Heliyon 7 (2021) e08263

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Evaluation of CMIP6 GCMs for simulations of temperature over Thailand and nearby areas in the early 21st century

Suchada Kamworapan [a], Pham Thi Bich Thao [b,c], Shabbir H. Gheewala [b,c], Sittichai Pimonsree [d], Kritana Prueksakorn [a,e,*]

[a] Andaman Environment and Natural Disaster Research Center (ANED), Faculty of Technology and Environment (FTE), Prince of Songkla University, Phuket Campus, Phuket, 83120, Thailand
[b] Joint Graduate School of Energy and Environment (JGSEE), King Mongkut's University of Technology, Thonburi, Bangkok 10140, Thailand
[c] Center of Excellence on Energy Technology and Environment (CEE), PERDO, Ministry of Higher Education, Science, Research and Innovation, Bangkok, Thailand
[d] Atmospheric Pollution and Climate Change Research Unit, School of Energy and Environment, University of Phayao, Phayao, 56000, Thailand
[e] Faculty of Environment and Resource Studies, Mahidol University, Nakhon Phathom, 73170, Thailand

ARTICLE INFO

Keywords:
Global climate models
CMIP6
Temperature
Thailand

ABSTRACT

This study evaluates the performance of 13 global climate models (GCMs) from the Coupled Model Intercomparison Project Phase 6 (CMIP6) for simulating the temperature over Thailand during 2000–2014, for land-only, sea-only, and both land and sea. Both observation and reanalysis datasets are employed to compare with the GCMs, evaluated by five performance metrics including mean annual temperature, mean bias errors, mean seasonal cycle amplitude, correlation coefficient, and root mean square error. GCMs are ranked by relative error of all performance metrics. Results show that the temperatures from most GCM simulations are below the mean reference data (i.e., average of ground-based and reanalysis datasets), with north to south gradient in the range from 19 °C to 33 °C. In addition, all the GCM biases range from -0.07 °C to 2.78 °C and show severity of the temperature changes in spatial pattern ranging from -5 °C to 15 °C. The correlations of most GCMs range from 0.70 to 0.95, while the magnitudes of error are less than 2 °C. Study cases point out that the 13-MODEL ENSEMBLE, CESM2, and CNRM-CM6-1 perform better than the other models in simulating the temperature over Thailand for land-only and sea-only, and both land and sea cases, respectively, while MIROC6 performs the worst for all study cases in this study area. From the designed methodology, CNRM-CM6-1 has the best performance and is the most appropriate choice to simulate the temperature for the overall Thailand area.

1. Introduction

Thailand is a tropical country in Southeast Asia, located near the equator. The temperature and precipitation in Thailand are strongly affected by the El Niño–Southern Oscillation (ENSO) and tropical monsoons. Hence, this area has a high variability in its climate; it often encounters floods and droughts that are difficult to predict for climate and disaster preparedness (Haraguchi and Lall, 2015; Tan and Pereira, 2010).

Global circulation/climate models (GCMs) are mathematical models representing the physics of the climate system, including atmosphere, ocean, land surface, and cryosphere (Randall et al., 2007). GCMs are beneficial tools widely utilized to study the global climate system for both historical and future periods. Researchers from several institutions around the world have continually developed GCMs from 1995 to the present, to create efficient and reliable models for simulating climate systems (Edwards, 2011). Climate variables of GCMs are provided in grid cell format coverage of the globe (Flato et al., 2019).

GCM outputs from the climate modeling groups around the world have been collected by the World Climate Research Programme (WCRP) Working Group on Coupled Modelling (WGCM) to generate climate model experiments that are under the Coupled Model Intercomparison Project (CMIP). Coupled Model Intercomparison Project Phase 6 (CMIP6) is the most recent update. CMIP6 was planned and designed since 2013, and it was officially published in the middle of the year 2017 (Stouffer et al., 2017). Results from CMIP simulations are used to evaluate climate for several international projects, such as the assessment report of the Intergovernmental Panel on Climate Change (IPCC) for the years 2001, 2007, and 2013 (IPCC, 2014; Taylor et al., 2012); in addition, CMIP6

* Corresponding author.
E-mail address: kritana.pru@mahidol.ac.th (K. Prueksakorn).

https://doi.org/10.1016/j.heliyon.2021.e08263
Received 2 June 2021; Received in revised form 29 August 2021; Accepted 22 October 2021

simulation results are anticipated to appear in the IPCC Sixth Assessment Report (AR6) (Grose et al., 2020).

Several studies have used the climate variables of GCMs for simulating climate trends in the past, present, and future, over various regions of the world. For instance, Yan et al. (2013) assessed 25 Coupled Model Intercomparison Project Phase 5 (CMIP5) historical simulations of temperature over China and found that the multi-model ensemble of CMIP5 could well simulate the spatial pattern. Kumar et al. (2013) analyzed trends and long-term persistence of temperature and precipitation from 19 CMIP5 GCMs over continental areas (68°S–68°N) and found good agreement between the simulation and observation results. Rupp et al. (2013) evaluated the performance of 41 CMIP5 GCMs in simulating temperature and precipitation with observation data over the United States Pacific Northwest (PNW) using performance metrics. Based on the design criteria and results, CNRM-CM5 performed the best among the selected models. Miao et al. (2014) used criteria of the correlation, the centered root mean square error, and the amplitude of the standard deviations to evaluate 24 CMIP5 GCM simulations and projections over Northern Eurasia; the study discovered that most of the simulations overestimated the annual mean temperature. Siew et al. (2014) evaluated the performance of 10 CMIP5 GCMs over Southeast Asia and found large bias magnitudes for all models in the simulation of winter monsoon precipitation. Xu et al. (2017) evaluated the ability of 14 CMIP5 GCMs in simulating air temperature, specific humidity, geopotential height, and wind over Tibetan Plateau, using various criteria, including spatial correlation coefficient, spatial mean error, and standard deviation. Their ranking scores showed that CCSM4 and CNRM-CM5 performed better than the other models. Xuan et al. (2017) evaluated maximum and minimum air temperature, precipitation, wind speed, solar radiation, and relative humidity from 1971 to 2000 over Zhejiang Province in China, using 18 CMIP5 GCMs ranked by criteria of correlation coefficient, root mean square error, and model estimation bias in percentage. Six variables of all GCMs presented different spatial patterns, while five variables (except wind speed) of most simulations had similar patterns in seasonal variations. Bannister et al. (2017) evaluated the changes in recent and future temperature over Sichuan Basin in China, using 47 CMIP5 GCMs and found that most GCMs could not fairly demonstrate temperature trends, especially for seasonal cases. Agyekum et al. (2018) evaluated 18 CMIP5 GCMs in simulating various timescales of precipitation over the Volta Basin with the conclusion that ensemble means of all chosen models had better performance in the annual, seasonal, and monthly timescales than the individual models. Huang et al. (2019) analyzed vector winds for performance assessment of 37 CMIP5 GCMs and multi-model ensembles in the Asian-Australian monsoon region, using criteria of mean, annual cycle, and interannual variability and found that the multi-model ensembles had the best performance.

As aforementioned, CMIP6 GCMs are the project's last release, with just a few CMIP6 assessments in the current stage. Xin et al. (2020) simulated summer precipitation from eight CMIP6 GCMs over China and East Asia in summer during the period 1961–2005; moreover, they compared the simulations of all the CMIP6 GCMs and eight previous CMIP5 GCMs. The climatology of their study was assessed by interannual variation and linear trends, and most of the CMIP6 GCMs could simulate the interannual rainfall pattern over Eastern China better than previous CMIP5 GCMs. In summary, their study reported that the multi-model ensemble of CMIP6 had better performance than the multi-model ensemble of CMIP5. Almazroui et al. (2020) analyzed projection change in temperature and precipitation over Africa of 27 CMIP6 GCMs for the period 2030–2059, as well as for the period 2070–2099. Then, the two future periods were compared with historical climate (1981–2010). Under the highest emission scenario in CMIP6, it was predicted that the mean annual temperature at the end of the 2100 would rise by 5.6 °C over the Sahara area and by 3.5 °C over Central East Africa.

For the previous studies within the boundary of Thailand, Watanabe et al. (2014) studied monsoon precipitation using nine models of CMIP5 over Thailand, and the projection of all GCMs was compared with three reference datasets (Climate Prediction Center Merged Analysis of Precipitation: CMAP, Global Precipitation Climatology Project: GPCP, and Asian Precipitation – Highly Resolved Observational Data Integration Towards Evaluation: APHRODITE). It was emphasized that September would be the month with the largest difference (possibly between 60% and 90%) in river discharge between the retrospective simulation (1980–1999) and the future forecast (2080–2099). Supharatid (2015) analyzed the performance of Coupled Model Intercomparison Project Phase 3 (CMIP3) and CMIP5 GCMs for simulating precipitation over the Chao Phraya River Basin, Thailand, and found that the mean precipitation of GCMs in CMIP5 was closer to observation data than that of GCMs in CMIP3.

Since GCMs participating in CMIPs projects are generated by different climate institutes around the world (Taylor et al., 2012), there are different physical parameterizations and development strategies in each GCM (Hourdin et al., 2006). Moreover, the topography and climate characterization of each region are different (White et al., 2009). The internal determinations of GCMs are key to the different performances of GCMs by region, as it is difficult to input the topography and climate datasets that cover all of the globe. This is why researchers have evaluated the capacity of the GCMs regionally or locally around the world (Kumar et al., 2013; Lovino et al., 2018; Miao et al., 2014; Moise et al., 2015; Raghavan et al., 2018; Rupp et al., 2013; Su et al., 2013). Among climate parameters, the temperature variable is a common factor in many disasters frequently occurring in Thailand such as drought causing forest fires, which in turn affect human health, agriculture, and corals (Brown et al., 1996; Phongsuwan et al., 2013; Rerngnirunsathit, 2012). In order to take advantage of the updated version of CMIP, this study focuses on evaluations over Thailand, generally having large climate fluctuations (Tan and Pereira, 2010).

Regarding the period of study, climatologists have traditionally defined the climate normal as a 20- to 30-year average period (WMO, 2017); this definition has prevailed since the early twentieth century (Arguez and Vose, 2011; Trewin, 2007). However, given the current global warming crisis, it is an inescapable fact that climate variability and mechanisms have changed significantly (Le Treut et al., 2007). Anthropogenic greenhouse gas emissions from the burning of fossil fuels have led to an increase in global average surface temperatures and caused the current climate change (Trenberth et al., 2000). Space studies according to the Goddard Institute records from National Space Science Data Centre (NASA) have shown that the warmest years have occurred since 2000 and annual temperatures continue to rise (NASA/GISS, 2020), marking the beginning of the warmest decade in the records of National Oceanic and Atmospheric Administration (NOAA) Global Temp: National Centers for Environmental Information (Smith et al., 2008), NASA Goddard Institute for Space Studies (Hansen et al., 2010), and Met Office Hadley Centre and Climatic Research Unit (Morice et al., 2012). Therefore, it is possible that rising temperatures cause new weather patterns that are different from those of the last century (IPCC, 2021). In addition, the accuracy of the reference data used to assess the model's simulation results is also critical (Gampe et al., 2019); therefore; not only have the models been adjusted, but the reference data have been continuously collected with more accuracy (Dee et al., 2011; Harris et al., 2020; Rienecker et al., 2011; Willmott and Matsuura, 2001). Hence, data collected recently are likely to be more accurate than data collected over decades or centuries due to the smaller number of measurement sites and technologies in the past (Sun et al., 2018).

Due to these reasons, the study of climate over period (such as 15 years) has been considered acceptable. Fiedler et al. (2020) demonstrated this by comparing simulation results from CMIP6 GCMs with observations for historical tropical precipitation between 1900 to 1999 and 2000 to 2014. They found that there were only small differences in the spatial mean statistics between the two periods, about 0.01 mm/day (Fiedler et al., 2020). The value of the spatial correlation coefficient of CMIP6 and the observation over the tropics for both 1900 to 1999 and 2000 to 2014 was 0.84, while their root mean square error values were

1.55 and 1.57, respectively. They also examined the spatial correlation and root mean square error between CMIP6 for 1900 to 1999 and CMIP6 for 2000 to 2014 with a correlation of 0.998 and a root mean square error of 0.19 mm/day, respectively (Fiedler et al., 2020). The efficiency for the simulations of CMIP6 between the long period (20th century) and the short period (early 21st century) is quite close. Therefore, these results support the idea that climate research may no longer need a 20–30-year study period.

Given the evidence of temperature rise scenarios, the recent 15-year period presents a new challenge to the capability of GCMs from CMIP6 to simulate current temperature; therefore, the most recent decade of the historical CMIP6 experiment, the short-term period around the turn of the century (2000–2014), was chosen. The capability of 13 CMIP6 GCMs for simulations of temperature over this study area is illustrated to verify the reliability of the evaluations and to support further climate studies in Thailand.

2. Methodology

2.1. Research framework

The overall research framework of this study is shown in Figure 1. CMIP6 GCM outputs and the reference datasets (i.e., ground-based datasets and climate reanalysis datasets) evaluated in this research were linearly interpolated to the same grid having 0.15° resolution. The historical outputs of 13 CMIP6 GCMs were evaluated over Thailand for the years 2000–2014 by five statistical metrics, separately for land, sea, and both land and sea.

2.2. Study area

The study covers Thailand between 5°S–21°N latitudes and 96°E–107°E longitudes, in the tropical zone (Figure 2). The topography of Thailand has high mountains, a central plain, and an upland plateau. The study area is presented in two main parts that are land area and sea area. The land area is located in the middle of Southeast Asia, while the sea area is located between the Pacific and Indian oceans (Tantrakarnapa, 2018).

2.3. CMIP6 models

The performance of 13 GCMs in the Atmospheric Model Intercomparison Project Simulation (AMIPhist) of CMIP6 from several institutions around the world, covering the year 2014, are evaluated in this study. Data associated with this study have been deposited at https://esgf-node.llnl.gov/search/cmip6/, by choosing the features as: Source ID = GCMs' names in Table 1, Variable = tas, Frequency = mon, Experiment ID = amip-hist. Ensemble variants of each model are used to uniquely distinguish each member of an ensemble based on realizations (r), initialization schemes (i), different physics (p), and forcing (f) indices (Papalexiou et al., 2020). In this work, CMIP6 GCMs with multiple members are created by the same model but with different initial conditions. This study included all 13 available GCMs at the time of starting this research (1 March 2020). The spatial resolution and ensemble members of each GCM are presented in Table 1. Because the ensemble members of each model have different sizes, the ensemble members from each model are averaged before they are all averaged in a multi-model ensemble.

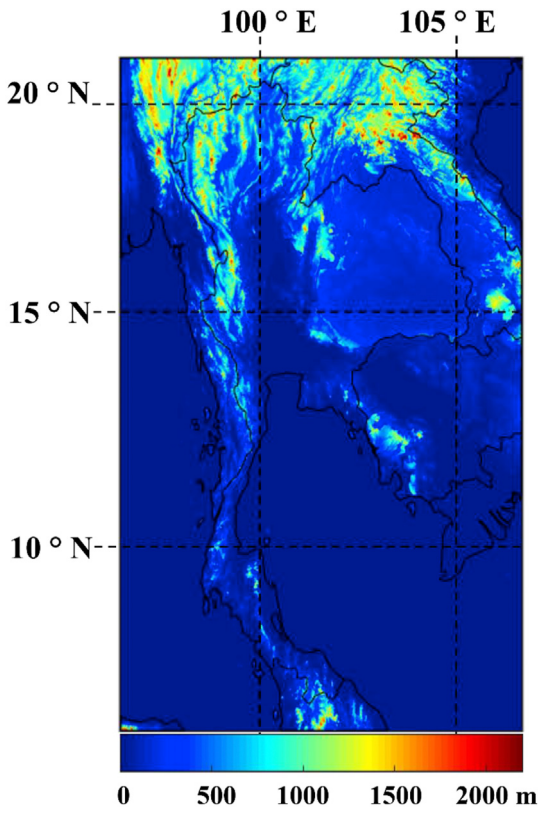

Figure 2. Topography (elevation in meters (m)) of the study area.

The multi-model ensemble is a method created by multiple model simulations. Considering the summary by several previous studies, this method can reduce the biases and uncertainties of the simulations associated with GCMs (Ahmed et al., 2018; He et al., 2019; Raju and

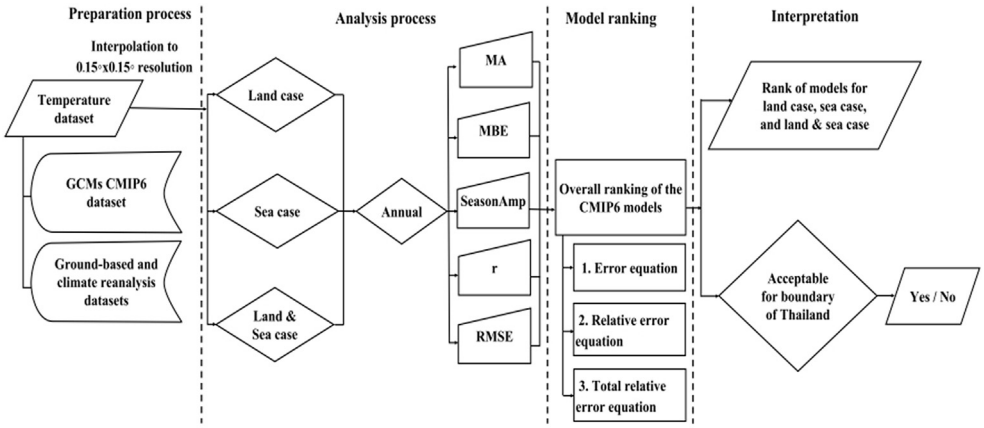

Figure 1. Research framework of this study (MA: Mean annual, MBE: Mean bias error, SeasonAmp: Mean seasonal cycle amplitude, r: the correlation coefficient, RMSE: Root mean square error).

Table 1. CMIP6 models used in this study.

GCM Name	Institution	Resolution lon × lat	Variant label	Number of Ensemble Members
BCC-CSM2-MR	Beijing Climate Center of China Meteorological Administration (BCC-CMA), China	1.12 ° × 1.12 °	r (1)i1p1f1	1
CAMS-CSM1-0	Chinese Academy of Meteorological Sciences (CAMS), China	1.12 ° × 1.12 °	r (1,2,3)i1p1f1	3
CanESM5	Canadian Centre for Climate Modelling and Analysis (CCCMA), Canada	2.8 ° × 2.8 °	r (1,2,3,4,5,6,7,...8,9,10) i1p2f1	10
CESM2	National Center of Atmospheric Research (NCAR), USA	1.25 ° × 0.9 °	r (1,2,3)i1p1f1	3
CNRM-CM6-1	Centre National de Recherches Meteorologiques, France, Centre Europeen de Recherche et de Formation Avancee en Calcul Scientifique (CNRM-CERFACS), France	1.4 ° × 1.4 °	r (1,2,3,4,5,6,7,...8,9,10) i1p1f2	10
CNRM-CM6-1-HR	Centre National de Recherches Meteorologiques, France, Centre Europeen de Recherche et de Formation Avancee en Calcul Scientifique (CNRM-CERFACS), France	0.5 ° × 0.5 °	r1i1p1f2	1
CNRM-ESM2-1	Centre National de Recherches Meteorologiques, France, Centre Europeen de Recherche et de Formation Avancee en Calcul Scientifique (CNRM-CERFACS), France	1.4 ° × 1.4 °	r1i1p1f2	1
FGOALS-f3-L	Chinese Academy of Sciences (CAS), China	1.25 ° × 1 °	r (1,2,3)i1p1f1	3
FIO-ESM-2-0	First Institute of Oceanography, Ministry of Natural Resources, Qingdao National Laboratory for Marine Science and Technology (FIO-QLNM), China	1.25 ° × 0.9 °	r (1,2,3)i1p1f1	3
GFDL-CM4	NOAA Geophysical Fluid Dynamics Laboratory (NOAA GFDL), USA	1 ° × 1.3 °	r (1,2,3)i1p1f1	3
IPSL-CM6A-LR	Institute Pierre Simon Laplace (IPSL), France	2.5 ° × 1.3 °	r (1,2,3)i1p1f1	3
MIROC6	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute, National Institute for Environmental Studies (MIROC), Japan	1.4 ° × 1.4 °	r (1,2,3,4,5)i1p1f1	5
MRI-ESM2-0	Max Planck Institute for Meteorology (MPI-M), Germany	1.12 ° × 1.12 °	r (1,2,3,4,5)i1p1f1	5

Kumar, 2020; Yan et al., 2015). In addition, several researchers have demonstrated that the multi-model ensemble method can improve climate simulation performance compared to single models (Chhin and Yoden, 2018; Hughes et al., 2014; Kamworapan and Surussavadee, 2019; Raju and Kumar, 2014, 2015).

In this study, simple mean techniques are used to calculate the average of all ensemble members for each GCM, as well as the average of the multi-model ensemble for multiple GCMs. Eq. (1) is used to calculate the average of all members of a single model, where GCMx is the average of all ensemble members of GCM x, n is the total number of ensemble members, and em_i is the simulation value of ensemble member i. Then, the multi-model ensemble of 13 GCMs (13-MODEL ENSEMBLE) is calculated by Eq. (2), where n is the total number of GCMs (Ahmed et al., 2019).

$$GCM_x = \frac{1}{n}\sum_{i=1}^{n} em_i \tag{1}$$

$$13-\text{MODEL ENSEMBLE} = \frac{1}{n}\sum_{i=1}^{n} GCM_x \tag{2}$$

To examine the ability of model simulations, monthly near-surface air temperature (Tas) of GCMs from 2000 to 2014 are extracted and compared with four reference grid datasets using performance metrics.

2.4. Reference datasets

Reference datasets are employed to compare the GCM performances; they are summarized in Table 2. The reference datasets in this study can be divided into two categories, as described below.

Table 2. Reference datasets used in this study.

Data	Resolution lon × lat	Source	Reference
UDEL v5.01	0.5 ° × 0.5 °	Ground-based	Willmott and Matsuura (2001)
CRU TS v4.02	0.5 ° × 0.5 °	Ground-based	Harris et al. (2020)
MERRA2	0.625 ° × 0.5 °	Reanalysis	Rienecker et al. (2011)
ERA-interim	0.75 ° × 0.75 °	Reanalysis	Dee et al. (2011)

2.4.1. The ground-based products

The ground-based products are meteorological instruments providing reliable accuracy because they can directly detect conditions at the weather station and have a high temporal frequency of measurements (Gilewski and Nawalany, 2018). Ground-based datasets used in this study include two monthly datasets for temperature over global land areas, for instance, (1) University of Delaware (UDEL) Air Temperature v5.01, with data available from 1900–2017 (Willmott and Matsuura, 2001) and (2) University of East Anglia Climatic Research Unit Time Series (CRU TS) v4.02, with data available from 1901 to 2017 (Harris et al., 2020).

2.4.2. The reanalysis products

The climate reanalysis product is a dataset that has integrated observations and satellite data to create climate variables with a high resolution over a long-term data record, to simulate the best representative climate estimates for all places on the globe. Furthermore, climate reanalysis products can estimate historical climates over several decades or more (Decker et al., 2012; Wargan et al., 2017). Two reanalysis datasets used in this study are (1) the Modern-Era Retrospective Analysis for Research Applications (MERRA), Version 2, with data available from 1979 to present (Bosilovich et al., 2015; Rienecker et al., 2011) and (2) the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim reanalysis (ERA-Interim), with data available from 1979 to present (Dee et al., 2011).

2.5. Performance metrics

Statistical metrics are normally used for assessing differences between model simulations and reference datasets, and each metric can show the overall (relative) performances differently for the model simulations (Gleckler and Taylor, 2008). Additionally, statistical metrics are often found in the comparison of the climate simulation results between GCMs and reference datasets (Kamworapan and Surussavadee, 2019; Li et al., 2019; McMahon et al., 2015; Miao et al., 2014; Raghavan et al., 2018; Rupp et al., 2013; Xu et al., 2017). The 13 GCMs, evaluated by five performance metrics for analysis of Tas variable, are divided into three study cases: land-only, sea-only, and both land and sea using different dataset (land case: CRU, UDEL, MERRA and ERA-Interim; sea case:

MERRA and ERA-Interim; land and sea case: CRU, UDEL, MERRA, and ERA-Interim). This is because the thermodynamic characteristics (endothermic vs. exothermic) for land area and sea area are different. For example, Tas of the land area is higher than of the sea area during the day but reduces rapidly during the night; hence, the Tas above the land area changes faster than that of the sea area (Trenberth et al., 2007). Ocean currents also affect Tas and precipitation (Reid et al., 2009), which show apparent differences seasonally in Tas (Crespo et al., 2019). The climate variables from land-atmosphere and ocean-atmosphere interactions, and the consideration of the cases separately are of interest in this study. Each study case uses different reference datasets because the climate data from CRU and UDEL are available over land only; hence, these ground-based datasets are not analyzed in the sea-only case. Results of each performance metric are averaged over the study area (i.e., 7738 grid cells). The statistical equations used in this study are shown in Table 3 with the details as follows:

(1) Mean annual (MA) refers to the sum of each mean annual Tas, divided by total number of years. It is used to assess the difference of spatial Tas by comparing between model outputs and mean reference data. For the best GCM, the MA Tas over the domain should be closest to the mean reference data.
(2) Mean bias error (MBE) is a measure of the bias, indicating positive or negative differences of the GCM data from the reference data. In other words, MBE indicates whether the GCM simulation overestimates (warm bias) or underestimates (cold bias) the mean reference data; therefore, MBE = 0 means no bias (Lovino et al., 2018).
(3) Mean seasonal cycle amplitude (SeasonAmp) is a metric indicating the difference between warmest and coldest month. It can be used to display the severity of Tas changes in each area of the study domain.
(4) Correlation coefficient (r) is employed to assess the relationship degree or to assess the similarity between the spatial patterns in GCM data and in reference data. For evaluation of model performance, the highest r values between simulated and observed results of Tas in the former studies are 0.93 for tropical (20°S–20°N) (Gleckler and Taylor, 2008), 0.85 for Tibetan Plateau (Su et al., 2013), 0.95 for China (Yan et al., 2013), 0.56 for Northern Eurasia (Miao et al., 2014), 0.96 for Northeastern Argentina (Lovino et al., 2018), 0.97 for Lower Mekong Basin (Ruan et al., 2019), and 0.99 for the global land surface (Fan et al., 2020). The highest r values from these studies range from 0.56 to 0.99, and most of them are higher than 0.9. Therefore, r value defined as good correlation in this work is equal to or more than 0.9. The statistical significance of the correlation coefficients is set at the p-value < 0.05.
(5) Root mean square error (RMSE) is a statistical metric of the magnitude of the error between the GCM and the reference data. RMSE summarizes errors between the GCM data and reference data (magnitude). The lower RMSE value, the less the error is found from simulation (Lovino et al., 2018).

Table 3. Statistical equations of the performance metrics.

Performance metric	Equation	Reference
MA	$\frac{\sum_{y=1}^{N} T_y}{N}$	Ruan et al. (2019)
MBE	$MA_m - MA_r$	Su et al. (2013)
SeasonAmp	$M_{max} - M_{min}$	Rupp et al. (2013)
r	$\frac{\sum_i \sum_j (MA_{m_{ij}} - \overline{MA_m})(MA_{r_{ij}} - \overline{MA_r})}{\sqrt{\left(\sum_i \sum_j (MA_{m_{ij}} - \overline{MA_m})^2\right)\left(\sum_i \sum_j (MA_{r_{ij}} - MA_r)^2\right)}}$	MathWorks (2001)
RMSE	$\sqrt{\frac{\sum_{p=1}^{n} (MA_{m_{ij}} - MA_{r_{ij}})^2}{n}}$	Hebeler (2020)

Notes. T_y is annual Tas for year y. N is total number of the years. MA_m is mean annual Tas of model for years 2000–2014. MA_r is mean annual Tas of reference data for years 2000–2014. M_{max} is mean warmest month of model for years 2000–2014. M_{min} is mean coldest month of model for years 2000–2014. $MA_{m_{ij}}$ is mean annual Tas of model at row i and column j. $MA_{r_{ij}}$ is mean annual Tas of reference data at row i and column j. $\overline{MA_m}$ is the average of mean annual Tas of model. $\overline{MA_r}$ is the average of mean annual Tas of reference data. n is total number of the pixels.

2.6. Model ranking

All the GCMs are ranked by performance metrics that identify the strengths and weaknesses of each model. Because of the different objects and results of each metric, the relative error is used in overall ranking of the models in this study. Model rankings that consider several metrics of the climate variables are used widely (Gleckler and Taylor, 2008; Kamworapan and Surussavadee, 2019; Radić and Clarke, 2011; Rupp et al., 2013; Waugh and Eyring, 2008; Xu et al., 2017).

The first step in finding the error for each GCM x and each performance metric y ($e_{x,y}$) is defined in Eq. (3), where $w_{r,y}$ and $w_{x,y}$ are the reference metric and the GCM result for y, respectively. Next, the relative error of for each GCM x and each performance metric y ($e^r_{x,y}$) is calculated by Eq. (4). Lastly, Eq. (5) is used to calculate the sum of the relative error of GCM x ($e_{x,total}$) by all performance metrics, when N is the total number of performance metrics.

$$e_{x,y} = |w_{r,y} - w_{x,y}| \tag{3}$$

$$e^r_{x,y} = \frac{e_{x,y} - \min(e_{x,y})}{\max(e_{x,y}) - \min(e_{x,y})} \tag{4}$$

$$e_{x,total} = \sum_{y=1}^{N} e^r_{x,y} \tag{5}$$

All performance metrics are assigned equal weights, as applied to evaluate the GCMs in previous studies (Gleckler and Taylor, 2008; Kamworapan and Surussavadee, 2019; Rupp et al., 2013), as well as in the model ranking by overall performance metrics (Gleckler and Taylor, 2008; Kamworapan and Surussavadee, 2019; Rupp et al., 2013). Robustness or confidence for ranking of performance metrics in this study is grouped "highest" and "higher" as recommended by Rupp et al. (2013).

3. Results and discussion

3.1. Annual Tas

3.1.1. Mean annual (MA)

Figure 3 shows spatial annual Tas for mean reference data, 13-MODEL ENSEMBLE, and the individual 13 GCMs for land and sea case, as this case can represent the overall spatial pattern for both the land-only cases and sea-only cases. Spatially, all GCMs had the Tas gradient in the same direction with low Tas in the north of Thailand (especially over high mountains) and high Tas in Southern Thailand. The mean Tas of the 13 CMIP6 GCMs ranged from 26.22 °C to 29.06 °C. The simulation result evaluated by MA showed that GFDL-CM4 was very similar to the mean reference data; the difference was only 0.07 °C, whereas the mean Tas of MIROC6 performed the worst in terms of both magnitude and spatial pattern, being very different from the mean reference data. The Tas simulation of MIROC6 was 2.77 °C higher than the mean reference data.

MA was used to assess CMIP5 GCMs over Lower Mekong Basin (Ruan et al., 2019) and Eastern Tibetan Plateau (Su et al., 2013) during two periods (i.e., 1961–2004 and 1961–2005). GFDL-CM3 was ranked the

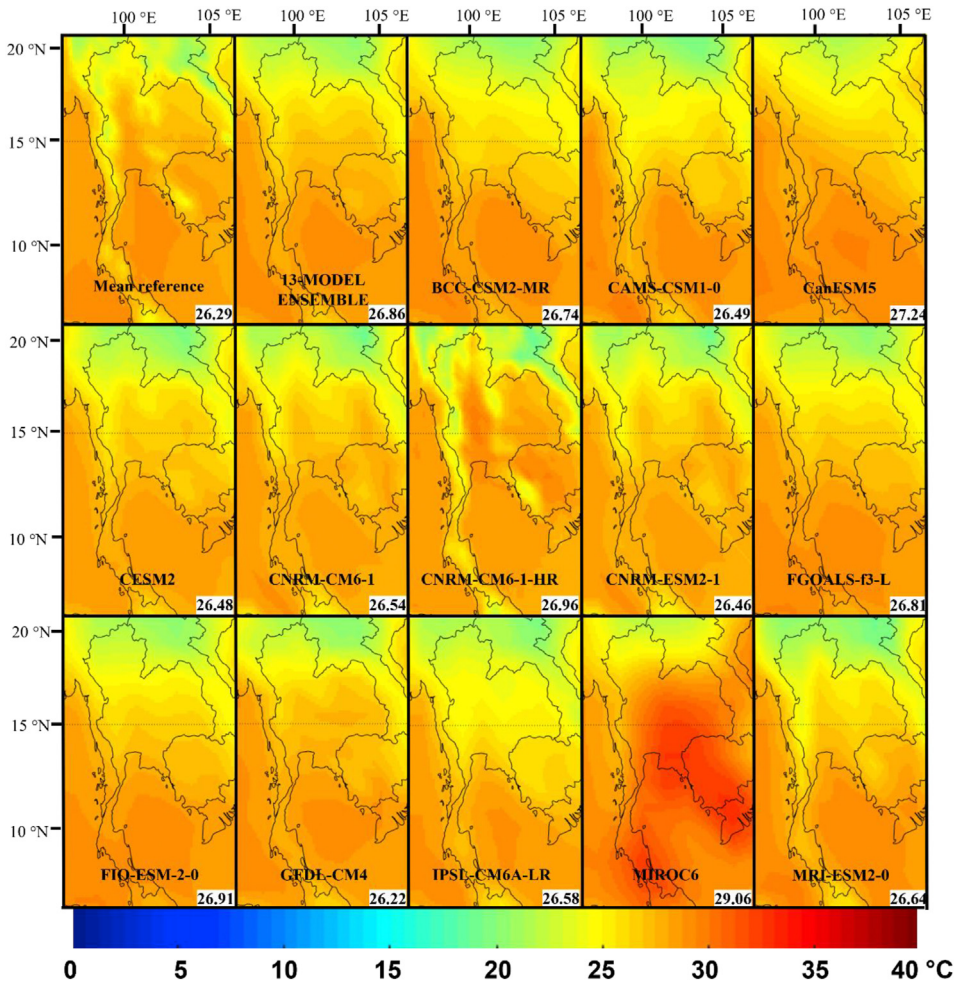

Figure 3. Mean annual Tas (MA) (°C) of mean observations, 13-MODEL ENSEMBLE, and 13 GCMs during 2000–2014, with corresponding mean values at the bottom-right of each sub-plot.

Table 4. The best CMIP6 GCM in this study and CMIP5/CMIP3 GCMs ranking results from the same institution in former studies.

Metrics	Model with the best performance	Former studies				
	CMIP6 & Institution	reference	variable	year	area	CMIP & rank
MA	GFDL-CM4 (NOAA GFDL)	Ruan et al. (2019)	annual Tas	1961–2004	Lower Mekong Basin	GFDL-CM3* (23/34)
		Su et al. (2013)	annual Tas	1961–2005	eastern Tibetan Plateau	GFDL-CM3* (20/24)
MBE	GFDL-CM4 (NOAA GFDL)	Rupp et al. (2013)	annual Tas	1960–1999	Pacific Northwest USA	GFDL-CM3* (20/41)
		Miao et al. (2014)	annual Tas	1901–2005	Northern Eurasia	GFDL-CM3* (3/24)
		Xuan et al. (2017)	annual maximum Tas	1971–2000	Zhejiang Province in China	GFDL-CM3* (18/18)
			annual minimum Tas	1971–2000	Zhejiang Province in China	GFDL-CM3* (17/18)
		Miao et al. (2014)	monthly Tas	1901–2005	Northeastern Argentina	GFDL-CM3* (15/25)
		Su et al. (2013)	annual Tas	1901–2005	Eastern Tibetan Plateau	GFDL-CM3* (20/24)
SeasonAmp	CNRM-CM6-1-HR (CNRM-CERFACS)	Rupp et al. (2013)	annual Tas	1960–1999	Pacific Northwest USA	CNRM-CM5* (11/41) CNRM-CM5-2* (13/41)
r	CNRM-CM6-1 CNRM-CM6-1-HR CNRM-ESM2-1 (CNRM-CERFACS	Ruan et al. (2019)	annual Tas	1961–2004	Lower Mekong Basin	CNRM-CM5* (23/34)
		Zhou and Yu (2006)	annual Tas	1880–1999	Global Northern Hemispheric China	CNRM-CM3** (9/19) CNRM-CM3** (9/19) CNRM-CM3** (3/19)
		Xu et al. (2017)	annual Tas	1979–2005	Tibetan Plateau	CNRM-CM5* (1/14)
RMSE	CNRM-ESM21 (CNRM-CERFACS)	Miao et al. (2014)	monthly Tas	1901–2005	Northeastern Argentina	CNRM-CM5-2* (10/25)
		Ruan et al. (2019)	annual Tas	1961–2004	Lower Mekong Basin	CNRM-CM5* (6/34)
		Xuan et al. (2017)	annual maximum Tas	1971–2000	Zhejiang Province in China	CNRM-CM5* (11/18)
			annual minimum Tas	1971–2000	Zhejiang Province in China	CNRM-CM5*(5/18)
		Grose et al. (2014)	annual Tas	1980–1999	Pacific Ocean region	CNRM-CM5-1* (8/27)

* is a model in CMIP5, while ** is a model in CMIP3.

worst performing model group, ranking 23 out of 34 models and 20 out of 24 models in Ruan et al. (2019) and in Su et al. (2013), respectively (Table 4). Thus, the ranking results of both those studies demonstrated that GFDL-CM3 was not adequate for simulating the mean annual Tas over tropical and subtropical zones. As for the performance comparisons of Tas simulation between GFDL-CM3 in CMIP5 of Ruan et al. (2019) and Su et al. (2013), and GFDL-CM4 in CMIP6 of this study, the findings were very different, although the study area of Ruan et al. (2019) overlapped with this current study. GFDL-CM4, the GCM from NOAA GFDL institution, in this study showed the best performance in MA, especially for the simulation over the tropical zone. The ability of GFDL-CM4 had significant improvement, which may be attributed to a new version of the physical climate model along with spatial resolution more than double of GFDL-CM3 (resolution 2.5 ° × 2 °) (Held et al., 2019). These updates could be the key factors of the performance results because GFDL-CM4 had higher efficiency than the previous version.

When considering the ranking of the GCM from MIROC institution (the worst performing in MA), the model ranking results of Ruan et al. (2019) and Su et al. (2013) were different from this current study, because MIROC5 in Ruan et al. (2019) and Su et al. (2013) was ranked 7th out of 34 models and 8th out of 24 models, respectively (Table 5). Thus, MIROC5 in CMIP5 was a well-performing model group for Tas simulation. On the other hand, MIROC6 performance (the latest version from the MIROC institution) under the CMIP6 project showed the worst performance in this current study, although MIROC6 already determined new physical parameterizations in sub-modules (Tatebe et al., 2019). In addition, the horizontal resolution of MIROC6 was not higher than the previous version in CMIP5. Tatebe et al. (2019) reported that increasing the horizontal resolution of the model also means an increased computational cost; however, many GCMs in CMIP6 were developed in this part while MIROC6 was not (Boucher et al., 2020; Grise and Davis, 2020; Held et al., 2019; Séférian et al., 2019).

3.1.2. Mean bias error (MBE)

Figure 4 shows the spatial distribution of bias in MA Tas over the study area relative to mean reference data, 13-MODEL ENSEMBLE, and the individual 13 GCMs. Numbers in each sub-image are biases of spatial annual Tas for each simulation. The MBE of 13 CMIP6 GCMs ranged from -0.07 °C to 2.78 °C and most of CMIP6 GCMs showed overestimates of the mean reference values, except for GFDL-CM4 that underestimated Tas with the mean bias of -0.07 °C (or mean bias of -3.10 %); moreover, GFDL-CM4 had the lowest bias in this study among the GCMs tested. Most models showed MBE in the range of ±1 °C. MIROC6 was the only GCM showing MBE above 2 °C, the highest MBE in this study. MIROC6 showed a positive direction (warm bias), and over the study area, its mean bias was 31.50%. Reviewing the bias of each season, it was found that summer and rainy seasons had a significant effect on simulating temperature in Thailand. All GCMs exhibited a positive bias in simulated summer and rainy temperatures, with summer bias ranging from 0.76 °C to 4.47 °C and rainy season bias ranging from 1.53 °C to 4.07 °C. In contrast, all GCMs had the lowest bias values in the winter season ranging from -0.46 °C to 3.05 °C. However, the seasonal performance of 13-MODEL ENSEMBLE was able to reduce the relatively large bias that appeared in summer and rainy of all GCMs to about -0.17 °C and 0.04 °C, respectively. For temperature in the study area, it is possible that the bias of each season is caused by the amount of precipitation in the summer and rainy months.

CMIP6 models have improved their climate simulations compared to previous generations; however, they still have persistent biases and uncertainties, especially a warm bias over tropical regions (Kim et al., 2020). In addition, Arias et al. (2021) discovered that CMIP6 GCMs have limitations in simulating air surface temperatures in areas with complex topography. Looking at the temperature bias in this study, almost all CMIP6 GCMs show a similar spatial pattern of their biases, with surface temperature generally being underestimated in low-altitude areas, while overestimations are observed in high-altitude areas. Hence, the overestimation bias is also likely due to model deficiencies associated with topographic parameters.

Although GCMs are designed to simulate atmospheric and ocean processes, the structure of GCMs clearly separates the elements of the climate system: atmosphere, oceans, cryosphere, biosphere, and geosphere (Le Treut et al., 2007). The internal climate sub-models of the GCMs, which focus on simulating each component of the climate system, are also different (Gent, 2012). In addition, Zhao and Li (2015) reported

Table 5. The worst CMIP6 GCM in this study and CMIP5/CMIP3 GCMs ranking results from the same institution in former studies.

Metrics	Model with the worst performance	Former studies				
	CMIP6 & Institution	reference	variable	Year	area	CMIP & rank
MA	MIROC6 (MIROC)	Ruan et al. (2019)	annual Tas	1961–2004	Lower Mekong Basin	MIROC5* (7/34)
		Su et al. (2013)	annual Tas	1961–2005	Eastern Tibetan Plateau	MIROC5* (8/24)
MBE	MIROC6 (MIROC)	Rupp et al. (2013)	annual Tas	1960–1999	Pacific Northwest USA	MIROC5* (22/41)
		Miao et al. (2014)	annual Tas	1901–2005	Northern Eurasia	MIROC5* (15/24)
		Xuan et al. (2017)	annual maximum Tas	1971–2000	Zhejiang Province in China	MIROC5* (1/18)
			annual minimum Tas	1971–2000	Zhejiang Province in China	MIROC5* (10/18)
		Miao et al. (2014)	monthly Tas	1901–2005	Northeastern Argentina	MIROC5* (15/25)
		Su et al. (2013)	annual Tas	1961–2005	Eastern Tibetan Plateau	MIROC5* (8/24)
SeasonAmp	FGOALS-f3-L (CAS)	Rupp et al. (2013)	annual Tas	1960–1999	Pacific Northwest USA	FGOALS-s2* (41/41)
r	MIROC6 (MIROC)	Zhou and Yu (2006)	Annual Tas	1960–1999	Global Northern Hemispheric China	MIROC3.2** (10/19) MIROC3.2** (6/19) MIROC3.2** (6/19)
		Ruan et al. (2019)	annual Tas	1961–2004	Lower Mekong Basin	MIROC5* (5/34)
		Xu et al. (2017)	annual Tas	1979–2005	Tibetan Plateau	MIROC4h* (6/14)
		Su et al. (2013)	annual Tas	1961–2005	Eastern Tibetan Plateau	MIROC5* (14/24)
RMSE	MIROC6 (MIROC)	Xuan et al. (2017)	annual maximum Tas	1971–2000	Zhejiang Province in China	MIROC5* (2/18)
			annual minimum Tas	1971–2000	Zhejiang Province in China	MIROC5* (9/18)
		Miao et al. (2014)	monthly Tas	1901–2005	Northeastern Argentina	MIROC5* (8/25)
		Ruan et al. (2019)	annual Tas	1961–2004	Lower Mekong Basin	MIROC5** (30/34)
		Su et al. (2013)	annual Tas	1961–2005	Eastern Tibetan Plateau	MIROC5** (9/24)
		Grose et al. (2014)	annual Tas	1980–1999	Pacific Ocean region	MIRCO5** (13/27)

* is a model in CMIP5, while ** is a model in CMIP3.

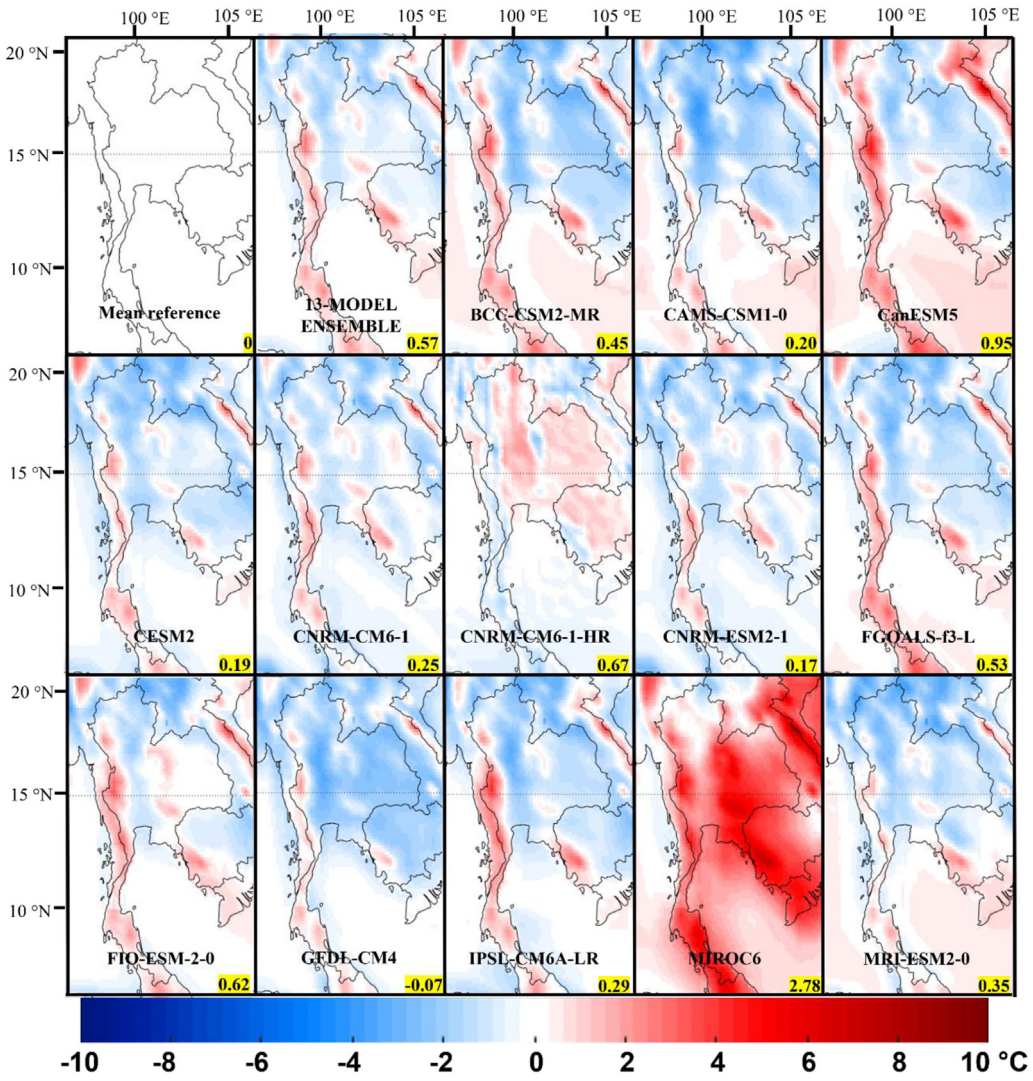

Figure 4. Spatial distributions of bias in mean annual Tas (MBE) (°C) of mean observations, 13-MODEL ENSEMBLE, and 13 GCMs during 2000–2014, with corresponding MBE values at the bottom-right of each sub-plot.

that the land-atmosphere interactions are usually difficult to simulate due to the influence of complex topography and altitude differences. Therefore, the bias patterns of temperature in Figure 4 are much higher or lower over land than over ocean areas, which is due to the performance of sub-models to simulate land-atmosphere and ocean-atmosphere interactions and the terrain characteristics of the focused areas.

MBE was the statistical metric used to assess the Tas simulation of GCMs over several regions in previous studies, for example, Pacific Northwest USA (Rupp et al., 2013), Northern Eurasia (Miao et al., 2014), Zhejiang Province in China (Xuan et al., 2017), Northeastern Argentina (Lovino et al., 2018), and Eastern Tibetan Plateau (Su et al., 2013) (Table 4). In GCMs ranked by MBE, Miao et al. (2014) found that GFDL-CM3 in CMIP5 had a good performance; it was ranked 3rd out of 24 models, while Kumar et al. (2013) and Lovino et al. (2018) found that it had moderate performance. In addition, Xuan et al. (2017) and Su et al. (2013) reported that GFDL-CM3 was a poor-performing model for Tas simulation, and it was also ranked as the group of worst performing models. These previous studies revealed interesting points, namely that GFDL-CM3 could well simulate the Tas over the temperate zone as well as the polar zone. On the other hand, the performance of GFDL-CM3 was extremely poor in simulating over the tropical and subtropical zones.

NOAA GFDL institution, the generator of GFDL-CM group, had put a focus on developing GFDL-CM4 for better simulation of Tas over the tropical zone (Held et al., 2019). As a result, GFDL-CM4 participating in the CMIP6 project had a better performance in simulating Tas over Thailand. The previous models by the MIROC institution (the worst performing in MBE) had different performances by region. Miao et al. (2014), Ruan et al. (2019), Rupp et al. (2013), Su et al. (2013), and Xuan et al. (2017) reported that MIROC5 was good to moderate when ranked by MBE. It received an excellent ranking by Xuan et al. (2017) and Su et al. (2013). On the other hand, the evaluation of MIROC6 model, the next version of MIROC5 model, showed different simulation performance from those studies. MIROC6 showed the highest MBE value compared to the simulations of other models, and this finding is consistent with the result of Fan et al. (2020).

It may be noted that the study areas of well-performing groups in previous studies (GFDL-CM3 in Miao et al. (2014) and MIROC5 in Xuan et al. (2017) and Su et al. (2013)) were also near the institutions that had developed the software, both for GFDL-CM3 and MIROC5; these are institutions in the United States and Japan, respectively. It might be that the physical parameters of these models were tuned for Tas simulations based on their own locations as a priority.

3.1.3. Mean seasonal cycle amplitude (SeasonAmp)

Figure 5 shows Tas gradient spatial patterns for all simulations ranging from -5 °C to 15 °C. Numbers in each sub-image are the changes of spatial annual Tas in each simulation. Out of the 13 GCMs, all obviously showed that the northern region of Thailand had more Tas changes than the southern region; these results were consistent with the reality of Thailand's climate (Tan and Pereira, 2010). This study found that CNRM-CM6-1-HR was the best in simulating the mean Tas changes, with the value closest to the mean reference data in terms of magnitude (higher only 0.1 °C). Although CNRM-CM6-1-HR was the best GCM that could simulate the change of mean Tas in terms of magnitude (total of the area), it was not the best simulation in terms of shape. 13-MODEL ENSEMBLE was the best simulation showing consistency with the mean reference data in terms of shape, noticeable in the northern region of the study area, while its mean change value was only 0.35 °C lower than the mean reference. FGOALS-f3-L, generated by CAS institution, was GCM that had the highest change from the mean reference data, with the change value of 6.29 °C (2.45 °C higher than the mean reference).

The SeasonAmp metric was reported by Rupp et al. (2013) as one of the highest confidence metrics for CMIP5 ranking over Pacific Northwest USA. They reported that multi-model simulation could simulate the severity of T change close to the observed data, with Tas gradient as well as Tas change value of the multi-models within 1 °C (Rupp et al., 2013). In addition, they found that CNRM-CM5 and CNRM-CM5-1 created by CNRM-CERFACS institution were similarly ranked 11th and 13th out of 41 models, respectively. FGOALS -s2 was found to have the worst performance for simulating the severity of Tas change. FGOALS-s2 was ranked 41st out of 41 models (lowest ranked), consistent with the ranking in this current study, as FGOALS-s2 of this study also was ranked the lowest. Although the latest model from the CAS institution participating CMIP6 had updated both the atmospheric and oceanic model (Guo et al., 2020), FGOALS-f3-L still performed worst in simulating the severity of Tas change when compared with the other GCMs, especially over Thailand.

3.1.4. Correlation coefficient (r)

Correlation coefficient (r) was used in this study for measuring the relationship level between model and reference data in each grid-cell. Table 6 shows r values of mean Tas for the individual 13 GCMs and for the 13-MODEL ENSEMBLE. These metrics were computed for the three study cases including land-only, sea-only, and both land and sea. The correlation coefficient is used for measuring the relationship level between model and reference data in each grid cell. The correlation values in Table 6 showed that most GCMs had correlations in the range good, with coefficients in the numerical range from 0.70 to 0.95. Evaluation of CMIP6 GCMs by r showed that three models from CNRM-CERFACS institution had the highest correlations to the reference data. CNRM-CM6-1-HR was the first rank for the land-only case, while CNRM-CM6-1 and CNRM-ESM2-1 were the first rank for both land and sea case, presenting the best correlation (r = 0.96). For the sea-only case, 13-MODEL ENSEMBLE

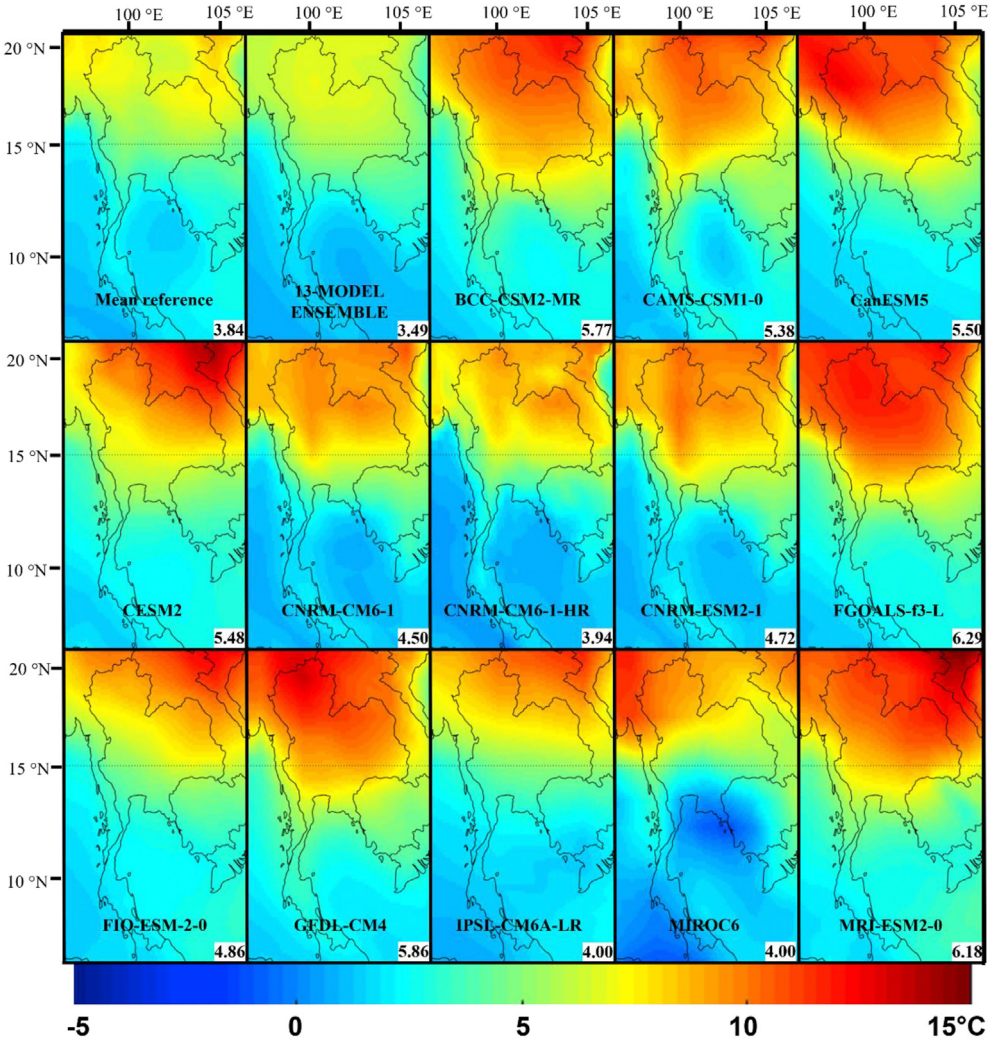

Figure 5. Comparisons of the mean seasonal cycle amplitudes of Tas (SeasonAmp) (°C) of mean observations, 13-MODEL ENSEMBLE, and 13 GCMs during 2000–2014, with corresponding SeasonAmp values at the bottom-right of each sub-plot.

Table 6. r and RMSE between observations and model simulations of mean annual Tas for years 2000–2014, for three cases: land only, sea only, and both land and sea.

Model		land-only		sea-only		both land and sea	
		r	RMSE (°C)	r	RMSE (°C)	r	RMSE (°C)
1.	BCC-CSM2-MR	**0.82**	1.21	0.94	1.15	**0.89**	1.19
2.	CAMS-CSM1-0	**0.87**	1.15	**0.92**	1.09	**0.91**	1.13
3.	CanESM5	**0.75****	1.51	**0.83**	1.34	**0.84**	1.44
4.	CESM2	**0.91**	0.90	**0.94**	0.62	**0.94**	0.80
5.	CNRM-CM6-1	**0.94**	0.75*	**0.92**	0.54	**0.96***	0.67
6.	CNRM-CM6-1-HR	**0.95***	1.07	**0.87**	0.76	**0.95**	0.95
7.	CNRM-ESM2-1	**0.94**	0.76	**0.92**	0.49*	**0.96***	0.66*
0	FGOALS-f3-L	**0.85**	1.16	**0.89**	1.10	**0.91**	1.13
9.	FIO-ESM-2-0	**0.90**	1.06	**0.93**	0.94	**0.93**	1.01
10.	GFDL-CM4	**0.88**	1.11	**0.95**	0.75	**0.91**	0.98
11.	IPSL-CM6A-LR	**0.84**	1.17	**0.85**	0.79	**0.90**	1.03
12.	MIROC6	**0.82**	3.63**	0.14**	2.45**	**0.74****	3.19**
13.	MRI-ESM2-0	**0.93**	1.01	**0.93**	1.14	**0.95**	1.06
	13-MODEL ENSEMBLE	**0.92**	0.88	**0.96***	0.95	**0.95**	0.91

* is the best value for each metric's evaluation, ** is the worst value for each metric's evaluation, a significance level of 0.05 is shown in boldface.

showed the best correlation, while only MIROC6 showed an obviously unsatisfactory correlation (r = 0.14) value in this case.

Zhou and Yu (2006) used the r metric to assess the performance of CMIP3 GCMs and found that the CNRM-CM3 simulations had different levels of relationship depending on the study area, ranging from moderate to good. It was ranked 3rd to simulate the temperature over the whole of China with a good correlation, but it was ranked 9th out of 19 models over global and the Northern Hemisphere, with a moderate correlation (Table 4). Moreover, the r metric was used to assess the performances of CMIP5 GCMs over the Tibetan Plateau (Xu et al., 2017) and Lower Mekong Basin (Ruan et al., 2019). Xu et al. (2017) and Ruan et al. (2019) found that CNRM-CM5 created by CNRM-CERFACS institution had correlations of 0.89 and 0.91, respectively, considered as "good." Xu et al. (2017) showed that CNRM-CM5 performed best in simulating Tas and was ranked 1st. The CNRM-CM5 evaluated by Ruan et al. (2019) had a high correlation of 0.91, but it was still ranked 23rd out of 34 models. Hence, in the perspective of ranking-based performance of Tas simulations, the previous studies found that the GCMs created by CNRM-CERFACS institution under both CMIP3 and CMIP5 projects might depend on the study area. Moreover, previous versions of CNRM performed well over subtropical and temperate zones, but predictions for the Lower Mekong Basin located in the tropical zone, covering a part of this study area, had a poor performance. In this study, CNRMCM6-1, CNRM-CM6-1-HR, and CNRM-ESM2-1 in CMIP6 were listed as the three highest-ranked models by the r metric. Séférian et al. (2019) and Voldoire et al. (2019) reported that CNRM-family participating in CMIP6 project had updated atmosphere and land surface components. This update of CNRM-family might improve the simulation performance over Thailand.

The previous version of MIROC model under CMIP3 and CMIP5 in former studies showed quite a satisfactory ranking for Tibetan Plateau (Xu et al., 2017), global, Northern Hemisphere, China (Zhou and Yu, 2006), and Eastern Tibetan Plateau (Su et al., 2013). Moreover, MIROC5 in CMIP5 had outstanding ranking in the Lower Mekong Basin (Ruan et al., 2019), ranking 5th out of 34 models.

In this current study, the r metric for MIROC6 was the worst for the sea-only case and both land and sea, and was the second worst for the land-only case. A major factor might be the design of MIROC6 as Atmosphere General Circulation Model (AGCM), while MIROC5 in CMIP5 project was Coupled Atmosphere-Ocean General Circulation Model (AOGCM). This difference might be the reason that climate simulation in MIROC5 was better than in MIROC6; however, AOGCMs had disadvantage in requiring more data and computation time than AGCMs (Tatebe et al., 2019).

3.1.5. Root mean square error (RMSE)

RMSE was used to compare between GCMs and reference data, to indicate the magnitudes of errors in each grid cell. The evaluation of GCMs by RMSE for the three study areas revealed that 12 of 13 CMIP6 GCMs showed magnitudes of overall error less than 2 °C. The two GCMs with the lowest errors were CNRM-ESM2-1 and CNRM-CM6-1, with magnitudes of error less than 1 °C. CNRM-ESM2-1 performed the best for sea-only case and for both land and sea case, with RMSE of 0.49 °C and 0.66 °C, respectively, while CNRM-CM6-1 performed the best for the land-only case, with RMSE of 0.75 °C (Table 6). In contrast, the GCM that showed the largest errors was MIROC6, with RMSE higher than 2 °C.

GCM rankings of Ruan et al. (2019) and Grose et al. (2014) by RMSE showed that CNRM-CM5 could perform very well over the Lower Mekong Basin and the Pacific Ocean, ranking 6th out of 34 models and 8th out of 27 models, respectively, whereas Lovino et al. (2018) and Xuan et al. (2017) found that GCMs from the CNRM-CERFACS institution had good-to-moderate performance for Tas simulation. The evaluation of GCMs by RMSE by these former studies showed that the CNRM-CM family of models as a group had good performance for simulating Tas over the tropical zone and moderate performance over the subtropical zone. Moreover, this study confirmed that the three new GCMs generated by the CNRM institution under the CMIP6 project had an obvious improvement having the first rank for three metrics: SeasonAmp, r, and RMSE.

MIROC5 performed poorly in simulating the Lower Mekong Basin and was ranked 30th out of 34 models, as reported by Ruan et al. (2019). However, MIROC5 could perform satisfactorily and had a moderate ranking over Northeastern Argentina (Lovino et al., 2018) and Eastern Tibetan Plateau (Su et al., 2013), while MIROC5, according to Xuan et al. (2017), performed 2nd best for annual maximum Tas over Zhejiang Province in China. Therefore, the evaluation of GCMs with RMSE in previous studies and this study indicate that MIROC6 might not be suitable for simulating Tas over the tropical zone.

CMIP6 GCMs generated by the NOAA GFDL and the CNRM-CERFACS institution were clearly prominent in this study area. Looking at the GFDL-CM4 and CNRM-CM families, it was discovered that their evolution was focused on the sub-model component.

The physical climate system of GFDL-CM4 has evolved from CM3 to CM4.0. In addition, the horizontal resolution in the atmosphere of GFDL-CM4 is about 100 km, which is the finer resolution than GFDL-CM3 (~200 km) (Held et al., 2019). The Atmosphere Model version 4 (AM4) for GFDL, used as the atmospheric component of CM4.0, has improved horizontal resolution, new convection and mountain drag parameterization with radiative transfer, and aerosol-cloud interactions

significantly updated (Zhao et al., 2018). Thus, the horizontal resolution and model physics in AM4 are the key factors to improve the tropical temperature simulation capability of the model.

The physical core of CNRM-CM6-1-HR and CNRM-ESM2-1 is similar to that of CNRM-CM6-1, but CNRM-CM6-1-HR is a higher resolution version, while CNRM-ESM2-1 adds a representation of the global carbon cycle, atmospheric chemistry, and aerosols. All sub-model components in three CNRM models were updated. However, there were no significant changes in the parameterizations for the ocean and sea ice components; on the other hand, the atmospheric and land surface components were completely redesigned using new parameterizations. Voldoire et al. (2019) reported that the equilibrium climate sensitivity of three CNRMs in CMIP6 is much higher than that of the previous version in CMIP5. These major upgrades are most likely a key reason for the better performance over older GCMs in the same institution.

In MIROC6, the atmospheric component is a sub-model component that has major changes compared to the last version of MIROC5 in CMIP5, which is the implementation of a parameterization of shallow convective processes, the higher model top, and the vertical resolution in the stratosphere. The sub-models of MIROC6 consist of three main components (atmosphere, land, and sea ice–ocean), while that of MIROC5 consist of four main components (atmosphere, land, ocean, and sea ice) (Tatebe et al., 2019; Watanabe et al., 2010). The reduced performance of MIROC6 is most likely due to the combination of sub-model components for the ocean and sea ice.

Compared to previous versions, FGOALS-f3-L has higher resolution as well as a larger upgrade in the sub-model component of the atmospheric and ocean model (Guo et al., 2020). However, these changes are not sufficient to improve the performance of the model, considering the difference between the warmest and coldest months.

For all the above reasons, the different performances of GCMs might depend on spatial resolution and sub-model components; in addition, the performance of GCMs might depend on the area where the GCM developer had the highest interest, via parametrizations in components of the climate system. These are just some noteworthy aspects found in this study; however, in order to improve the poorly performing GCMs, further investigation is needed.

3.2. Model ranking by Tas performance metrics

In Figure 6, the GCMs are sorted from highest to lowest total relative error from top to bottom. The evaluation of GCMs by five performance metrics showed that 13-MODEL ENSEMBLE, CESM2, and CNRMCM6-1 were the best GCMs for land-only, sea-only, and both land and sea cases, with total relative errors 0.51, 0.56, and 0.43, respectively. Besides, the CNRM group performed well for simulating Tas over Thailand because they had good ranks. The results of the CNRM group for both land and sea cases of this study were consistent with the CNRM-CM5 evaluation results over Northwest USA by Rupp et al. (2013) and CNRM-CM5-2 over Southeast Asia by Kamworapan and Surussavadee (2019). On the other hand, MIROC6 performed the worst in all study cases. Moreover, the findings also reveal that the performance of 13-MODEL ENSEMBLE was obviously outstanding. It was discovered to be consistent with the reference data and to be capable of reducing simulation biases and uncertainties, especially its performance ranking result in land-only cases and both land and sea cases. The evaluation of the general metrics of 13-MODEL ENSEMBLE is quite good, while that of a single model fluctuate significantly. Frequently, some metrics from a single model perform well, while the other metrics from that also perform poorly.

4. Conclusion

This study evaluated the performances of 13 CMIP6 GCMs in simulating Tas over Thailand for the 15- year period of 2000–2014. Ground-based products (UDEL v5.01 and CRU TS v.4.02) and reanalysis products (MERRA2 and ERA-interim) were used to evaluate the GCMs. GCM ranking in this study was calculated by taking the summation of relative errors from five performance metrics including MA, MBE, SeasonAmp, r, and RMSE. Most CMIP6 GCMs (except for MIROC6) were able to simulate Tas over Thailand reasonably well. However, CMIP6 GCMs (except GFDL-CM3) tended to overestimate (positive direction of error). Differences of highest and lowest temperatures from the GCMs were larger than in the reference data, especially in the northern part of Thailand. Twelve out of 13 models (except MIROC6) had high correlations with the reference data. RMSE values showed that the GCMs had very similar performances, except for MIROC6. The evaluation of the 13 CMIP6 GCMs by five performance metrics indicated that GFDL-CM4 and CNRM group participating under the CMIP6 project were the best by performance. In particular, GFDL-CM4 was ranked first by MA and MBE, while the CNRM group was ranked first by SeasonAmp, r, and RMSE. On the other hand, MIROC6 performed the worst in almost all measures (except for SeasonAmp, where MIROC6 was third ranked out of 13 models). In summary, model ranking by the total relative error showed that CNRM-CM6-1 was the best performing model followed by CNRM-ESM2-1, CNRM-CM6-1-HR, 13-MODEL ENSEMBLE, IPSLCM6A-LR, and CESM2 in this order, while MIROC6 was the worst performing model. The results of this study can benefit regional climate scientists. Input datasets from the best GCMs of this study could be downscaled to a finer spatial resolution over Thailand, for simulating the other climate variables (for example,

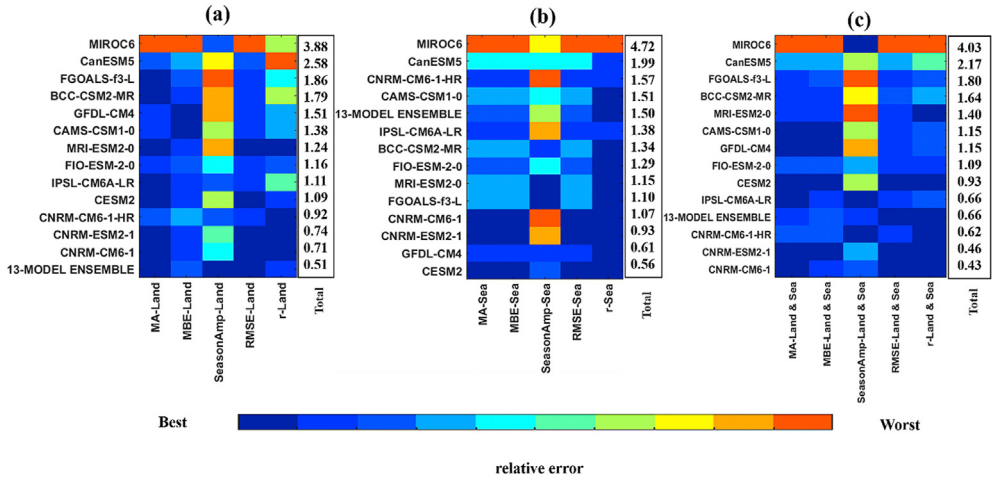

Figure 6. Relative error metrics of Tas variable for five statistical metrics (horizontal ordinate) for each CMIP model and model ensemble mean (vertical ordinate) for (a) land-only, (b) sea-only, and (c) both land and sea. The last column is total score of relative error over five performance metrics.

temperature, precipitation, wind) and for projections of future climate. In future work, this study could be extended to 1) cover other areas in tropical zone and 2) to include other latest CMIP6 GCMs releases.

Declarations

Author contribution statement

Suchada Kamworapan: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Pham Thi Bich Thao & Shabbir H. Gheewala: Analyzed and interpreted the data.

Sittichai Pimonsree: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Kritana Prueksakorn: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

This study was supported by (1) Faculty of Environment and Resource Studies, Mahidol University, (2) Andaman Environment and Natural Disaster research center (ANED), Faculty of Technology and Environment (FTE), Prince of Songkla University, (3) National Research Council of Thailand (Grant no. R63035), (4) Joint Graduate School of Energy and Environment (JGSEE), King Mongkut's University of Technology Thonburi, and (5) Center of Excellence on Energy Technology and Environment (CEE), PERDO, Ministry of Higher Education, Science, Research and Innovation.

Data availability statement

Data associated with this study has been deposited at https://esgf-node.llnl.gov/search/cmip6/by choosing the features as: Source ID = GCMs' names in Table 1, Variable = tas, Frequency = mon, Experiment ID = amip-hist.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

Acknowledgements

The authors acknowledge climate modeling groups for generating their model output and also the World Climate Research Programme's Working Group on Coupled Modelling for sharing CMIP6 GCM data. Moreover, we would like to thank Dr. Jinda Sawattawee, Dr. Tanwa Arpornthip, Ms. Thanchanok Noosook and Ms. Hong Anh Thi Nguyen for their technical suggestions and valuable supports. We also thank to the Research and Development office (RDO), Prince of Songkla University for their language proofreading and editing support.

References

Agyekum, J., Annor, T., Lamptey, B., Quansah, E., Agyeman, R.Y.K., 2018. Evaluation of CMIP5 global climate models over the Volta Basin: precipitation. Adv. Meteorol. 2018, 1–24.

Ahmed, K., Sachindra, D.A., Shahid, S., Demirel, M.C., Chung, E.S., 2018. Selection of multi-model ensemble of GCMs for the simulation of precipitation based on spatial assessment metrics. Hydrol. Earth Syst. Sci. Discuss. 1–35.

Ahmed, K., Sachindra, D.A., Shahid, S., Demirel, M.C., Chung, E.S., 2019. Selection of multi-model ensemble of general circulation models for the simulation of precipitation and maximum and minimum temperature based on spatial assessment metrics. Hydrol. Earth Syst. Sci. 23, 4803–4824.

Almazroui, M., Saeed, F., Saeed, S., Islam, M., Ismai, l.M., Kluste, N.A.B., Siddiqui, M., 2020. Projected change in temperature and precipitation over Africa from CMIP6. Earth Syst Environ 4, 455–475.

Arguez, A., Vose, R.S., 2011. The definition of the standard WMO climate normal: the key to deriving alternative climate normals. Bull. Am. Meteorol. Soc. 92, 699–704.

Arias, P.A., Ortega, G., Villegas, L.D., Martínez, J.A., 2021. Colombian Climatology in CMIP5/CMIP6 Models: Persistent Biases and Improvements, 100. Revista Facultad De Ingeniería Universidad De Antioquia, pp. 75–96.

Bannister, D., Herzog, M., Graf, H.-F., Hosking, J.S., Short, C.A., 2017. An assessment of recent and future temperature change over the Sichuan Basin, China, using CMIP5 climate models. J. Clim. 30, 6701–6722.

Bosilovich, M.G., Akella, S., Coy, L., Cullather, R., Draper, C., Gelaro, R., Kovach, R., Liu, Q., Molod, A., Norris, P., et al., 2015. MERRA-2: Initial Evaluation of Climate, p. 139. https://gmao.gsfc.nasa.gov/pubs/docs/Bosilovich803.pdf. (Accessed 5 April 2020). Accessed.

Boucher, O., Servonnat, J., Albright, A.L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., et al., 2020. Presentation and evaluation of the IPSL- CM6A- LR climate model. J. Adv. Model 12 (e2019MS002010), 1–52.

Brown, B.E., Dunne, R.P., Chansang, H., 1996. Coral bleaching relative to elevated seawater temperature in the Andaman Sea (Indian Ocean) over the last 50 years. Coral Reefs 15, 151–152.

Chhin, R., Yoden, S., 2018. Ranking CMIP5 GCMs for model ensemble selection on regional scale: case study of the Indochina Region. J. Geophys. Res. Atmos. 123, 8949–8974.

Crespo, L.R., Keenlyside, N., Koseki, S., 2019. The role of sea surface temperature in the atmospheric seasonal cycle of the equatorial Atlantic. Clim. Dynam. 52, 5927–5946.

Decker, M., Brunke, M.A., Wang, Z., Sakaguchi, K., Zeng, X.B., Bosilovich, M.G., 2012. Evaluation of the reanalysis products from GSFC, NCEP, and ECMWF using flux tower observations. J. Clim. 25, 191–1944.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., et al., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137, 553–597.

Edwards, P.N., 2011. History of climate modeling. Wiley Interdiscip. Rev. Clim. Change 2, 128–139.

Fan, X., Duan, Q., Shen, C., Wu, Y., Xing, C., 2020. Global surface air temperatures in CMIP6: historical performance and future changes. Environ. Res. Lett. 15, 104056.

Fiedler, S., Crueger, T., D'Agostino, R., Peters, K., Becker, T., Leutwyler, D., et al., 2020. Simulated tropical precipitation assessed across three major phases of the Coupled Model Intercomparison Project (CMIP). Mon. Weather Rev. 148, 3653–3680.

Flato, G., Gillett, N., Arora, V., Cannon, A., Anstey, J., 2019. Modelling future climate change. In: Bush, E., Lemmen, D.S. (Eds.), Chapter 3 in Canada's Changing Climate Report. Government of Canada, Ottawa, Ontario, pp. 74–111.

Gampe, D., Schmid, J., Ludwig, R., 2019. Impact of reference dataset selection on RCM evaluation, bias correction, and resulting climate change signals of precipitation. J. Hydrometeorol. 20 (9), 1813–1828.

Gent, R.P., 2012. Coupled climate and earth system models. In: Rasch, P. (Ed.), Climate change modeling methodology. Springer, New York, pp. 5–30.

Gilewski, P., Nawalany, M., 2018. Inter-comparison of rain-gauge, radar, and satellite (IMERG GPM) precipitation estimates performance for rainfall-runoff modeling in a mountainous catchment in Poland. Water 10, 1665.

Gleckler, P.J., Taylor, K.E., 2008. Doutriaux, C. Performance metrics for climate models. J. Geophys. Res. 113, D06104.

Grise, K.M., Davis, S.M., 2020. Hadley cell expansion in CMIP6 models. Atmos. Chem. Phys. 20, 5249–5268.

Grose, M.R., Brown, J.N., Narsey, S., Brown, J.R., Murphy, B.F., Langlais, C., Gupta, A.S., Moise, A.F., Irving, D.B., 2014. Assessment of the CMIP5 global climate model simulations of the western tropical Pacific climate system and comparison to CMIP3. Int. J. Climatol. 34, 3382–3399.

Grose, M.R., Narsey, S., Delage, F.P., Dowdy, A.J., Bador, M., Boschat, G., Chung, C., Kajtar, J.B., Rauniyar, S., Freund, M.B., Lyu, K., Rashid, H., Zhang, X., Wales, S., Trenham, C., Holbrook, N.J., Cowan, T., Alexander, L., Arblaster, J.M., Power, S., 2020. Insights from CMIP6 for Australia's future climate. Earth's Future. 8 e2019EF001469.

Guo, Y., Yu, Y., Lin, P., Liu, H., He, B., Bao, Q., Zhao, S., Wang, X., 2020. Overview of the CMIP6 historical experiment datasets with the climate system model CAS FGOALS-f3-L. Adv. Atmos. Sci. 37, 1057–1066.

Hansen, J., Ruedy, R., Sato, M., Lo, K., 2010. Global surface temperature change. Rev. Geophys. 48, RG4004.

Haraguchi, M., Lall, U., 2015. Flood risks and impacts: a case study of Thailand's floods in 2011 and research questions for supply chain decision making. Int. J. Disaster Risk Reduct. 14, 256–272.

Harris, I., Osborn, T.J., Jones, P., Lister, D., 2020. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. Sci Data 7, 1–18.

He, W., Zhao, S., Wu, Q., Jiang, Y., Wan, S., 2019. Simulating evaluation and projection of the climate zones over China by CMIP5 models. Clim. Dynam. 52, 2597–2612.

Hebeler, F., 2020. RMSE. MATLAB Central File Exchange. https://www.mathworks.com/matlabcentral/Fileexchange/21383-rmse. (Accessed 16 January 2020). Accessed.

Held, I.M., Guo, H., Adcroft, A., Dunne, J.P., Horowitz, L.W., Krasting, J., et al., 2019. Structureand performance of GFDL's CM4.0 climate model. J. Adv. Model. 11, 3691–3727.

Hourdin, F., Musat, I., Bony, S., et al., 2006. The LMDZ4 general circulation model: climate performance and sensitivity to parametrized physics with emphasis on tropical convection. Clim. Dynam. 27, 787–813.

Huang, F., Xu, Z., Guo, W., 2019. Evaluating vector winds in the Asian-Australian monsoon region simulated by 37 CMIP5 models. Clim. Dynam. 53, 491–507.
Hughes, D.A., Mantel, S., Mohobane, T., 2014. An assessment of the skill of downscaled GCM outputs in simulating historical patterns of rainfall variability in South Africa. Nord. Hydrol 45, 134–147.
IPCC, 2014. Annex II: glossary. Mach KJ, Planton S, von Stechow C. In: Pachauri, R.K., Meyer, L.A. (Eds.), Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. in: Core Writing Team. IPCC, Geneva, Switzerland, pp. 117–130.
IPCC, 2021. Summary for policymakers. In: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J.B.R., Maycock, T.K., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B. (Eds.), Climate Change 2021: the Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press. In Press.
Kamworapan, S., Surussavadee, C., 2019. Evaluation of CMIP5 global climate models for simulating climatological temperature and precipitation for Southeast Asia. Adv. Meteorol. 2019, 1–18.
Kim, Y.H., Min, S.K., Zhang, X., Sillmann, J., Sandstad, M., 2020. Evaluation of the CMIP6 multi-model ensemble for climate extreme indices. Weather Clim. Extrem. 29, 100269.
Kumar, S., Merwade, V., Kinter, J.L., Niyogi, D., 2013. Evaluation of temperature and precipitation trends and long-term persistence in CMIP5 twentieth-century climate simulations. J. Clim. 26, 4168–4185.
Le Treut, H., Somerville, R., Cubasch, U., Ding, Y., Mauritzen, C., Mokssit, A., Peterson, T., Prather, M., 2007. Historical overview of climate change. In: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L. (Eds.), Climate Change 2007: the Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
Li, J., Liu, Z., Yao, Z., Wang, R., 2019. Comprehensive assessment of Coupled Model Intercomparison Project Phase 5 global climate models using observed temperature and precipitation over mainland Southeast Asia. Int. J. Climatol. 39, 4139–4153.
Lovino, M.A., Müller, O.V., Berbery, E.H., Müller, G.V., 2018. Evaluation of CMIP5 retrospective simulations of temperature and precipitation in northeastern Argentina. Int. J. Climatol. 38, e1158–e1175.
MathWorks, Inc, 2001. Image Processing Toolbox: for Use with MATLAB. Mathworks Incorporated.
McMahon, T.A., Peel, M.C., Karoly, D.J., 2015. Assessment of precipitation and temperature data from CMIP3 global climate models for hydrologic simulation. Hydrol. Earth Syst. Sci. 19, 361–377.
Miao, C., Duan, Q., Sun, Q., Huang, Y., Kong, D., Yang, T., Ye, A., Di, Z., Gong, W., 2014. Assessment of CMIP5 climate models and projected temperature changes over Northern Eurasia. Environ. Res. Lett. 9, 55007.
Moise, A., Wilson, L., Grose, M., Whetton, P., Watterson, I., Bhend, J., et al., 2015. Evaluation of CMIP3 and CMIP5 models over the Australian region to inform confidence in projections. Aust. Meteorol. Oceanogr. J 65, 19–53.
Morice, C.P., Kennedy, J.J., Rayner, N.A., Jones, P.D., 2012. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 dataset. J. Geophys. Res. Atmos. 117, D08101.
NASA/GISS, 2020. NASA, NOAA Analyses Reveal 2019 Second Warmest Year on Record. https://climate.nasa.gov/vital-signs/global-temperature/%3e%20. (Accessed 1 July 2021). accessed.
Papalexiou, S.M., Rajulapati, C.R., Clark, M.P., Lehner, F., 2020. Robustness of CMIP6 historical global mean temperature simulations: trends, long-term persistence, autocorrelation, and distributional shape. Earth's Future 8 (10), e2020EF001667.
Phongsuwan, N., Chankong, A., Yamarunpatthana, C., Chansang, H., Boonprakob, R., Petchkumnerd, P., Thongtham, N., Paokantha, S., Chanmethakul, T., Panchaiyapoom, P., Bundit, O.A., 2013. Status and changing patterns on coral reefs in Thailand during the last two decades. Deep-Sea Res. Pt. II 96, 19–24.
Radić, V., Clarke, G.K.C., 2011. Evaluation of IPCC models' performance in simulating late- twentieth century climatologies and weather patterns over North America. J. Clim. 24, 5257–5274.
Raghavan, V., Liu, J., Nguyen, N.S., Vu, M.T., Liong, S.-Y., 2018. Assessment of CMIP5 historical simulations of rainfall over Southeast Asia. Theor. Appl. Climatol. 132, 989–1002.
Raju, K.S., Kumar, D.N., 2014. Ranking of global climate models for India using multicriterion analysis. Clim. Res. 60, 103–117.
Raju, K.S., Kumar, D.N., 2015. Ranking general circulation models for India using TOPSIS. J. Water Clim. Change. 6, 288–299.
Raju, K.S., Kumar, D.N., 2020. Review of approaches for selection and ensembling of GCMs. J. Water Clim. Change. 11, 577–599.
Randall, D.A., Wood, R.A., Bony, S., Colman, R., Fichefet, T., Fyfe, J., Kattsov, V., Pitman, A., Shukla, J., Srinivasan, J., Stouffer, R.J., Sumi, A., Taylor, K.E., 2007. Climate models and their evaluation. In: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L. (Eds.), Climate Change 2007: the Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, pp. 589–662.
Reid, P.C., Fischer, A., Lewis-Brown, E., Meredith, M.P., Sparrow, M., Andersson, A.J., Antia, A., Bathmann, U., Beaugrand, G., Brix, H., et al., 2009. Impacts of the oceans on climate change. Adv. Mar. Biol. 56, 1–150.
Rerngnirunsathit, P., 2012. Thailand Country Profiles 2011. Department of Disaster Prevention and Mitigation, Ministry of Interior, Bangkok, Thailand, pp. 1–23. http://www.adrc.asia/countryreport/THA/2011/FY2011B_THA_CR.pdf.
Rienecker, M.M., Suarez, M.J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M.G., Schubert, S.D., Takacs, L., Kim, G.-K., et al., 2011. MERRA: NASA's modern-era retrospective analysis for research and applications. J. Clim. 24, 3624–3648.
Ruan, Y., Liu, Z., Wang, R., Yao, Z., 2019. Assessing the performance of CMIP5 GCMs for projection of future temperature change over the Lower Mekong Basin. Atmosphere 10, 93.
Rupp, D.E., Abatzoglou, J.T., Hegewisch, K.C., Mote, P.W., 2013. Evaluation of CMIP5 20th century climate simulations for the Pacific Northwest USA. J. Geophys. Res. Atmos. 118, 10884–10906.
Séférian, R., Nabat, P., Michou, M., Saint Martin, D., Voldoire, A., Colin, J., Decharme, B., Delire, C., Berthet, S., Chevallier, M., et al., 2019. Evaluation of CNRM Earth- System model, CNRM- ESM2- 1: role of Earth system processes in present- day and future climate. J. Adv. Model. 11, 4182–4227.
Siew, J.H., Tangang, F.T., Juneng, L., 2014. Evaluation of CMIP5 coupled atmosphere–ocean general circulation models and projection of the Southeast Asian winter monsoon in the 21st century. Int. J. Climatol. 34, 2872–2884.
Smith, T.M., Reynolds, R.W., Peterson, T.C., Lawrimore, J., 2008. Improvements to NOAA's historical merged land–ocean surface temperature analysis (1880–2006). J. Clim. 21, 2283–2296.
Stouffer, R.J., Eyring, V., Meehl, G.A., Bony, S., Senior, C., Stevens, B., Taylor, K.E., 2017. CMIP5 scientific gaps and recommendations for CMIP6. Bull. Am. Meteorol. Soc. 98, 95–105.
Su, F., Duan, X., Chen, D., Hao, Z., Cuo, L., 2013. Evaluation of the global climate models in the CMIP5 over the Tibetan Plateau. J. Clim. 26, 3187–3208.
Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., Hsu, K.L., 2018. A review of global precipitation data sets: data sources, estimation, and intercomparisons. Rev. Geophys. 56 (1), 79–107.
Supharatid, S., 2015. Assessment of CMIP3-CMIP5 climate models precipitation projection and implication of flood vulnerability of Bangkok. Am. J. Clim. Change 4, 140–162.
Tan, C.T., Pereira, J.J., 2010. Climate change adaptation: an overview of Southeast Asia. Research Publishing Services, Singapore Asian J. Environ. Disaster Manage 2, 371–395.
Tantrakarnapa, K., 2018. Second Biennial Update Report of Thailand. Office of Natural Resources and Environmental Policy and Planning, Bangkok, Thailand, pp. 1–108.
Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., Sudo, K., Sekiguchi, M., Abe, M., Saito, F., et al., 2019. Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. Geosci. Model Dev. (GMD) 12, 2727–2765.
Taylor, K.E., Stouffer, R.J., Meehl, G.A., 2012. An overview of CMIP5 and the experiment design. Bull. Am. Meteorol. Soc. 93, 485–498.
Trenberth, K.E., Miller, K., Mearns, L., Rhodes, S., 2000. Effects of Changing Climate on Weather and Human Activities.Understanding Global Change: Earth Science and Human Impacts Series, Global Change Instruction Program. UCAR. University Science Books, pp. 1–44.
Trenberth, K.E., Jones, P.D., Ambenje, P., Bojariu, R., Easterling, D., Klein, T., Parker, D., Rahimzadeh, F., Renwick, J.A., Rusticucci, M., Soden, B., Zhai, P., 2007. Observations: surface and atmospheric climate change. In: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M.C., Averyt, K.B., Tignor, M., Miller, H.L. (Eds.), Climate Change 2007. The Physical Science Basis. Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, pp. 235–336.
Trewin, B.C., 2007. The Role of Climatological Normals in a Changing Climate. World Meteorological Organization.
Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., Colin, J., Guérémy, J., Michou, M., Moine, M., et al., 2019. Evaluation of CMIP6 DECK experiments with CNRM-CM6-1. J. Adv. Model. Earth Syst. 11, 2177–2213.
Wargan, K., Labow, G., Frith, S., Pawson, S., Livesey, N., Partyka, G., 2017. Evaluation of the ozone fields in NASA's MERRA-2 reanalysis. J. Clim. 30, 2961–2988.
Watanabe, M., Suzuki, T., O'ishi, R., Komuro, Y., Watanabe, S., Emori, S., Takemura, T., et al., 2010. Improved climate simulation by MIROC5: mean states, variability, and climate sensitivity. J. Clim. 23, 6312–6335.
Watanabe, S., Hirabayashi, Y., Kotsuki, S., Hanasaki, N., Tanaka, K., Mateo, C., Oki, T., 2014. Application of performance metrics for climate models to project future river discharge in Chao Phraya River Basin. Hydrol. Res. Lett. 8, 33–38.
Waugh, D.W., Eyring, V., 2008. Quantitative performance metrics for stratospheric-resolving chemistry-climate models. Atmos. Chem. Phys. 8, 5699.
White, P., Hilario, F.D., de Guzman, R.G., Cinco, T.A., 2009. A Review of Climate Change Model Predictions and Scenario Selection for Impacts on Asian Aquaculture.
Willmott, C.J., Matsuura, K., 2001. Terrestrial Air Temperature and Precipitation: Monthly and Annual Time Series (1950–1999). http://climate.geog.udel.edu/∼climate/html_pages/README.ghcn_ts2.html. (Accessed 30 January 2020). Accessed.
World Meteorological Organization, 2017. WMO Guidelines on the Calculation of Climate Normals (WMO-No. 1203). World Meteorological Organization, Geneva, Switzerland. Retrieved from. https://library.wmo.int/doc_num.php?explnum_id=4166.
Xin, X., Wu, T., Zhang, J., Yao, J., Fang, Y., 2020. Comparison of CMIP6 and CMIP5 simulations of precipitation in China and the East Asian summer monsoon. Int. J. Climatol. 40, 6423–6440.
Xu, J., Gao, Y., Chen, D., Xiao, L., Ou, T., 2017. Evaluation of global climate models for downscaling applications centered over the Tibetan plateau. Int. J. Climatol. 37, 657–671.
Xuan, W., Ma, C., Kang, L., Gu, H., Pan, S., Xu, Y.-P., 2017. Evaluating historical simulations of CMIP5 GCMs for key climatic variables in Zhejiang Province, China. Theor. Appl. Climatol. 128, 207–222.

Yan, G., Wen-Jie, D., Fu-Min, R., Zong-Ci, Z., Jian-Bin, H., 2013. Surface air temperature simulations over China with CMIP5 and CMIP3. Adv. Clim. Change Res. 4 (3), 145–152.
Yan, R., Huang, J., Wang, Y., Gao, J., Qi, L., 2015. Modeling the combined impact of future climate and land use changes on streamflow of Xinjiang Basin, China. Nord. Hydrol 47, 356–372.
Zhao, W., Li, A., 2015. A review on land surface processes modelling over complex terrain. Adv. Meteorol. 2015, 1–17.
Zhao, M., Golaz, J.-C., Held, I.M., Guo, H., Balaji, V., Benson, R., et al., 2018. The GFDL global atmosphere and land model AM4.0/LM4.0: 1. Simulation characteristics with prescribed SSTs. J. Adv. Model. Earth Syst. 10, 691–734.
Zhou, T., Yu, R., 2006. Twentieth-century surface air temperature over China and the globe simulated by coupled climate models. J. Clim. 19, 5843–5858.