. GST_{ESM} obtained from the MME5 at all sites was significantly positively correlated with GST_{REF} ($R^2 > 0.88$) during the

Fig. 4. Average SSM bias between the 19 CMIP6 ESMs and *in situ* measurements at QT08 and QT09 of QTP from July to August in 2013−2014.
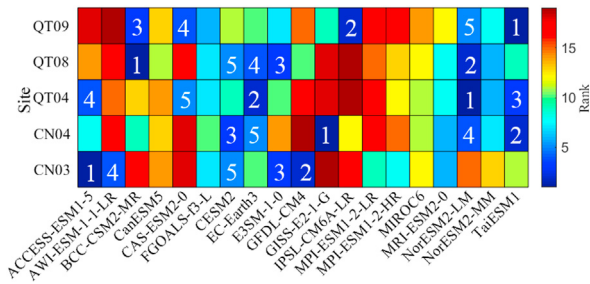
Fig. 5. Portrait diagrams of the model ranking based on the composite rating indicator for GST_{ESM} of all sites from 2010 to 2014(The colours of the symbols represent the model's rank to characterise the synthetic performance of GST_{ESM}. Only the Top 5 ESMs for each site were labelled with the sequence number).

calibration periods, which indicated that the calibration models performed well.

The error metrics between the original and calibrated GST and GST_{REF} in 2014 are also shown in Table 3. The calibrated GST exhibited better performance with lower bias (median bias = 0.41 °C) and ubRMSE (median ubRMSE = 0.95 °C) than the uncalibrated GST (median bias = −0.60 °C and median ubRMSE = 1.21 °C). After calibration, the absolute median bias and median ubRMSE values were reduced by 0.19 and 0.17 °C, respectively. Furthermore, the linear calibration models improved the performance of the original GST at the five sites because the large errors (i.e. bias and ubRMSE) of the original GST obtained from MME5 have been corrected (Fig. 7). Thus, the optimal multi-model

ensemble and calibration models are recommended before applying the CMIP6 GST_{ESM} for permafrost modelling.

To assess the validity of the GST calibration models shown in Table 3, we calibrated the GSTs of the three other sites (i.e. Tuotuohe (alpine steppe), Qumalai (alpine meadow) and Maduo (alpine wet meadow)) in the surrounding area using the same calibration methods by their vegetation types (i.e. CN03 (alpine steppe), QT04 (alpine meadow) and CN04 (alpine wet meadow)), with the hypothesis that the grids with similar vegetation types were subjected to similar bias in GST_{ESM}. The MME5 GST values for the three sites were first calculated using the Top 5 ESMs confirmed in Fig. 3 and then calibrated using the GST calibration models with the corresponding vegetation types. The results of MME5 and the calibrated GST were evaluated against the *in situ* measurements at the three sites.

Fig. 8 shows the error metrics of MME5 and the calibrated GST against the *in situ* measurements at the three sites. Notably, the calibrated GST (bias ranges from 0.67 to 2.17 °C and RMSE ranges from 2.99 to 3.70 °C) performed better than MME5 (bias ranges from 3.5 to 4.98 °C and RMSE ranges from 4.21 to 5.43 °C) with lower bias and RMSE, indicating that our GST calibration models are effective. Moreover, large errors and variabilities of the original GST from the 19 ESMs (Fig. A3; bias ranges from 2 to 13 °C and RMSE ranges from 2 to 14 °C), are detected, indicating that the reconstructed long-term GST obtained from MME5 and the calibration models could be used for further applications. Nevertheless, some uncertainties may exist in the GST calibrations. For example, the GST calibration models for QT08 and QT09
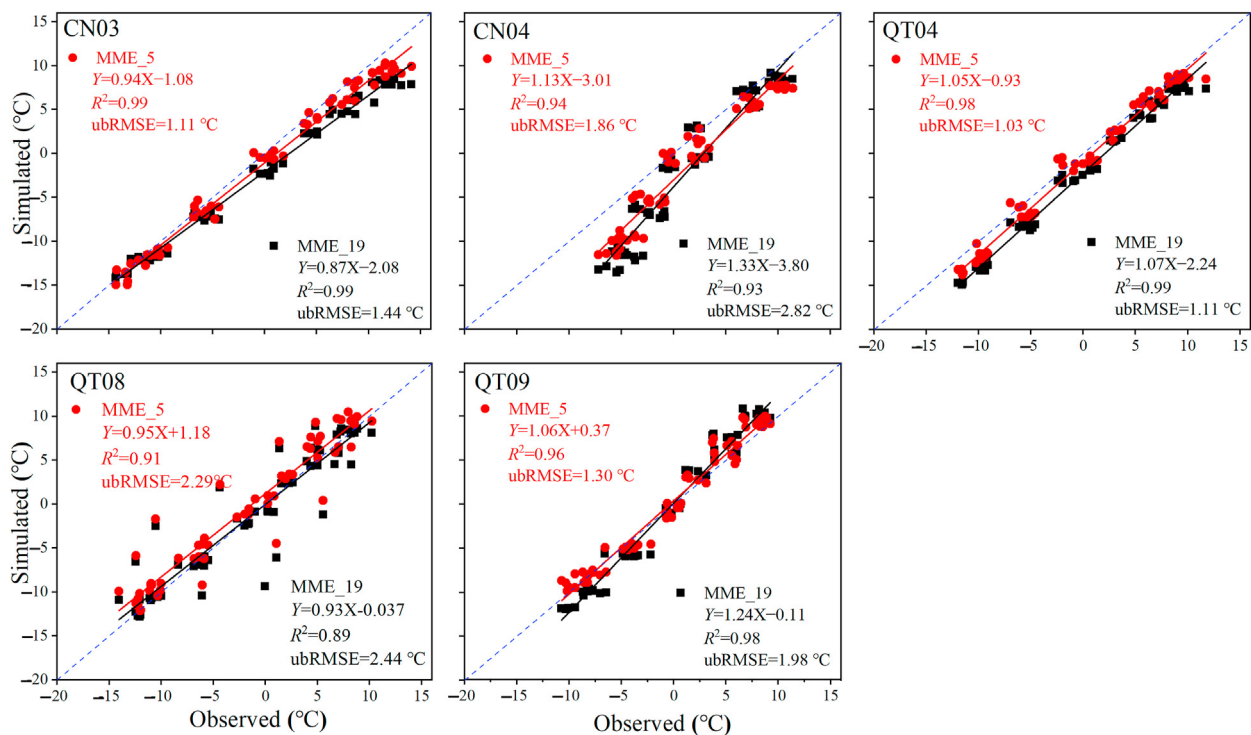

Fig. 6. Comparisons between monthly GST simulated by the averaged multi-model ensemble (MME19) and optimised multi-model ensemble (MME5) against GST_{REF} from 2010 to 2014.

Table 3
Error metrics of the original and calibrated GST against observations.

Site	Calibration models, R^2	Bias (°C)		ubRMSE (°C)	
		Original	Calibrated	Original	Calibrated
CN03	$y = 1.07x + 1.24$, $R^2 = 0.99$	−0.63	0.55	0.81	0.88
CN04	$y = 0.84x + 2.48$, $R^2 = 0.94$	−2.60	0.00	2.01	1.48
QT04	$y = 0.92x + 0.89$, $R^2 = 0.97$	−0.60	0.41	1.12	0.95
QT08	$y = 0.96x − 1.24$, $R^2 = 0.89$	1.55	0.33	0.61	0.56
QT09	$y = 0.94x − 0.23$, $R^2 = 0.95$	1.00	0.78	1.50	1.22
Median	−	−0.6	0.41	1.12	0.95

Note: *All regressions are significant ($p < 0.01$). The periods 2002−2013 for CN03 and CN04, 2007−2013 for QT04, 2010−2013 for QT08 and 2008−2013 for QT09 were used to establish the linear calibration models, and 2014 was used for all sites to calculate bias and ubRMSE.

were established with relatively short training periods, which may ignore the signals of some strong climate changes during the training period.

3.2. Annual and seasonal GST variations

Based on the GST calibration models established in Section 3.1.2, the monthly MME5 GST_{REF} from 1900 to 2014 was

revised. Similar warming patterns were observed in the annual and seasonal calibrated GSTs at all sites, although some differences in peaks and troughs related to variability are noted (Fig. 9a−e). The warming rates of the calibrated annual and seasonal GSTs at the five sites from 1900 to 2014 are presented in Fig. 9f. The average warming rate of the annual GST at all sites was 0.04 °C per decade. The highest warming rate was detected at CN03 and CN04 (0.06 °C per decade), whereas the lowest warming rate was detected at QT09 (0.01 °C per decade). Regarding seasonal variations, the calibrated GST warmed faster in winter, slower in spring and autumn, and slowest in summer at all sites, except for CN04 and QT04, which had higher warming trends. This finding is consistent with those of Wu et al. (2013), Guo (2017), Lu et al. (2017), Zhu et al. (2017) and Liu and Chen (2000), who reported major changes in winter for shallow soil and air temperatures.

Furthermore, we compared the warming rates with that derived from the optimal five ESMs for each site from 1900 to 2014 (Fig. 10) and the GLDAS-Noah GST product from 1961 to 2014 (Fig. A4). Notably, the average warming rates of the calibrated GST were lower than those of most original optimal five ESMs (Fig. 10), indicating that the original optimal five ESMs overestimated soil temperature in the present-day permafrost region as previous studies (Burke et al., 2020; Guo et al., 2012). Meanwhile, similar warming patterns and
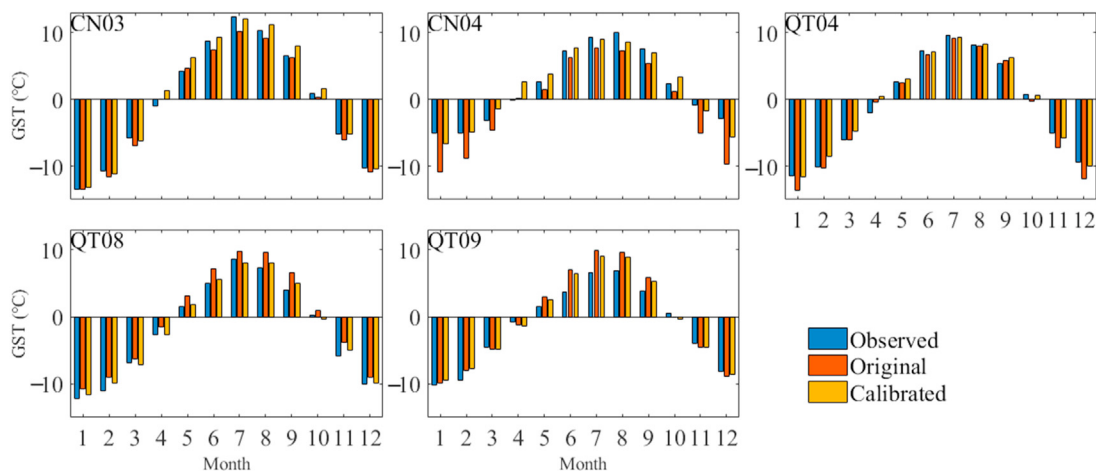

Fig. 7. Monthly GST of *in situ* measurements, original GST obtained from MME5 and calibrated GST at the five sites in 2014.
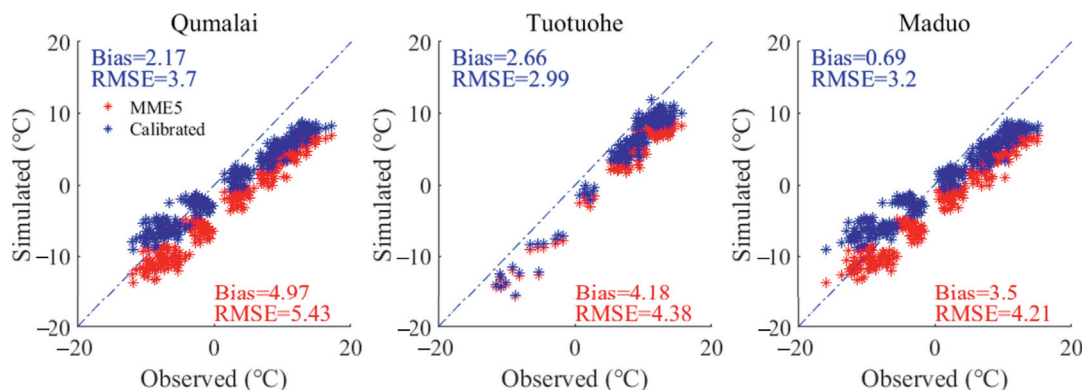

Fig. 8. Comparisons between the monthly GST simulated by the optimised multi-model ensemble (MME5) and the calibrated GST at the three sites.
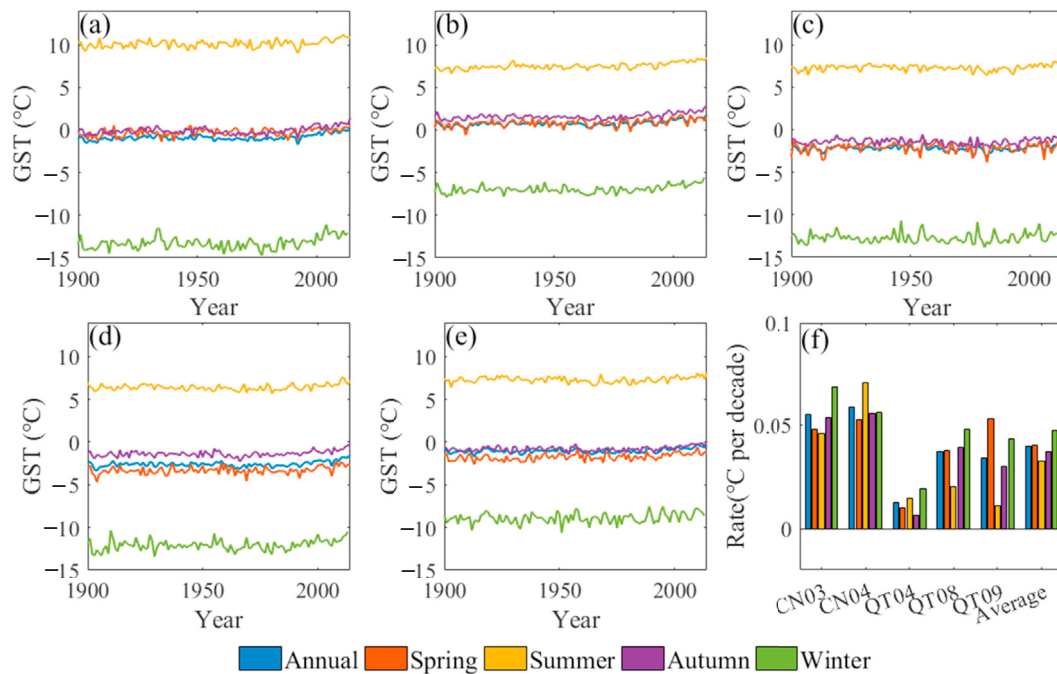
Fig. 9. Time series of the annual and seasonal calibrated GSTs at the five sites (a−e) from 1900 to 2014, and (f) warming rates (All are statistically significant at the 0.05 level).

warming rates were observed in both GLDAS-Noah and revised GST at the five sites from 1961 to 2014, which reconfirmed the sound performance of the established long-term GST datasets (Fig. A4). Nevertheless, our results showed a lower warming rate than the previous studies. For instance, the observed soil temperature warming rate from meteorological sites on the QTP was 0.58 °C per decade at a depth of 5 cm from 1983 to 2013 (Zhu et al., 2017) partly

because most of the meteorological sites are distributed on the seasonally frozen ground. Because the release and absorption of heat during the freeze−thaw cycle of the active layer reduces the annual soil temperature dynamic, the warming rate of permafrost is lower than that of seasonally frozen ground (Zhao et al., 2004).

Vegetation significantly impacted the GST over the permafrost regions (Lu et al., 2006). Our results showed that
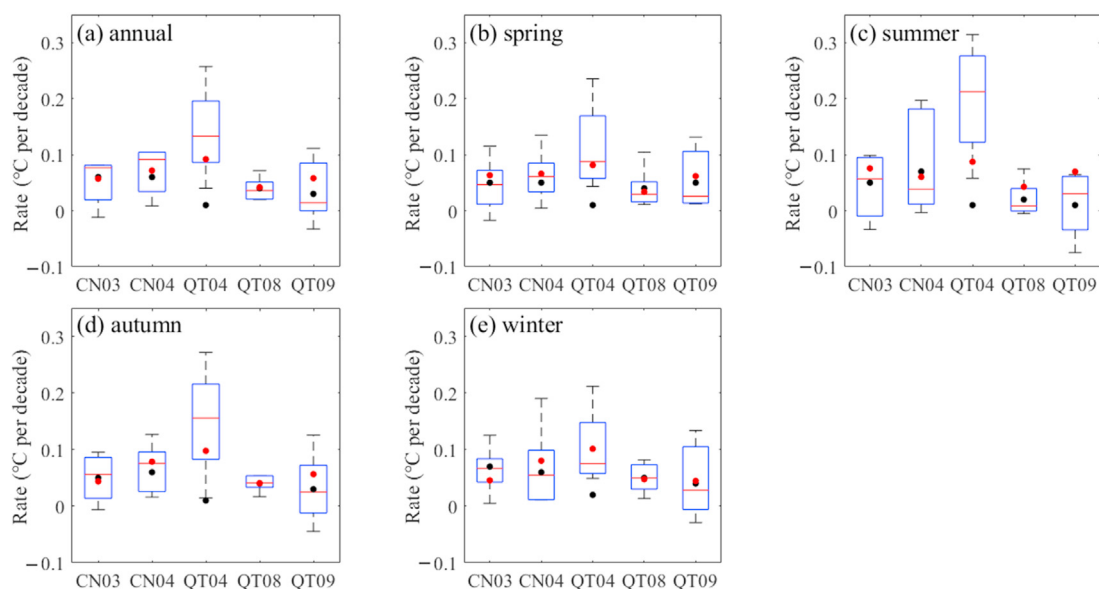

Fig. 10. Box plots of the GST warming rate of the (a) annual and (b−e) seasonal GSTs at the five sites from the original optimal five ESMs from 1900 to 2014 (The median warming rates of the original optimal five ESMs are denoted by the red line, and the upper and lower boundaries indicate the maximum and minimum values, respectively. The average warming rate of the original optimal five ESMs and the calibrated GST are denoted by the red and black dots, respectively. All linear trends are statistically significant ($p < 0.05$)).

the annual warming rate was higher at CN03 and CN04 (0.06 °C per decade) with alpine steppe and alpine wet meadow than at the three other sites from 1900 to 2014 and lowest at QT04 (0.01 °C per decade) with alpine meadow. The *in situ* measurement of CN04 with alpine wet meadows did not exhibit a low warming rate, which is inconsistent with that reported by Hu et al. (2019a). This finding could be attributed to the coarse spatial resolution (1.0° × 1.0°) of the CMIP6 ESMs and the particular position of CN04, located in the lower boundary of the permafrost regions, which increases the warming rate of CN04 by introducing the influence of the surrounding seasonally frozen ground. Thus, model simulations with enhanced spatial resolutions are urgently needed to obtain accurately modelled GST.

4. Conclusions

The study assesses a suite of 19 CMIP6 ESMs in GST simulation and calibrates and reconstructs long-term GST time series. The main conclusions are as follows: 1) The GST of the 19 CMIP6 ESMs at the five sites exhibited large cold bias (cRMSE ranged from 1.14 to 4.98 °C). This finding could be partly attributed to the cold bias of TAS derived from surface climate and the wet bias of SSM simulated by the underlying LSMs. ACCESS-ESM1-5, CESM2, E3SM-1-0, EC-Earth3, FGOALS-f3-L, NorESM2-LM, NorESM2-MM and TaiESM1 performed better than the other ESMs for all sites. Meanwhile, AWI-ESM-1-1-LR, GFDL-CM4, GISS-E2-1-G, IPSL-CM6A-LR, MPI-ESM1-2-LR and MPI-ESM1-2-HR exhibited poor performance. 2) Compared with the monthly GST from the MME19, higher correlation coefficients ($r > 0.91$), lower ubRMSE (ubRMSE < 2.30 °C) and better slope (closer to 1) than the monthly GST from the MME5 and *in situ* measurements at all sites were obtained. Compared with the original GST obtained from MME5, the accuracy of the calibrated GST was effectively improved with a lower median bias (bias = 0.41 °C) and ubRMSE (ubRMSE = 0.95 °C). 3) Both annual and seasonal GSTs exhibited warming trends. The warming rate was highest in winter and spring, followed by autumn, and lowest in summer from 1900 to 2014. The highest warming rate was detected at CN03 with the alpine steppe, and the lowest warming rate was detected at QT04 with the alpine meadow.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.accre.2023.01.007.

References

Al-Yaari, A., Wigneron, J.P., Kerr, Y., et al., 2016. Testing regression equations to derive long-term global soil moisture datasets from passive microwave observations. Rem. Sens. Environ. 180 (7), 453−464. https://doi.org/10.1016/j.rse.2015.11.022.

Albergel, C., Dorigo, W., Reichle, R.H., et al., 2013. Skill and global trend analysis of soil moisture from reanalyses and microwave remote sensing. J. Hydrometeorol. 14 (4), 1259−1277. https://doi.org/10.1175/jhm-d-12-0161.1.

Anisimov, O.A., Nelson, F.E., 1996. Permafrost distribution in the Northern Hemisphere under scenarios of climatic change. Global Planet. Change 14 (1), 59−72. https://doi.org/10.1016/0921-8181(96)00002-1.

Boucher, O., Servonnat, J., Albright, A.L., et al., 2020. Presentation and evaluation of the IPSL−CM6A−LR climate model. J. Adv. Model. Earth Syst. 12 (7), 1−52. https://doi.org/10.1029/2019ms002010.

Burke, E.J., Zhang, Y., Krinner, G., 2020. Evaluating permafrost physics in the coupled model intercomparison project 6 (CMIP6) models and their sensitivity to climate change. Cryosphere 14 (9), 3155−3174. https://doi.org/10.5194/tc-14-3155-2020.

Cao, B., Gruber, S., Zheng, D., et al., 2020. The ERA5−Land soil temperature bias in permafrost regions. Cryosphere 14 (8), 2581−2595. https://doi.org/10.5194/tc-14-2581-2020.

Chen, W., Jiang, Z., Li, L., 2011. Probabilistic projections of climate change over China under the SRES A1B scenario using 28 AOGCMs. J. Clim. 24 (17), 4741−4756. https://doi.org/10.1175/2011jcli4102.1.

Cheon, J.-Y., Ham, B.-S., Lee, J.-Y., et al., 2013. Soil temperatures in four metropolitan cities of Korea from 1960 to 2010: implications for climate change and urban heat. Environ. Earth Sci. 71 (12), 5215−5230. https://doi.org/10.1007/s12665-013-2924-8.

Cui, Y., Qin, J., Jing, W., et al., 2018. Applicability evaluation of merged soil moisture in GLDAS and CLDAS products over Qinghai−Tibetan Plateau. Plateau Meteorol. 37 (1), 123−136. https://doi.org/10.7522/j.issn.1000-0534.2017.00035.

Danabasoglu, G., Lamarque, J.F., Bacmeister, J., et al., 2020. The community Earth system model version 2 (CESM2). J. Adv. Model. Earth Syst. 12 (2), 1−35. https://doi.org/10.1029/2019ms001916.

Danek, C., Shi, X., Stepanek, C., et al., 2020. AWI AWI−ESM1.1LR model output prepared for CMIP6 CMIP historical. Earth Syst. Grid Feder. https://doi.org/10.22033/ESGF/CMIP6.9328.

Ding, X., Lai, X., Fan, G., et al., 2018. Analysis on the applicability of reanalysis soil temperature and moisture datasets over Qinghai−Tibetan Plateau. Plateau Meteorol. 37 (3), 626−641. https://doi.org/10.7522/j.issn. 1000−0534.2017.00060.

Doscher, R., Acosta, M., Alessandri, A., et al., 2020. The EC-Earth3 Earth system model for the climate model intercomparison project 6. Geosci. Model Dev. 15 (7), 2973−3020. https://doi.org/10.5194/gmd-2020-446.

Eyring, V., Bony, S., Meehl, G.A., et al., 2016. Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. Geosci. Model Dev. 9 (5), 1937−1958. https://doi.org/10.5194/gmd-9-1937-2016.

Fang, X., Luo, S., Lyu, S., 2018. Observed soil temperature trends associated with climate change in the Tibetan Plateau, 1960−2014. Theor. Appl. Climatol. 135 (1−2), 169−181. https://doi.org/10.1007/s00704-017-2337-9.

Gao, J., Du, J., Yang, C., Deqing, Z., et al., 2022. Evaluation and correction of climate simulations for the Tibetan Plateau using the CMIP6 models. Rem. Sens. 13 (12), 1947. https://doi.org/10.3390/atmos13121947.

Golaz, J.C., Caldwell, P.M., Van Roekel, L.P., et al., 2019. The DOE E3SM coupled model version 1: overview and evaluation at standard resolution. J. Adv. Model. Earth Syst. 11 (7), 2089−2129. https://doi.org/10.1029/2018ms001603.

Guo, D., 2017. Simulated change in soil temperature on the Tibetan Plateau from 1901 to 2010. Quat. Sci. 37 (5), 1101−1109.

Guo, D., Li, D., Hua, W., 2018. Quantifying air temperature evolution in the permafrost region from 1901 to 2014. Int. J. Climatol. 38 (1), 66−76. https://doi.org/10.1002/joc.5161.

Guo, D., Wang, H., 2016. CMIP5 permafrost degradation projection: a comparison among different regions. J. Geophys. Res. Atmos. 121 (9), 4499−4517. https://doi.org/10.1002/2015jd024108.

Guo, D., Wang, H., Li, D., 2012. A projection of permafrost degradation on the Tibetan Plateau during the 21st century. J. Geophys. Res. Atmos. 117 (D5), 1−15. https://doi.org/10.1029/2011jd016545.

Guo, D., Yang, M., Wang, H., 2011. Characteristics of land surface heat and water exchange under different soil freeze/thaw conditions over the central Tibetan Plateau. Hydrol. Process 25 (16), 2531−2541. https://doi.org/10.1002/hyp.8025.

Gusain, A., Ghosh, S., Karmakar, S., 2020. Added value of CMIP6 over CMIP5 models in simulating Indian summer monsoon rainfall. Atmos. Res. 232 (9), 1−15. https://doi.org/10.1016/j.atmosres.2019.104680.

He, B., Bao, Q., Wang, X., et al., 2019. CAS FGOALS-f3-L model datasets for CMIP6 historical atmospheric model intercomparison project simulation. Adv. Atmos. Sci. 36 (8), 771−778. https://doi.org/10.1007/s00376-019-9027-8.

Held, I.M., Guo, H., Adcroft, A., et al., 2019. Structure and performance of GFDL's CM4.0 climate model. J. Adv. Model. Earth Syst. 11 (11), 3691−3727. https://doi.org/10.1029/2019MS001829.

Hu, G., Zhao, L., Li, R., et al., 2019a. Variations in soil temperature from 1980 to 2015 in permafrost regions on the Qinghai−Tibetan Plateau based on observed and reanalysis products. Geoderma 337 (3), 893−905. https://doi.org/10.1016/j.geoderma.2018.10.044.

Hu, G., Zhao, L., Li, R., et al., 2019b. Simulation of land surface heat fluxes in permafrost regions on the Qinghai−Tibetan Plateau using CMIP5 models. Atmos. Res. 220 (1), 155−168. https://doi.org/10.1016/j.atmosres.2019.01.006.

Hu, G., Zhao, L., Wu, X., et al., 2017. Comparison of the thermal conductivity parameterizations for a freeze−thaw algorithm with a multi-layered soil in permafrost regions. Catena 156, 244−251. https://doi.org/10.1016/j.catena.2017.04.011.

Hu, G., Zhao, L., Zhu, X., et al., 2020. Review of algorithms and parameterizations to determine unfrozen water content in frozen soil. Geoderma 368 (2), 1−13. https://doi.org/10.1016/j.geoderma.2020.114277.

Huang, X., Han, S., Shi, C., 2022. Evaluation of three air temperature reanalysis datasets in the alpine region of the Qinghai−Tibet Plateau. Rem. Sens. 14 (18), 4447. https://doi.org/10.3390/rs14184447.

Koven, C.D., Riley, W.J., Stern, A., et al., 2013. Analysis of permafrost thermal dynamics and response to climate change in the CMIP5 earth system models. J. Climate 26 (6), 1877−1900. https://doi.org/10.1175/jcli-d-12-00228.1.

Lawrence, D.M., Slater, A.G., Swenson, S.C., et al., 2012. Simulation of present-day and future permafrost and seasonally frozen ground conditions in CCSM4. J. Climate 25 (7), 2207−2225. https://doi.org/10.1175/jcli-d-11-00334.1.

Lee, W.L., Liang, H.C., 2020. AS-RCEC TaiESM1.0 model output prepared for CMIP6 CMIP piControl. Earth Syst. Grid Feder. https://doi.org/10.22033/ESGF/CMIP6.9798.

Li, C., Lu, H., Yang, K., et al., 2017. Evaluation of the Common Land Model (CoLM) from the perspective of water and energy budget simulation: towards inclusion in CMIP6. Atmosphere 8 (8), 1−33. https://doi.org/10.3390/atmos8080141.

Li, R., Zhao, L., Ding, Y., et al., 2012. Temporal and spatial variations of the active layer along the Qinghai−Tibet highway in a permafrost region. Chin. Sci. Bull. 57 (35), 4609−4616. https://doi.org/10.1007/s11434-012-5323-8.

Li, W., Zhang, Y., Shi, X., et al., 2019. Development of land surface model BCC_AVIM2.0 and its preliminary performance in LS3MIP/CMIP6. J.

Meteorol. Res. 33 (5), 851−869. https://doi.org/10.1007/s13351-019-9016-y.

Li, Y., Yan, D., Peng, H., et al., 2020. Evaluation of precipitation in CMIP6 over the Yangtze River Basin. Atmos. Res. 253 (3), 1−12. https://doi.org/10.1016/j.atmosres.2020.105406.

Liu, C., Yu, Y., Xie, J., et al., 2015. Applicability of soil temperature and moisture in several datasets over Qinghai−Xizang Plateau. Plateau Meteorol. 34 (3), 653−665.

Liu, X., Chen, B., 2000. Climatic warming in the Tibetan Plateau during recent decades. Int. J. Climatol. 20 (14), 1729−1742. https://doi.org/10.1002/1097-0088(20001130)20:14<1729::AID-JOC556>3.0.CO;2-Y.

Lu, Z., Wu, Q., Sheng, Y., et al., 2006. Heat and water difference of active layers beneath different surface conditions near Beiluhe in Qinghai−Xizang Plateau. J. Glaciol. Geocryol. 28 (5), 642−647.

Lu, Q., Zhao, D., Wu, S., 2017. Simulated responses of permafrost distribution to climate change on the Qinghai−Tibet Plateau. Sci. Rep. 7 (1), 3845. https://doi.org/10.1038/s41598-017-04140-7.

Lun, Y., Liu, L., Cheng, L., et al., 2021. Assessment of GCMs simulation performance for precipitation and temperature from CMIP5 to CMIP6 over the Tibetan Plateau. Int. J. Climatol. 41 (7), 3994−4018. https://doi.org/10.1002/joc.7055.

Luo, S., Lyu, S., Zhang, Y., et al., 2009. Soil thermal conductivity parameterization establishment and application in numerical model of central Tibetan Plateau. Chin. J. Geophys. 52 (1), 919−928. https://doi.org/10.3969/j.issn.0001-5733.2009.04.008.

Ma, J., Li, R., Huang, Z., et al., 2022. Evaluation and spatio−temporal analysis of surface energy flux in permafrost regions over the Qinghai−Tibet Plateau and Arctic using CMIP6 models. Int. J. Digital Earth 15 (1), 1948−1966. https://doi.org/10.1080/17538947.2022.2142307.

Mauritsen, T., Bader, J., Becker, T., et al., 2019. Developments in the MPI−M Earth system model version 1.2 (MPI−ESM1.2) and its response to increasing CO_2. J. Adv. Model. Earth Syst. 11 (4), 998−1038. https://doi.org/10.1029/2018MS001400.

Meehl, G.A., Boer, G.J., Covey, C., et al., 1997. Intercomparison makes for a better climate model. Eos, Trans. Am. Geophys. Union 78 (41), 445−451. https://doi.org/10.1029/97EO00276.

Miller, R.L., Schmidt, G.A., Nazarenko, L.S., et al., 2021. CMIP6 historical simulations (1850−2014) with GISS−E2.1. J. Adv. Model. Earth Syst. 13 (1), 1−35. https://doi.org/10.1029/2019ms002034.

Mu, C., Zhang, T., Zhao, Q., et al., 2017. Permafrost affects carbon exchange and its response to experimental warming on the northern Qinghai−Tibetan Plateau. Agric. For. Meteorol. 247 (15), 252−259. https://doi.org/10.1016/j.agrformet.2017.08.009.

Pascoe, C., Lawrence, B.N., Guilyardi, E., et al., 2019. Designing and documenting experiments in CMIP6. Geosci. Model Dev. 1−27. https://doi.org/10.5194/gmd-2019-98.

Qin, Y., Liu, W., Guo, Z., et al., 2020. Spatial and temporal variations in soil temperatures over the Qinghai−Tibet Plateau from 1980 to 2017 based on reanalysis products. Theor. Appl. Climatol. 140 (3−4), 1055−1069. https://doi.org/10.1007/s00704-020-03149-9.

Qin, Y., Wu, T., Wu, X., et al., 2017. Assessment of reanalysis soil moisture products in the permafrost regions of the central of the Qinghai−Tibet Plateau. Hydrol. Process 31 (26), 4647−4659. https://doi.org/10.1002/hyp.11383.

Shang, W., Duan, K., Li, S., et al., 2021. Simulation of the dipole pattern of summer precipitation over the Tibetan Plateau by CMIP6 models. Environ. Res. Lett. 16 (1), 1−10. https://doi.org/10.1088/1748-9326/abd0ac.

Slater, A.G., Lawrence, D.M., 2013. Diagnosing present and future permafrost from climate models. J. Clim. 26 (15), 5608−5623. https://doi.org/10.1175/jcli-d-12-00341.1.

Stouffer, R.J., Eyring, V., Meehl, G.A., et al., 2017. CMIP5 Scientific gaps and recommendations for CMIP6. Bull. Am. Meteorol. Soc. 98 (1), 95−105. https://doi.org/10.1175/bams-d-15-00013.1.

Sun, Z., Zhao, L., Hu, G., et al., 2019. Modeling permafrost changes on the Qinghai−Tibetan Plateau from 1966 to 2100: a case study from two boreholes along the Qinghai−Tibet engineering corridor. Permafr. Periglac. Process 31 (1), 156−171. https://doi.org/10.1002/ppp.2022.

Swart, N.C., Cole, J.N.S., Kharin, V.V., et al., 2019. The Canadian Earth system model version 5 (CanESM5.0.3). Geosci. Model Dev. 12 (11), 4823–4873. https://doi.org/10.5194/gmd-12-4823-2019.

Tatebe, H., Ogura, T., Nitta, T., et al., 2019. Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. Geosci. Model Dev. 12 (7), 2727–2765. https://doi.org/10.5194/gmd-12-2727-2019.

Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. Atmos. 106 (D7), 7183–7192. https://doi.org/10.1029/2000jd900719.

Taylor, K.E., Stouffer, R.J., Meehl, G.A., 2012. An overview of CMIP5 and the experiment design. Bull. Am. Meteorol. Soc. 93 (4), 485–498. https://doi.org/10.1175/bams-d-11-00094.1.

Taylor, K.E., Stouffer, R.J., Mitchell, J.F.B., et al., 2007. The WCRP CMIP3 multimodel dataset: a new era in climate change research. Bull. Am. Meteorol. Soc. 88 (9), 1383–1394. https://doi.org/10.1175/bams-88-9-1383.

Tjiputra, J.F., Schwinger, J., Bentsen, M., et al., 2020. Ocean biogeochemistry in the Norwegian Earth system model version 2 (NorESM2). Geosci. Model Dev. 13 (5), 2393–2431. https://doi.org/10.5194/gmd-13-2393-2020.

Walvoord, M.A., Kurylyk, B.L., 2016. Hydrologic impacts of thawing permafrost: a review. Vadose Zone J. 15 (6), 1–20. https://doi.org/10.2136/vzj2016.01.0010.

Wang, C., Zhao, L., Fang, H., et al., 2021. Mapping surficial soil particle size fractions in alpine permafrost regions of the Qinghai–Tibet Plateau. Rem. Sens. 13 (7), 1392. https://doi.org/10.3390/rs13071392.

Wu, Q., Hou, Y., Yun, H., et al., 2015. Changes in active-layer thickness and near-surface permafrost between 2002 and 2012 in alpine ecosystems, Qinghai–Xizang (Tibet) Plateau, China. Global Planet. Change 124 (1), 149–155. https://doi.org/10.1016/j.gloplacha.2014.09.002.

Wu, T., Zhao, L., Li, R., et al., 2013. Recent ground surface warming and its effects on permafrost on the central Qinghai–Tibet Plateau. Int. J. Climatol. 33 (4), 920–930. https://doi.org/10.1002/joc.3479.

Wu, Y., Liu, Y., Li, J., et al., 2020. Analysis of surface temperature bias over the Tibetan Plateau in the CAS FGOALS–f3–L model. Atmos. Oceanic Sci. Lett. 14 (1), 1–7. https://doi.org/10.1016/j.aosl.2020.100012.

Yang, S., Li, R., Wu, T., et al., 2020. Evaluation of reanalysis soil temperature and soil moisture products in permafrost regions on the Qinghai–Tibetan Plateau. Geoderma 377 (1), 114583. https://doi.org/10.1016/j.geoderma.2020.114583.

You, Q., Jiang, Z., Wang, D., et al., 2017. Simulation of temperature extremes in the Tibetan Plateau from CMIP5 models and comparison with gridded observations. Clim. Dynam. 51 (1–2), 355–369. https://doi.org/10.1007/s00382-017-3928-y.

Yukimoto, S., Koshiro, T., Kawai, H., et al., 2019. MRI MRI–ESM2.0 model output prepared for CMIP6 CMIP. Earth Syst. Grid Feder. https://doi.org/10.22033/ESGF/CMIP6.621.

Zhang, G., Nan, Z., Hu, N., et al., 2022. Qinghai–Tibet Plateau permafrost at risk in the late 21st century. Earth's Future 10 (6), 1–22. https://doi.org/10.1029/2022EF002652.

Zhang, H., Zhang, M., Jin, J., et al., 2020. Description and climate simulation performance of CAS–ESM version 2. J. Adv. Model. Earth Syst. 12 (12), 1–35. https://doi.org/10.1029/2020ms002210.

Zhang, W., Li, S., Pang, Q.-q., 2008. Variation characteristics of soil temperature over Qinghai–Xizang Plateau in the past 45 years. Acta Geograph. Sin. 63 (11), 1151–1159.

Zhao, L., Li, R., Ding, Y., et al., 2011. Soil thermal regime in Qinghai–Tibet Plateau and its adjacent regions during 1977–2006. Adv. Clim. Change Res. 7 (5), 307–316.

Zhao, L., Ping, C., Yang, D., et al., 2004. Changes of climate and seasonally frozen ground over the past 30 years in Qinghai–Xizang (Tibetan) Plateau, China. Global Planet. Change 43 (1–2), 19–31. https://doi.org/10.1016/j.gloplacha.2004.02.003.

Zhao, L., Zou, D., Hu, G., et al., 2021. A synthesis dataset of permafrost thermal state for the Qinghai–Xizang (Tibet) Plateau, China. Earth system science data. Data 13 (8), 4207–4218. https://doi.org/10.5194/essd-13-4207-2021.

Zhou, L.T., Du, Z., 2015. Regional differences in the surface energy budget over China: an evaluation of a selection of CMIP5 models. Theor. Appl. Climatol. 124 (1–2), 241–266. https://doi.org/10.1007/s00704-015-1407-0.

Zhu, F., Cuo, L., Zhang, Y., et al., 2017. Spatiotemporal variations of annual shallow soil temperature on the Tibetan Plateau during 1983–2013. Clim. Dynam. 51 (5–6), 2209–2227. https://doi.org/10.1007/s00382-017-4008-z.

Zhu, Y.Y., Yang, S., 2020. Evaluation of CMIP6 for historical temperature and precipitation over the Tibetan Plateau and its comparison with CMIP5. Adv. Clim. Change Res. 11 (3), 239–251. https://doi.org/10.1016/j.accre.2020.08.001.

Ziehn, T., Chamberlain, M.A., Law, R.M., et al., 2020. The Australian Earth system model: ACCESS–ESM1.5. J. South. Hemis. Earth Syst. Sci. 70 (1), 193–214. https://doi.org/10.1071/es19035.

Zou, D., Zhao, L., Sheng, Y., et al., 2017. A new map of permafrost distribution on the Tibetan Plateau. Cryosphere 11 (6), 2527–2542. https://doi.org/10.5194/tc-11-2527-2017.