

**MIT
Libraries**

| **DSpace@MIT**

MIT Open Access Articles

This is a supplemental file for an item in DSpace@MIT

Item title: Evaluation of CMIP5 Global
Climate Models for Simulating Climatological
Temperature and Precipitation for Southeast Asia

Link back to the item: <https://hdl.handle.net/1721.1/132947>



Massachusetts Institute of Technology

Research Article

Evaluation of CMIP5 Global Climate Models for Simulating Climatological Temperature and Precipitation for Southeast Asia

Suchada Kamworapan ¹ and Chinnawat Surussavadee ^{2,3}

¹Interdisciplinary Graduate School of Earth System Science and Andaman Natural Disaster Management, Prince of Songkla University, Phuket Campus, Phuket 83120, Thailand

²Telecommunications Engineering Department, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand

³Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02141, USA

Correspondence should be addressed to Chinnawat Surussavadee; pop@alum.mit.edu

Received 2 April 2019; Revised 27 August 2019; Accepted 10 September 2019; Published 25 September 2019

Academic Editor: Federico Porcù

Copyright © 2019 Suchada Kamworapan and Chinnawat Surussavadee. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study evaluates the performances of all forty different global climate models (GCMs) that participate in the Coupled Model Intercomparison Project Phase 5 (CMIP5) for simulating climatological temperature and precipitation for Southeast Asia. Historical simulations of climatological temperature and precipitation of the 40 GCMs for the 40-year period of 1960–1999 for both land and sea and those for the century of 1901–1999 for land are evaluated using observation and reanalysis datasets. Nineteen different performance metrics are employed. The results show that the performances of different GCMs vary greatly. CNRM-CM5-2 performs best among the 40 GCMs, where its total error is 3.25 times less than that of GCM performing worst. The performance of CNRM-CM5-2 is compared with those of the ensemble average of all 40 GCMs (40-GCM-Ensemble) and the ensemble average of the 6 best GCMs (6-GCM-Ensemble) for four categories, i.e., temperature only, precipitation only, land only, and sea only. While 40-GCM-Ensemble performs best for temperature, 6-GCM-Ensemble performs best for precipitation. 6-GCM-Ensemble performs best for temperature and precipitation simulations over sea, whereas CNRM-CM5-2 performs best over land. Overall results show that 6-GCM-Ensemble performs best and is followed by CNRM-CM5-2 and 40-GCM-Ensemble, respectively. The total errors of 6-GCM-Ensemble, CNRM-CM5-2, and 40-GCM-Ensemble are 11.84, 13.69, and 14.09, respectively. 6-GCM-Ensemble and CNRM-CM5-2 agree well with observations and can provide useful climate simulations for Southeast Asia. This suggests the use of 6-GCM-Ensemble and CNRM-CM5-2 for climate studies and projections for Southeast Asia.

1. Introduction

Global climate change has been observed and poses a fundamental threat to humanity. The evidence of rapid global climate change includes global temperature rise, warming oceans, shrinking ice sheets, glacial retreat, decreased snow cover, sea level rise, declining Arctic sea ice, more extreme weather events, and ocean acidification [1, 2]. Better understanding of climate change and ability to predict the future climate change and its potential impacts are important for climate change adaptation and mitigation. Since climate change varies from region to region, it affects

different regions of the world differently. Hence, climate change studies for each region in detail are important for increasing resilience of the society to climate change.

Southeast Asia is one of the most vulnerable regions to climate change, where its average temperatures have risen every decade since 1960. A report by Germanwatch on the global climate risk index [3] has listed Vietnam, Myanmar, the Philippines, and Thailand to be among ten countries in the world most affected by climate change during the period of 1997–2016. Vietnam has also been listed by the World Bank to be among five countries most likely to be affected by global warming in the future [4]. According to the Asian

Development Bank, Southeast Asia could suffer bigger losses than most regions in the world [5].

Southeast Asia's climate is tropical [6]. Its weather is mostly hot and humid with high annual precipitation amounts. Precipitation is mostly convective [7]. Intense convective precipitation could cause floods and landslides. Southeast Asia has often been affected by weather-related natural disasters, i.e., floods, droughts, landslides, and tropical cyclones. Since extreme weather events can be intensified by climate change [8], it is essential to understand and accurately project climate change and its impacts to the region. To accomplish this, a climate model that can provide useful climate simulations and projections for Southeast Asia is required.

Climate models are important tools for understanding and predicting the complex Earth's climate. Several global climate models (GCMs) have been developed by several research centers around the world. Forty GCMs from 20 research groups have participated in the Coupled Model Intercomparison Project Phase 5 (CMIP5) [9]. Their global climate simulations and projections are publicly available. Since there are many GCMs that can be used and they could perform differently for different regions of the world, the main objective of this study is to evaluate the performance of these GCMs in order to find GCMs that perform well in Southeast Asia and that should be employed for climate simulations and projections in the region.

Previous studies have evaluated the performances of these GCMs to be used for specific regions, e.g., eastern Tibetan Plateau [10], Australia [11], US Pacific Northwest [12], northeastern Argentina [13], northern Eurasia [14], US continental areas [15], and Southeast Asia [16, 17]. Since climate is different for different regions and GCMs also perform differently for different regions, results for different regions cannot be directly compared. Despite the importance of climate change studies for Southeast Asia as it is one of the most vulnerable regions to climate change, there are only few previous studies [16, 17] that evaluate the performances of CMIP5 GCMs in the region. Raghavan et al. [16] have evaluated the performance of CMIP5 GCMs for Southeast Asia with the focus on only historical precipitation simulations for 20 years of 1986–2005 without considering temperature simulations. Although results from [16] have shown that there is no particular model performing well for climatological precipitation simulations in Southeast Asia, it only evaluates 10 GCMs out of the total 40 CMIP5 GCMs and employs not many performance metrics.

Even though our preliminary study [17] has evaluated all 40 CMIP5 GCMs for climate simulations for Southeast Asia, it does not address detailed results for each performance metric and does not consider the performances of ensemble averages of different GCMs. The performances of all 40 CMIP5 GCMs for simulating climatological temperature and precipitation in the twentieth century are evaluated in further details in this study using observation and reanalysis datasets, where 19 performance metrics are employed for evaluation. The performance of the best GCM is also compared with the of ensemble averages of different GCMs.

Section 2 describes the research methodology employed in this study, which includes the study area, GCMs, observation and reanalysis datasets, and performance metrics. Section 3 presents the evaluation results. Section 4 summarizes and concludes the paper.

2. Research Methodology

2.1. Study Area. Southeast Asia is a subregion of Asia and consists of 2 main portions, i.e., the mainland and a string of archipelagoes to the south and east of the mainland. Southeast Asia is composed of 11 sovereign states, including Brunei Darussalam, Cambodia, East Timor, Indonesia, Laos, Malaysia, Myanmar, the Philippines, Singapore, Thailand, and Vietnam. Figure 1 shows topography (m) above the mean sea level of the study area, where the Shuttle Radar Topography Mission digital elevation model (SRTM) [18] with the spatial resolution of 90 m is employed. The study area covers latitudes from 12.75°S to 24.25°N and longitudes from 88.25°E to 144.75°E.

Southeast Asia lies in the tropics with tropical climate [6]. Since the incident angle of solar radiation is small, the temperature is generally hot and does not fluctuate much throughout the year. Employing observation and reanalysis datasets used in this study, where their details will be described later in Section 2.3, shows that the mean annual temperature for years 1960–1999 for the study area is 26.38°C. The 40-year monthly average temperatures for January–December for the study area are ranged from the minimum of 25.10°C, which occurs in January, to the maximum of 27.30°C, which occurs in May. The entire region is strongly affected by the southwest and northeast monsoons [19], which are due to differences in land and sea temperatures caused by solar radiation. The southwest monsoon is typically from late May to September. It particularly affects Thailand and Myanmar and causes the rainy season to be in the period. The northeast monsoon is typically from November to March. It brings relatively dry and cool air and little precipitation to the mainland and causes rain in the southern part of Southeast Asia in the period. Southeast Asia receives considerable annual precipitation. Employing observation and reanalysis datasets used in this study, where their details will be described later in Section 2.3, shows that the mean annual precipitation for years 1960–1999 for the study area is ~2,034.25 mm. The mean annual precipitation for each of the 40 y ranged from 1,810.20 mm in 1972 to 2,336.72 mm in 1996. Most precipitation in this region is strongly driven by convection [7]. Southeast Asia is often affected by weather-related natural disasters. The Philippines and Vietnam are often affected by tropical cyclones. Hence, climate change studies for this region are crucial.

2.2. Global Climate Models. The Coupled Model Intercomparison Project Phase 5 (CMIP5) [9] is a collaborative effort with the aim to improve the climate change knowledge. CMIP5 involves 20 climate modeling research groups around the world with 40 GCMs. CMIP5 outputs include

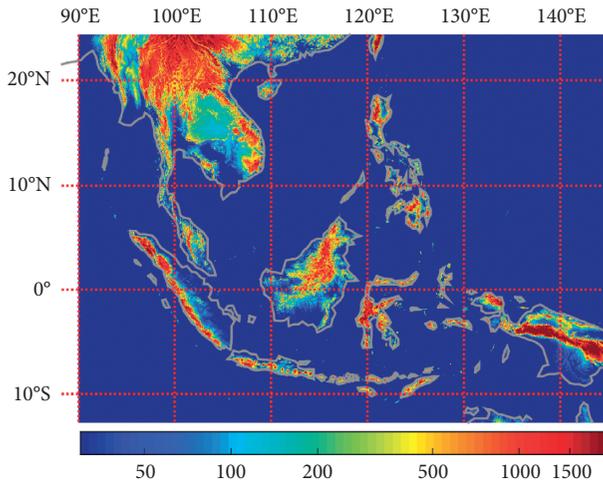


FIGURE 1: Topography (m) above the mean sea level of Southeast Asia, the study area.

historical climate simulations for years 1850–2005 and climate projections for near term (out to about 2035) and long term (out to 2100 and beyond) by considering 4 Representative Concentration Pathways (RCPs). To evaluate the performances of these GCMs, their simulated climatological temperature and precipitation for years 1901–1999 are employed.

The forty GCMs evaluated in this study together with their spatial resolutions and numbers of ensemble members driven by different initial conditions are shown in Table 1. For GCMs with more than one ensemble members, the average of all ensemble members is employed for evaluation. The GCM outputs employed in this study include monthly averages of near-surface air temperature, daily-minimum near-surface air temperature, daily-maximum near-surface air temperature, and surface precipitation.

2.3. Observation and Reanalysis Datasets. Two observation datasets and two reanalysis datasets are employed in this study to evaluate GCMs and are listed in Table 2. The two observation datasets include the University of Delaware Air Temperature and Precipitation (UD) version 3.01 [20] and the University of East Anglia Climatic Research Unit (CRU) TS3.10.01 [21]. The two reanalysis datasets are the National Centers for Environmental Prediction (NCEP)-National Center for Atmospheric Research (NCAR) 40-Year Reanalysis [22], which will be later called NCEP and the European Center for Medium-Range Weather Forecasts 40-Year Reanalysis (ERA40) [23].

UD global monthly temperature and precipitation data are on a $0.5^\circ \times 0.5^\circ$ grid and are available for years 1901–2010. UD is produced using observations from the Global Historical Climate Network and the archive of Legates and Willmott. CRU global monthly temperature and precipitation data are on a $0.5^\circ \times 0.5^\circ$ grid and are available for years 1901–2009. CRU is produced using observations from the National Meteorological Services and other external agents. Both UD and CRU are available over land only.

ERA40 monthly reanalysis is produced using a data assimilation system employing many sources of observations including radiosondes, balloons, aircraft, buoys, satellites, and scatterometers. It is available on a $2.5^\circ \times 2.5^\circ$ grid for 45 years from 1957 to 2002. NCEP monthly reanalysis is produced using a data assimilation system employing many sources of observations including land surface measurements, ships, rawinsonde, pibals, aircrafts, satellites, and other data. It is on a $1.9^\circ \times 1.9^\circ$ grid and is available from 1948 to 2012. NCEP and ERA40 are available for both land and sea.

Different observation datasets are different among themselves due to different original observations and methods employed. Although NCEP and ERA40 reanalysis datasets are produced using numerical models with observation assimilation, several studies have employed them for evaluating historical climate simulations [11, 12, 24–26]. Figure 2 compares average annual temperatures ($^\circ\text{C}$) for years 1960–1999 of CRU, UD, NCEP, and ERA40. It shows obvious differences among all datasets both in terms of value and resolution. Although CRU and UD are both observations, they are significantly different. To evaluate GCM temperature simulations, averages of CRU, UD, NCEP, and ERA40 temperature are employed. Since only CRU and NCEP provide monthly averages of daily-minimum and daily-maximum near-surface air temperature, their averages are employed for these parameters. Figure 3 compares average annual precipitation (mm) for years 1960–1999 of CRU, UD, NCEP, and ERA40. Precipitation from these datasets are also obviously different. Since ERA40 precipitation is significantly lower than others, it is not employed in this study. To evaluate GCM precipitation simulations, averages of CRU, UD, and NCEP precipitation are employed.

2.4. Performance Metrics. Several performance metrics for evaluating the performances of GCMs have been proposed. Most performance metrics employed in this study are from [12] and [27], where the root mean squared error (RMSE) is added. The performance metrics and time periods used for computing each performance metric are listed in Table 3.

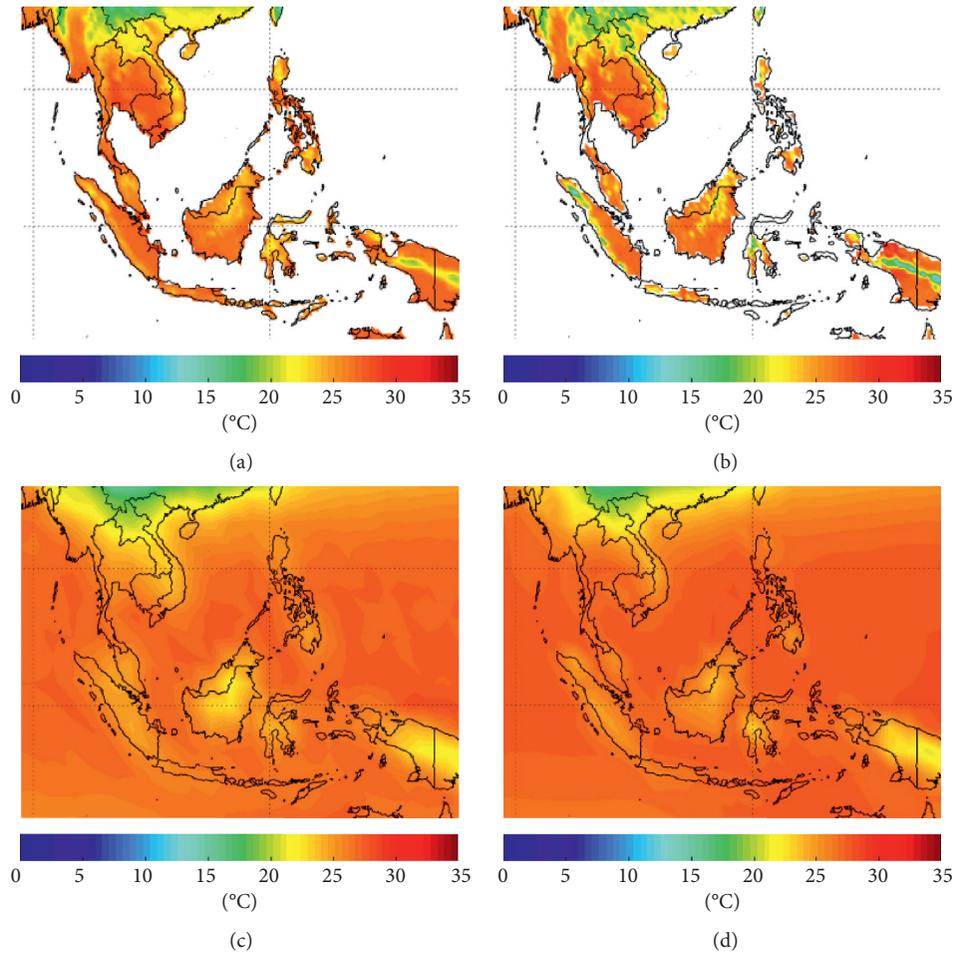
There are 19 performance metrics employed in this study. Eleven performance metrics are computed for land only, sea only, and both land and sea for the 40-year period of 1960–1999 when UD and CRU are available. The eleven performance metrics include (1) mean annual temperature (Mean-T), (2) mean annual precipitation (Mean-P), (3) mean diurnal temperature range (MDTR-MMM), where MMM designates a season, (4) mean seasonal cycle amplitude of temperature (Season-Amp-T) defined as the temperature difference between warmest and coldest months, (5) mean seasonal cycle amplitude of precipitation (Season-Amp-P) defined as the precipitation difference between wettest and driest months, (6) correlation coefficient between simulated and observed mean temperatures (Cor-MMM-T), (7) correlation coefficient between simulated and observed mean precipitation (Cor-MMM-P), (8) standard deviation of mean temperature

TABLE 1: Global climate models employed in this study together with their resolutions and numbers of ensemble members.

GCM	Research center	Resolution lon. \times lat.	Number of ensemble members
BCC-CSM1-1	Beijing Climate Center, China Meteorological Administration, China	2.8×2.8	3
BCC-CSM1-1-M	Beijing Climate Center, China Meteorological Administration, China	1.12×1.12	3
BNU-ESM	College of Global Change and Earth System Science, Beijing Normal University, China	2.8×2.8	1
CanESM2	Canadian Center for Climate Modeling and Analysis, Canada	2.8×2.8	5
CCSM4	National Center of Atmospheric Research, USA	1.25×0.94	6
CESM1-BGC	Community Earth System Model Contributors, USA	1.25×0.94	1
CESM1-CAM5	Community Earth System Model Contributors, USA	1.25×0.94	3
CESM1-FASTCHEM	Community Earth System Model Contributors, USA	1.25×0.94	3
CESM1-WACCM	Community Earth System Model Contributors, USA	2.5×1.89	4
CMCC-CESM	Centro Euro-Mediterraneo per I Cambiamenti Climatici, Italy	3.75×7.71	1
CMCC-CM	Centro Euro-Mediterraneo per I Cambiamenti Climatici, Italy	0.75×0.75	1
CMCC-CMS	Centro Euro-Mediterraneo per I Cambiamenti Climatici, Italy	1.88×1.87	1
CNRM-CM5	National Center of Meteorological Research, France	1.4×1.4	10
CNRM-CM5-2	National Center of Meteorological Research, France	1.4×1.4	4
CSIRO-Mk3-6-0	Commonwealth Scientific and Industrial Research Organization/Queensland Climate Change Center of Excellence, Australia	1.8×1.8	10
EC-EARTH	EC-EARTH consortium, The Netherlands/Ireland	1.13×1.12	14
FGOALS-g2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, China	2.8×2.8	5
FIO-ESM	The First Institute of Oceanography, SOA, China	2.81×2.79	3
GFDL-CM3	NOAA Geophysical Fluid Dynamics Laboratory, USA	2.5×2.0	5
GFDL-ESM2G	NOAA Geophysical Fluid Dynamics Laboratory, USA	2.5×2.0	1
GFDL-ESM2M	NOAA Geophysical Fluid Dynamics Laboratory, USA	2.5×2.0	1
GISS-E2-H	NASA Goddard Institute for Space Studies, USA	2.5×2.0	6
GISS-E2-H-CC	NASA Goddard Institute for Space Studies, USA	2.5×2.0	1
GISS-E2-R	NASA Goddard Institute for Space Studies, USA	2.5×2.0	6
GISS-E2-R-CC	NASA Goddard Institute for Space Studies, USA	2.5×2.0	1
HadCM3	Met Office Hadley Center, UK	3.75×2.5	10
HadGEM2-AO	Met Office Hadley Center, UK	1.88×1.25	1
HadGEM2-CC	Met Office Hadley Center, UK	1.88×1.25	3
HadGEM2-ES	Met Office Hadley Center, UK	1.88×1.25	4
INMCM4	Institute for Numerical Mathematics, Russia	2.0×1.5	1
IPSL-CM5A-LR	Institut Pierre Simon Laplace, France	3.75×1.8	6
IPSL-CM5A-MR	Institut Pierre Simon Laplace, France	2.5×1.25	3
IPSL-CM5B-LR	Institut Pierre Simon Laplace, France	3.75×1.8	1
MIROC5	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology, Japan	1.4×1.4	1
MIROC-ESM	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies, Japan	2.8×2.8	3
MIROC-ESM-CHEM	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies, Japan	2.8×2.8	3
MPI-ESM-LR	Max Planck Institute for Meteorology, Germany	1.88×1.87	3
MPI-ESM-MR	Max Planck Institute for Meteorology, Germany	1.88×1.87	3
MRI-CGCM3	Meteorological Research Institute, Japan	1.1×1.1	5
NorESM1-M	Norwegian Climate Center, Norway	2.5×1.9	3

TABLE 2: Observation and reanalysis datasets employed in this study.

Dataset	Research center	Resolution	Availability
CRU TS3.10.01	University of East Anglia Climatic Research Unit	$0.5^\circ \times 0.5^\circ$	1901–2009
UD v.3.01	University of Delaware Air Temperature and Precipitation v.3.01	$0.5^\circ \times 0.5^\circ$	1901–2010
NCEP	National Center for Environmental Prediction/ National Center for Atmospheric Research reanalysis	$\sim 1.9^\circ \times 1.9^\circ$	1948–2012
ERA40	European Center for Medium-Range Weather Forecasts 40-Year Reanalysis	$\sim 2.5^\circ \times 2.5^\circ$	Mid-1957 to mid-2002

FIGURE 2: Comparison of the mean annual temperatures ($^\circ\text{C}$) for years 1960–1999 of (a) CRU, (b) UD, (c) NCEP, and (d) ERA40.

(STD-MMM-T), (9) standard deviation of mean precipitation (STD-MMM-P), (10) root mean squared error of mean temperature (RMSE-MMM-T), and (11) root mean squared error of mean precipitation (RMSE-MMM-P). MDTR-MMM, Cor-MMM-T, Cor-MMM-P, STD-MMM-T, STD-MMM-P, RMSE-MMM-T, and RMSE-MMM-P are computed separately for 3 seasons, including the hot season from February to April, the rainy season from May to October, and the cold season from November to January. To compute these 11 metrics, 40-year averages for individual pixels are computed first and are then averaged for all pixels. For example, to compute Mean-T, 40-year mean annual temperatures for individual pixels are computed first and are then averaged to get Mean-T.

Mean-T and Mean-P are employed to evaluate the biases in model simulated temperatures and precipitation, respectively. MDTR-MMM is employed to evaluate performances of the models to simulate differences in seasonal daily-maximum and daily-minimum near-surface air temperatures and is computed using monthly averages of daily-maximum and daily-minimum near-surface air temperature provided by each GCM, CRU, and NCEP. Season-Amp-T is employed to evaluate performances of the models to simulate temperature differences between warmest and coldest months. Season-Amp-P is employed to evaluate performances of the models to simulate precipitation differences between wettest and driest months. Cor-MMM-T and Cor-MMM-P are employed to evaluate performances of the

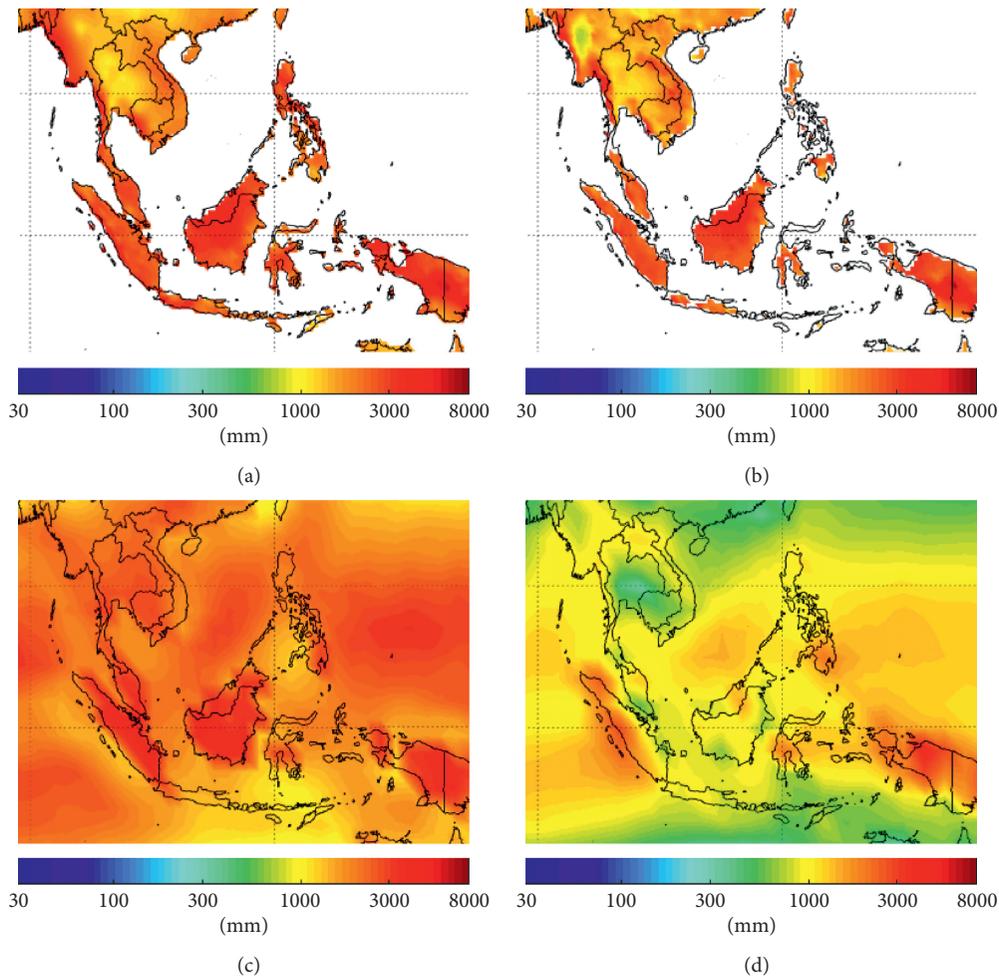


FIGURE 3: Comparison of the mean annual precipitation (mm) for years 1960–1999 of (a) CRU, (b) UD, (c) NCEP, and (d) ERA40.

models to simulate the spatial patterns of temperature and precipitation, respectively. STD-MMM-T and STD-MMM-P are employed to evaluate performances of the models to simulate the spatial variations of temperature and precipitation, respectively. RMSE-MMM-T and RMSE-MMM-P are employed to evaluate performances of the models to simulate the values of temperature and precipitation, respectively.

The other eight performance metrics evaluate the long-term performance of simulated climatological temperature and precipitation for the 99-year period of 1901–1999. Due to the availability of observations, they are computed for land only and include (1) variance of annual average temperature (Var-T) defined as the variance of 99-y annual average temperatures, (2) coefficient of variation of annual precipitation (CV-P) defined as the mean-normalized standard deviation of 99-year annual precipitation, (3) root mean squared error of annual average temperature (RMSE-T), (4) root mean squared error of annual precipitation (RMSE-P), (5) linear trend of annual average temperature (Trend-T) defined as the slope of the best-fit linear line for the time series of 99-year annual average temperature, (6) linear trend of annual precipitation (Trend-P) defined as the slope of the best-fit linear line for the time series of 99-year

annual precipitation, (7) correlation coefficient of the cold-season temperature mean and Niño3.4 index (ENSO-T), and (8) correlation coefficient of the cold-season precipitation and Niño3.4 index (ENSO-P). To compute these 8 metrics, the values for individual pixels for 99 year are computed first. Then, values for all pixels are averaged. For example, to compute Trend-T, the slope of the linear line that best fits annual average temperature values for 99 year for each pixel is first computed. The slopes for all pixels in the study area are then averaged to get Trend-T.

Var-T and CV-P are employed to evaluate performances of the models to simulate the 99-year variations of temperatures and precipitation, respectively. RMSE-T and RMSE-P are employed to evaluate performances of the models to simulate the values of temperature and precipitation, respectively. Trend-T and Trend-P are employed to evaluate performances of the models to simulate the 99-year trends of temperature and precipitation, respectively. Since the Niño3.4 index measures anomalies of sea surface temperatures in the east-central tropical Pacific, ENSO-T is employed to evaluate performances of the models to simulate the linear relationships between anomalies of sea surface temperatures in the east-central tropical Pacific and

TABLE 3: List of performance metrics, observation datasets, and time period employed for computing each performance metric.

Metric	Observation dataset	Time period
Mean-T-Land	CRU, UD, ERA40, NCEP	1960–1999
Mean-T-Sea	ERA40, NCEP	
Mean-T-Land-Sea	CRU, UD, ERA40, NCEP	
Mean-P-Land	CRU, UD, NCEP	1960–1999
Mean-P-Sea	NCEP	
Mean-P-Land-Sea	CRU, UD, NCEP	
MDTR-MMM-Land	CRU, NCEP	1960–1999
MDTR-MMM-Sea	NCEP	
MDTR-MMM-Land-Sea	CRU, NCEP	
Season-Amp-T-Land	CRU, UD, ERA40, NCEP	1960–1999
Season-Amp-T-Sea	ERA40, NCEP	
Season-Amp-T-Land-Sea	CRU, UD, ERA40, NCEP	
Season-Amp-P-Land	CRU, UD, NCEP	1960–1999
Season-Amp-P-Sea	NCEP	
Season-Amp-P-Land-Sea	CRU, UD, NCEP	
Cor-MMM-T-Land	CRU, UD, ERA40, NCEP	1960–1999
Cor-MMM-T-Sea	ERA40, NCEP	
Cor-MMM-T-Land-Sea	CRU, UD, ERA40, NCEP	
Cor-MMM-P-Land	CRU, UD, NCEP	1960–1999
Cor-MMM-P-Sea	NCEP	
Cor-MMM-P-Land-Sea	CRU, UD, NCEP	
STD-MMM-T-Land	CRU, UD, ERA40, NCEP	1960–1999
STD-MMM-T-Sea	ERA40, NCEP	
STD-MMM-T-Land-Sea	CRU, UD, ERA40, NCEP	
STD-MMM-P-Land	CRU, UD, NCEP	1960–1999
STD-MMM-P-Sea	NCEP	
STD-MMM-P-Land-Sea	CRU, UD, NCEP	
RMSE-MMM-T-Land	CRU, UD, ERA40, NCEP	1960–1999
RMSE-MMM-T-Sea	ERA40, NCEP	
RMSE-MMM-T-Land-Sea	CRU, UD, ERA40, NCEP	
RMSE-MMM-P-Land	CRU, UD, NCEP	1960–1999
RMSE-MMM-P-Sea	NCEP	
RMSE-MMM-P-Land-Sea	CRU, UD, NCEP	
Var-T-Land	CRU, UD	1901–1999
CV-P-Land	CRU, UD	1901–1999
RMSE-T-Land	CRU, UD	1901–1999
RMSE-P-Land	CRU, UD	1901–1999
Trend-T-Land	CRU, UD	1901–1999
Trend-P-Land	CRU, UD	1901–1999
ENSO-T-Land	CRU, UD	1901–1999
ENSO-P-Land	CRU, UD	1901–1999

MMM is the season designation. FMA: February, March, and April; MJJASO: May, June, July, August, September, and October; NDJ: November, December, and January.

temperatures in the study area. ENSO-P is employed to evaluate performances of the models to simulate the linear relationships between anomalies of sea surface temperatures in the east-central tropical Pacific and precipitation in the study area.

2.5. Evaluation Method. Since spatial resolutions and grid locations of different GCMs and observation and reanalysis datasets are different, they are bilinearly interpolated into the same $0.15^\circ \times 0.15^\circ$ grid covering the study area. To evaluate each performance metric, averages of observation and reanalysis datasets listed in Table 3 are employed.

From all 19 performance metrics listed in Table 3, the absolute error $A_{i,j}$ for each performance metric i and each global climate model j is computed as $A_{i,j} = |O_i - S_{i,j}|$, where O_i and $S_{i,j}$ are the performance metric i of observations and simulated performance metric i of the global climate model j , respectively. Due to different magnitude scales of different performance metrics, the relative error for each performance metric i and each global climate model j , i.e., $R_{i,j} = (A_{i,j} - A_{i,\min}) / (A_{i,\max} - A_{i,\min})$, is used, where $A_{i,\min}$ and $A_{i,\max}$ are minimum and maximum absolute errors for each performance metric i , respectively. The total error of a global climate model j is computed as $E_j = \sum_{i=1}^n R_{i,j}$, where n is the total number of performance metrics.

3. Results

3.1. Performances of 40 Global Climate Models. Figure 4 shows relative errors for all performance metrics for the 40 GCMs. The GCMs are listed on the left of the figure in the order of lowest to highest total errors from top to bottom, respectively. Relative errors are very different for different GCMs. Each GCM has mixed results for different performance metrics. Comparison of relative errors of 40 GCMs for all performance metrics obviously shows that CNRM-CM5-2 performs best.

Figure 5 shows total errors of the 40 GCMs computed using all performance metrics. Total errors of different GCMs vary considerably. This emphasizes the need for this study to find GCMs that can provide useful climatological temperature and precipitation for Southeast Asia. The six GCMs that perform best are CNRM-CM5-2, CNRM-CM5, BNU-ESM, CESM1-BGC, CESM-CAM5, and CCSM4, respectively. GISS-E2-R performs worst, where its total error is ~ 3.25 times higher than that of CNRM-CM5-2.

GCMs are further evaluated for 4 different categories, including temperature only, precipitation only, land only, and sea only, where only performance metrics for each category are considered. The six best GCMs for each category are listed in Table 4. The numbers shown in Table 4 are total errors for different categories. Results show that each GCM performs differently for different categories. For example, although CNRM-CM5 performs best and second best for sea and temperature, respectively, it performs sixth for precipitation. CCSM4 performs second best for precipitation, but it performs sixth for sea and worse than sixth for temperature and

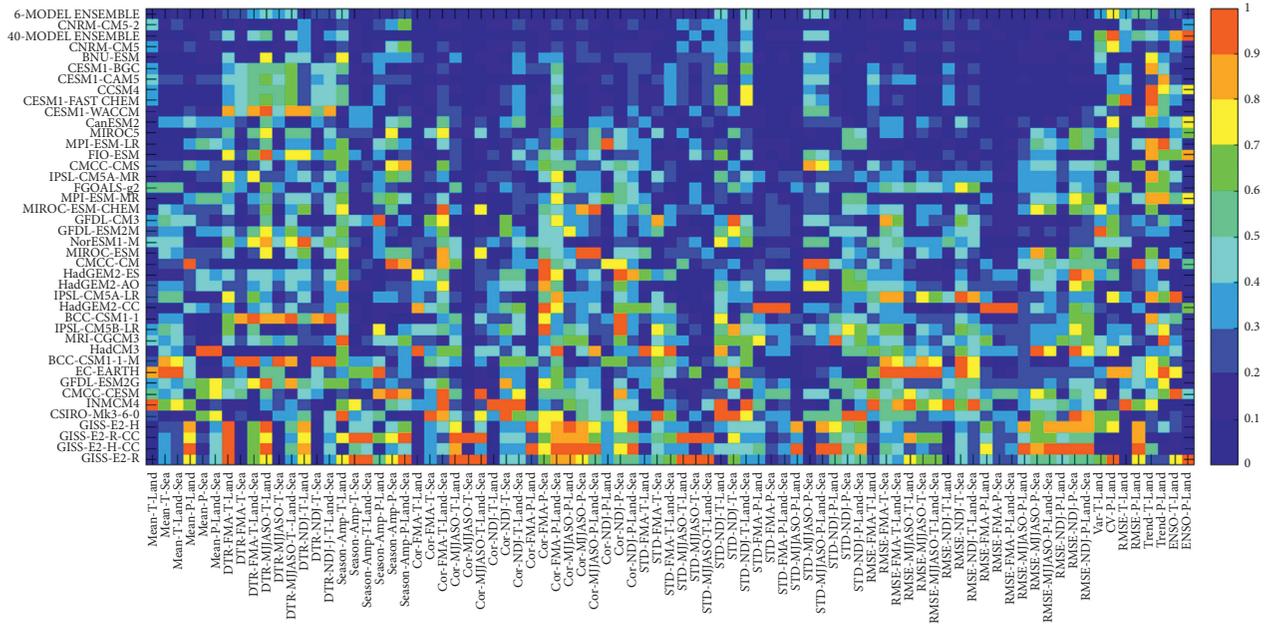


FIGURE 4: Relative error for each performance metric listed at the bottom for 40 GCMs, 6-GCM-Ensemble, and 40-GCM-Ensemble listed at the left in the order of lowest to highest total errors from top to bottom, respectively.

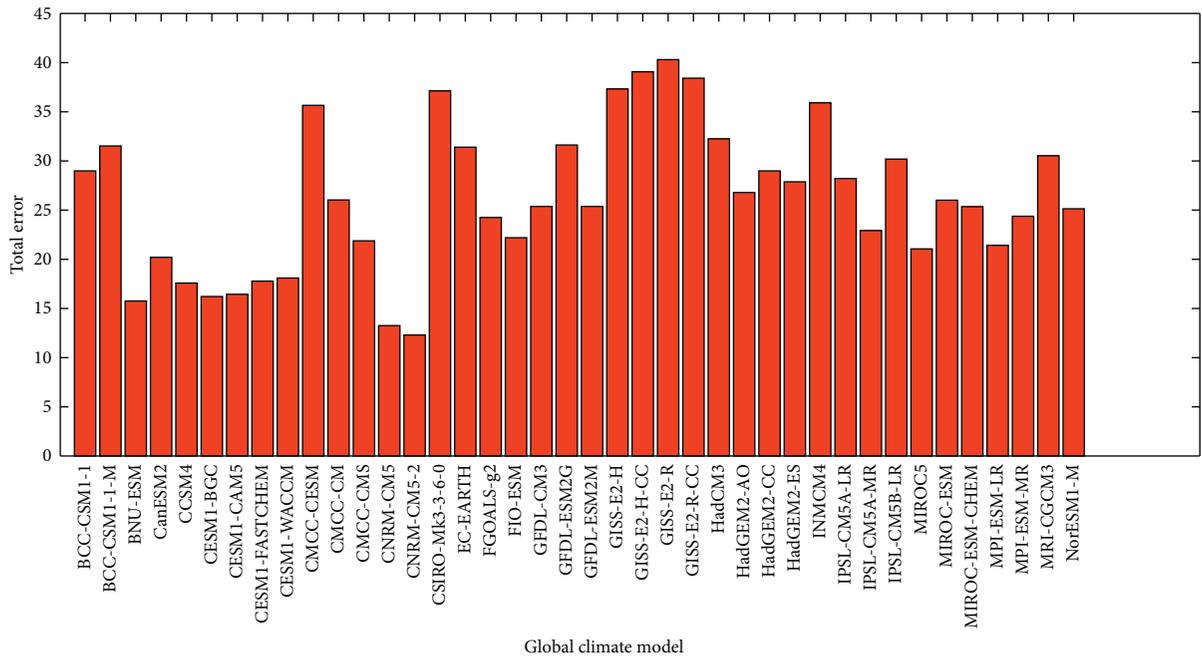


FIGURE 5: Total errors of 40 GCMs.

land. CNRM-CM5-2 is the only GCM that is in the top three for all categories. The best GCMs for temperature, precipitation, land, and sea are CNRM-CM5-2, CESM1-BGC, CNRM-CM5-2, and CNRM-CM5, respectively. CNRM-CM5 is the second best for temperature, and its total error is close to that of CNRM-CM5-2. The performances of the top six GCMs for simulations over sea are not that different. When all categories are considered, the total error of the second best

GCM, i.e., CNRM-CM5, is only 7.44% higher than that of CNRM-CM5-2, but then the total error of the third best jumps to 24.17% higher than that of CNRM-CM5-2.

Since there is no single GCM that performs best for all categories, overall performance of an ensemble average of different GCMs that perform well in each category could be better than that of a single GCM. It is observed from Table 4 that all top two GCMs for individual categories are within

TABLE 4: List of the six global climate models performing best for different categories, including temperature only, precipitation only, land only, and sea only.

Rank	Temperature only	Precipitation only	Land only	Sea only	All
1	CNRM-CM5-2 (7.82)	CESM1-BGC (3.60)	CNRM-CM5-2 (5.65)	CNRM-CM5 (3.12)	CNRM-CM5-2 (12.37)
2	CNRM-CM5 (7.95)	CCSM4 (4.18)	CNRM-CM5 (7.08)	BNU-ESM (3.21)	CNRM-CM5 (13.29)
3	MPI-ESM-LR (8.12)	CNRM-CM5-2 (4.54)	MIROC5 (7.15)	CNRM-CM5-2 (3.31)	BNU-ESM (15.36)
4	CMCC-CMS (8.93)	CESM1-FASTCHEM (4.99)	CESM1-CAM5 (7.40)	CESM1-BGC (3.47)	CESM1-BGC (16.18)
5	BNU-ESM (9.38)	CESM-CAM5 (5.30)	CESM1-BGC (7.59)	CESM1-FASTCHEM (3.59)	CESM-CAM5 (16.36)
6	MPI-ESM-MR (9.47)	CNRM-CM5 (5.34)	CanESM2 (7.92)	CCSM4 (3.60)	CCSM4 (17.69)

Numbers in parenthesis are total errors for different categories.

top six when results for all categories are combined. Hence, the next three sections compare the performance of the single best GCM, i.e., CNRM-CM5-2, with those of the ensemble average of the six best GCMs based on the total error, which will be called 6-GCM-Ensemble, and the ensemble average of all 40 GCMs, which will be called 40-GCM-Ensemble, for temperature, precipitation, and overall simulations, respectively.

3.2. Performances of CNRM-CM5-2 and GCM Ensembles for Simulating Temperature. Figure 6 compares the mean annual temperatures (Mean-Ts) ($^{\circ}\text{C}$) for years 1960–1999 of observations, CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble. They agree well with observations both in terms of temperature values and patterns. They all show temperature gradient from south to north of the mainland and lower temperature for highly elevated areas. All models are biased slightly lower than observations, where the mean errors (MEs; $E[\text{model}-\text{observations}]$) of CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble are -0.56 , -0.33 , and -0.18°C , respectively. Historical temperature simulations by averaging all 40 GCMs are least biased.

The mean diurnal temperature ranges (MDTRs) ($^{\circ}\text{C}$) of observations, CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble are 3.10, 3.02, 2.69, and 2.61°C , respectively, for the hot season, are 2.64, 2.43, 2.07, and 2.12°C , respectively, for the rainy season, and are 2.69, 2.77, 2.42, and 2.37°C , respectively, for the cold season. Results show that CNRM-CM5-2 performs best and is the best to provide the information about daily temperature variation for all seasons.

Figure 7 compares the mean seasonal cycle amplitudes of temperature (Season-Amp-Ts) ($^{\circ}\text{C}$) for years 1960–1999 of observations, CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble. Overall Season-Amp-T values and patterns of all models and observations agree well. The main difference is over the northern part of the mainland, where Season-Amp-T of observations is lower than that of all models. When the mean of Season-Amp-T for all pixels in the study area is computed, Season-Amp-T for observations, CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble are 2.20, 2.22, 2.23, and 2.19°C , respectively. All models perform comparably well in providing the temperature difference between warmest and coldest months.

Figure 8 compares average seasonal temperatures for years 1960–1999 of observations, CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble separately for hot

(FMA), rainy (MJJASO), and cold (NDJ) seasons. Temperature values and patterns of all models agree well with observations. Table 5 shows correlation coefficients (CCs) between observations and model simulations, standard deviations (STDs) normalized by the standard deviation of observations, and root mean squared errors (RMSEs) of average seasonal temperature for the three seasons for years 1960–1999 of CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble. Bold face shows the best value for each performance metric. CCs of all models are high and are almost the same for all seasons. Simulated seasonal temperatures of all models are strongly correlated with observations. Since STD presented in Table 5 is normalized by the standard deviation of observations, the number closest to 1.0 is the best. When the three seasons are considered, 40-GCM-Ensemble performs best in terms of STD, although CNRM-CM5-2 has the best STD for the hot season. 40-GCM-Ensemble also has the lowest RMSEs for all seasons and is followed by 6-GCM-Ensemble and CNRM-CM5-2, respectively.

There are 4 performance metrics employed for evaluating long-term temperature simulations for years 1901–1999. The 99-year standard deviations of annual average temperature (STD-Ts) for observations, CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble are 0.31, 0.34, 0.22, and 0.16, respectively. CNRM-CM5-2 agrees best with observations and is followed by 6-GCM-Ensemble and 40-GCM-Ensemble, respectively. The 99-year root mean squared errors of annual average temperature (RMSE-T) ($^{\circ}\text{C}$) of CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble are not much different and are 1.56, 1.46, and 1.44°C , respectively. The 99-year linear trends of annual average temperature (Trend-Ts) ($^{\circ}\text{C century}^{-1}$) of observations, CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble are 0.16, 0.34, 0.55, and 0.45, respectively. CNRM-CM5-2 is the best to provide the rate of change in long-term annual average temperatures in the study area.

Figure 9 compares the correlation coefficients of cold-season temperature and Niño3.4 index (ENSO-Ts) for years 1901–1999 of observations, CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble. The results show that CNRM-CM5-2 agrees best with observations and is the best to tell how well anomalies of sea surface temperatures in the east-central tropical Pacific represented by the Niño3.4 index are correlated with temperatures in the study area. When ENSO-Ts are averaged for all pixels in the study area, they are 0.28, 0.19, 0.40, and 0.69 for observations, CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble,

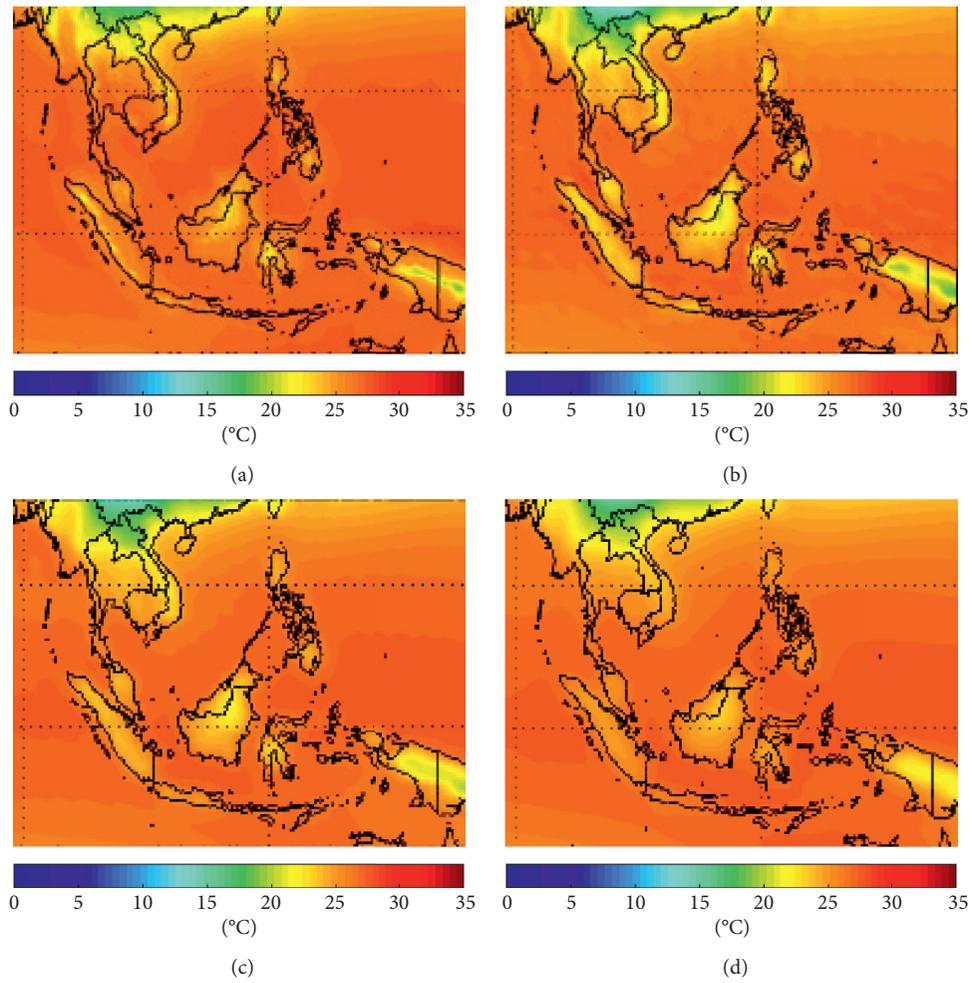


FIGURE 6: Comparisons of the mean annual temperatures (Mean-Ts) (°C) for years 1960–1999 of (a) observations, (b) CNRM-CM5-2, (c) 6-GCM-Ensemble, and (d) 40-GCM-Ensemble.

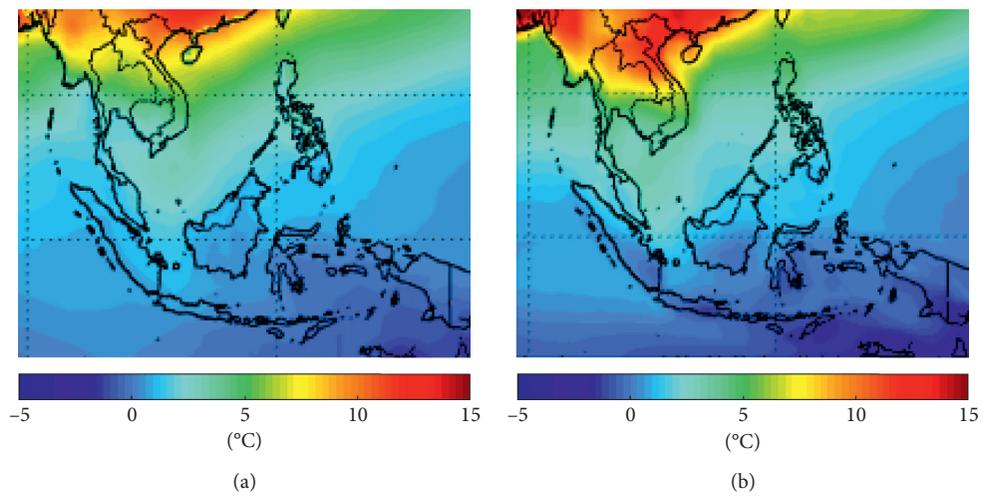


FIGURE 7: Continued.

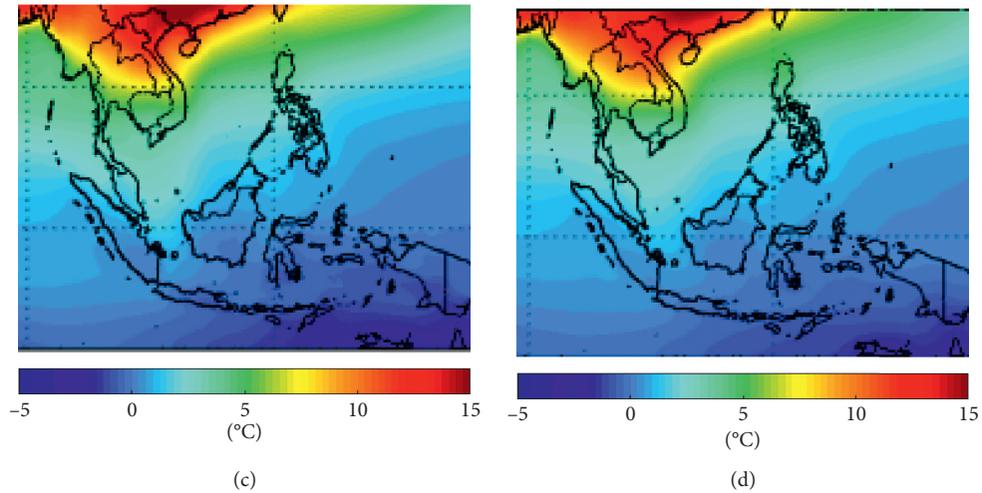


FIGURE 7: Comparisons of the mean seasonal cycle amplitudes of temperature (Season-Amp-Ts) for years 1960–1999 of (a) observations, (b) CNRM-CM5-2, (c) 6-GCM-Ensemble, and (d) 40-GCM-Ensemble.

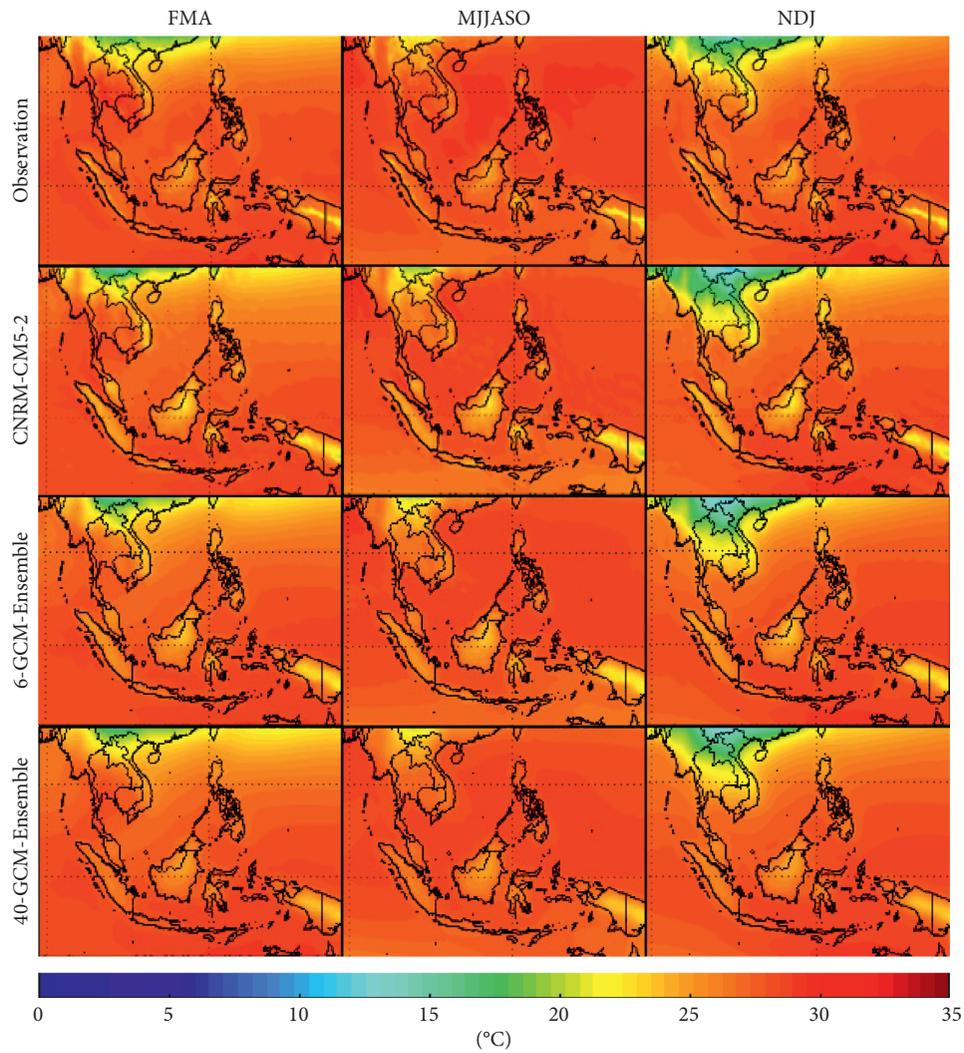


FIGURE 8: Comparisons of average seasonal temperatures (°C) for years 1960–1999 of observations, CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble, from top to bottom rows, respectively. Columns from left to right show hot (FMA), rainy (MJJASO), and cold (NDJ) seasons, respectively.

TABLE 5: Correlation coefficients (CCs) between observations and model simulations, standard deviations (STDs) normalized by the standard deviation of observations, and root mean squared errors (RMSEs) of average seasonal temperature for hot (FMA), rainy (MJJASO), and cold (NDJ) seasons for years 1960–1999 of CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble.

Model	CC			STD			RMSE		
	Hot	Rainy	Cold	Hot	Rainy	Cold	Hot	Rainy	Cold
CNRM-CM5-2	0.93	0.91	0.96	2.18	1.59	3.25	0.86	0.90	1.13
6-GCM-Ensemble	0.95	0.92	0.97	2.28	1.41	3.28	0.75	0.60	0.96
40-GCM-Ensemble	0.95	0.91	0.97	2.21	1.19	3.11	0.67	0.54	0.78

Bold face represents the best value for each performance metric.

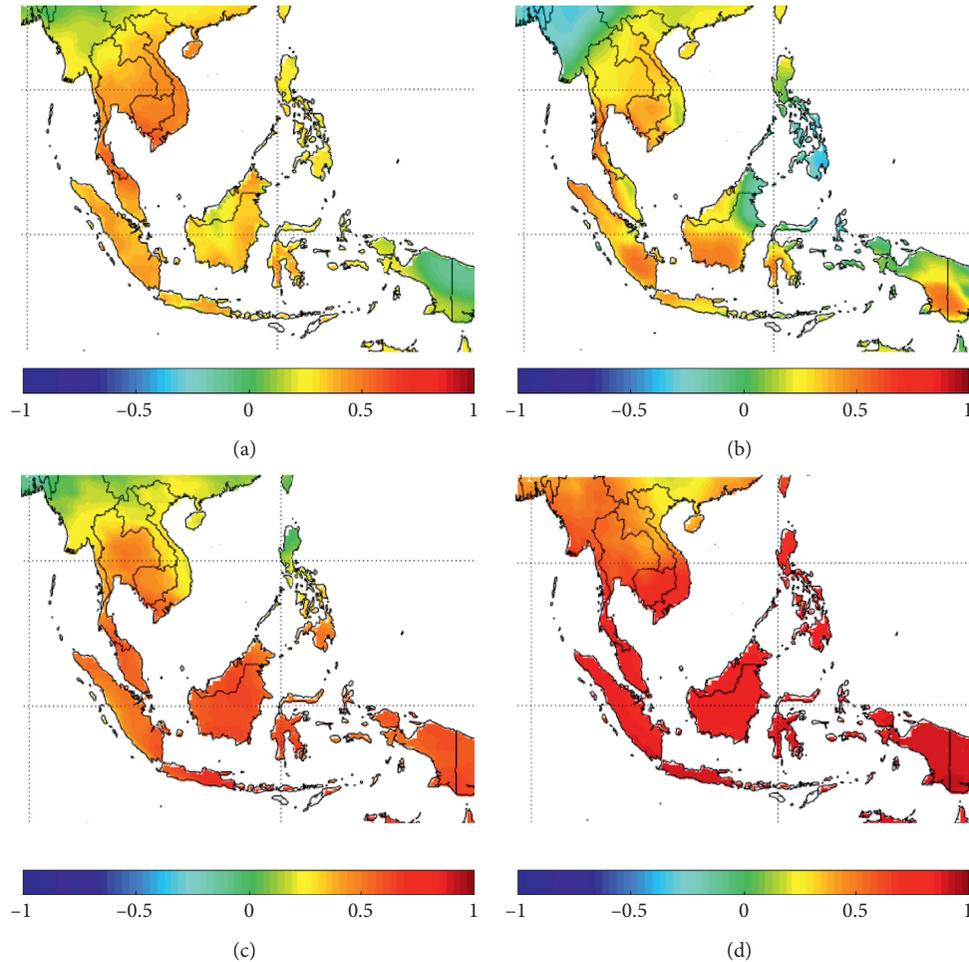


FIGURE 9: Correlation coefficients of cold-season temperature and Niño3.4 index (ENSO-Ts) for years 1901–1999 of (a) observations, (b) CNRM-CM5-2, (c) 6-GCM-Ensemble, and (d) 40-GCM-Ensemble.

respectively. Although ENSO-T of 6-GCM-Ensemble agrees well with that of observations over the mainland, it is obviously higher over archipelagoes to the south and east of the mainland, particularly for Papua New Guinea. ENSO-T of 40-GCM-Ensemble is significantly higher than that of observations for most of the study area.

3.3. Performances of CNRM-CM5-2 and GCM Ensembles for Simulating Precipitation. Figure 10 compares the mean annual precipitation (Mean-P) ($\text{mm}\cdot\text{y}^{-1}$) for years 1960–1999 of observations, CNRM-CM5-2, 6-GCM-Ensemble,

and 40-GCM-Ensemble. Overall precipitation patterns of all models agree well with observations. CNRM-CM5-2 and 6-GCM-Ensemble obviously have higher precipitation over high mountains in Papua New Guinea. MEs ($E[\text{model} - \text{observations}]$) of CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble are 158.83 , 167.77 , and $277.57 \text{ mm}\cdot\text{y}^{-1}$, respectively. Although CNRM-CM5-2 simulated mean annual precipitation has some wavy patterns, it is obviously the least biased.

Figure 11 compares the mean seasonal cycle amplitudes of precipitation (Season-Amp-Ps) (%) for years 1960–1999 of observations, CNRM-CM5-2, 6-GCM-Ensemble, and 40-

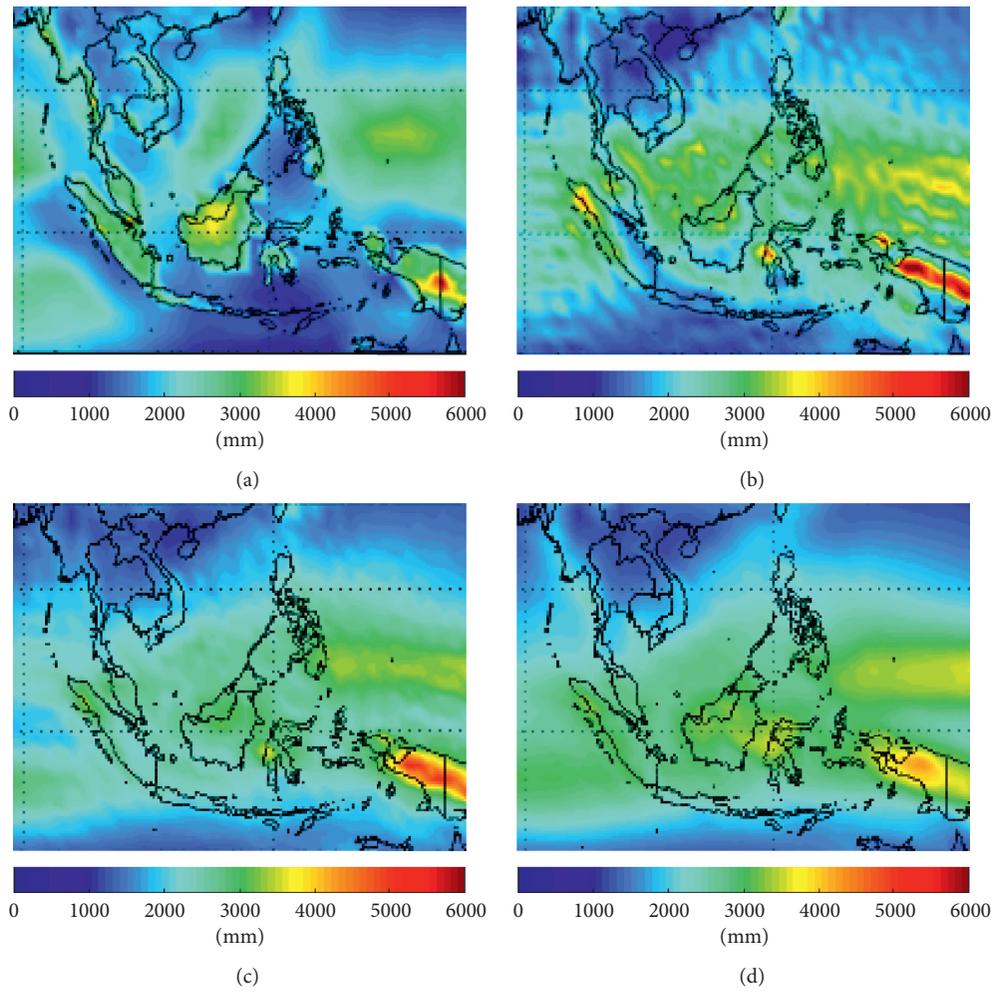


FIGURE 10: Comparisons of the mean annual precipitation (Mean-P) ($\text{mm}\cdot\text{y}^{-1}$) for years 1960–1999 of (a) observations, (b) CNRM-CM5-2, (c) 6-GCM-Ensemble, and (d) 40-GCM-Ensemble.

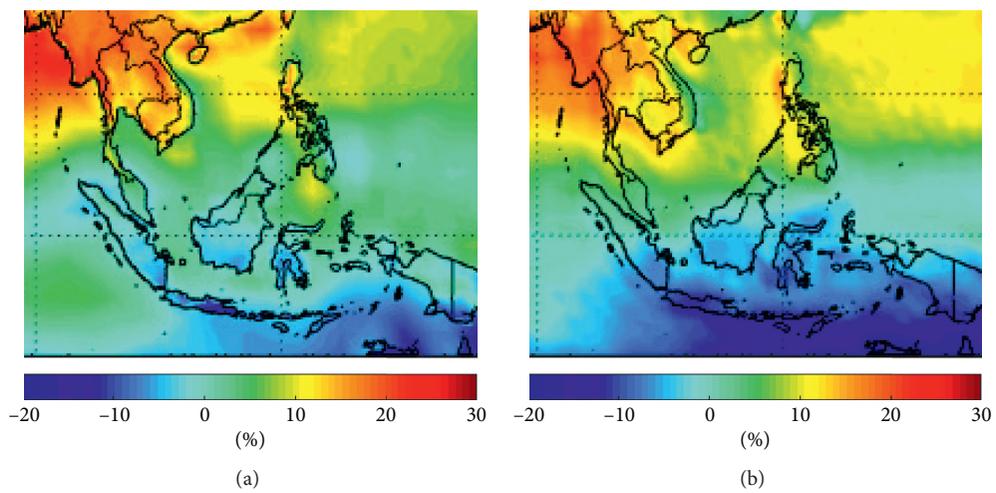


FIGURE 11: Continued.

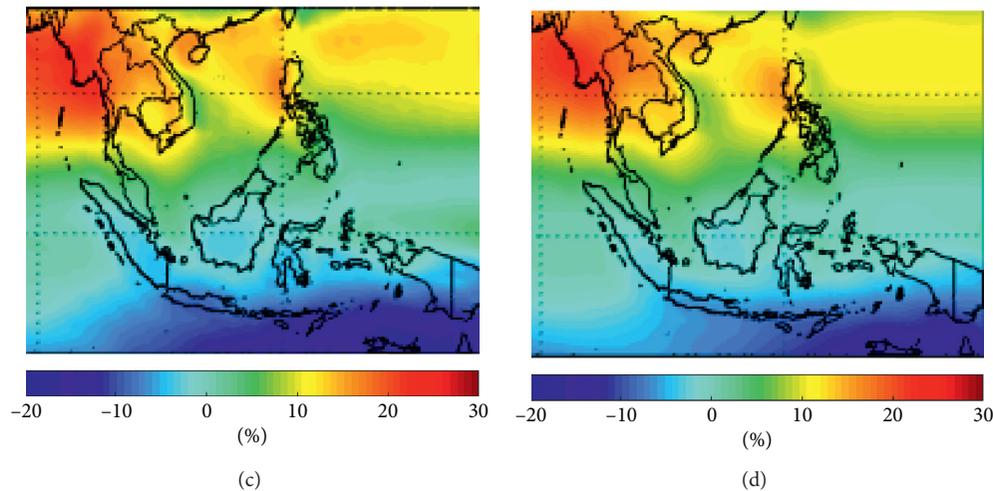


FIGURE 11: Comparisons of the mean seasonal cycle amplitudes of precipitation (Season-Amp-Ps) for years 1960–1999 of (a) observations, (b) CNRM-CM5-2, (c) 6-GCM-Ensemble, and (d) 40-GCM-Ensemble.

GCM-Ensemble. Season-Amp-P is calculated as percentage of the mean annual precipitation. Overall Season-Amp-P values and patterns of all models and observations agree well. When the mean of Season-Amp-P for all pixels in the study area is computed for each model and observations, Season-Amp-Ps for observations, CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble are 4.10, 2.53, 3.51, and 3.97%, respectively. 40-GCM-Ensemble agrees best with observations and is the best to provide difference in precipitation in wettest and driest months.

Figure 12 compares average seasonal precipitation for years 1960–1999 of observations, CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble for hot (FMA), rainy (MJJASO), and cold (NDJ) seasons. Overall seasonal precipitation values and patterns of observations and simulations of all models agree well. The main discrepancies for the average seasonal precipitation in the hot season include the following: (1) all models have higher precipitation than observations over high mountains in Papua New Guinea, and (2) 40-GCM-Ensemble has obvious higher precipitation over the areas in the lower half of the figure than other models and observations. The main discrepancies for the average seasonal precipitation in the rainy season include the following: (1) observations have higher precipitation than all models along the west coast of the mainland, and (2) CNRM-CM5-2 has higher precipitation than observations over the sea southern of the mainland. The main discrepancy for the average seasonal precipitation in the cold season is that CNRM-CM5-2 and 6-GCM-Ensemble have higher precipitation over high mountains in Papua New Guinea. CNRM-CM5-2 has wavy patterns for all three seasons.

Table 6 shows CCs between observations and model simulations, STDs normalized by the standard deviation of observations, which means that the best is the closest to 1.0, and RMSEs of average seasonal precipitation for hot (FMA), rainy (MJJASO), and cold (NDJ) seasons for years 1960–1999 of CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble. All models are highly correlated with observations for all seasons. CCs of 6-GCM-Ensemble are the

highest for all seasons and are very close of those of 40-GCM-Ensemble. When the three seasons are considered, 6-GCM-Ensemble performs best in terms of STD although CNRM-CM5-2 has the best STD for the rainy season. 6-GCM-Ensemble obviously has the lowest RMSEs for all seasons and is followed by 40-GCM-Ensemble and CNRM-CM5-2, respectively.

There are 4 performance metrics for evaluating long-term performance of precipitation simulations for years 1901–1999. The 99-year coefficients of variation of annual precipitation (CV-P) of observations, CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble are 0.13, 0.14, 0.05, and 0.02, respectively. CNRM-CM5-2 performs best and its CV-P almost equals to that of observations. The 99-year root mean squared errors of annual precipitation (RMSE-Ps) (mm) of CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble are 246.03, 590.37, and 548.92 mm, respectively. CNRM-CM5-2 performs much better than 6-GCM-Ensemble and 40-GCM-Ensemble. The 99-year linear trend of annual precipitation (Trend-P) ($\% \text{ century}^{-1}$) of observations, 40 GCMs, 6-GCM-Ensemble, and 40-GCM-Ensemble are 5.26, 4.17, 3.70, and 2.16, respectively. CNRM-CM5-2 is the best to provide the rate of change in long-term annual precipitation in the study area.

Figure 13 compares the correlation coefficients of cold-season precipitation and Niño3.4 index (ENSO-Ps) for years 1901–1999 of observations, CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble. Results show that CNRM-CM5-2 agrees best with observations and is the best to tell how well anomalies of sea surface temperatures in the east-central tropical Pacific represented by the Niño3.4 index are correlated with cold-season precipitation in the study area. When ENSO-Ps are averaged for all pixels in the study area, they are -0.10 , 0.19 , -0.20 , and 0.69 for observations, CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble, respectively. ENSO-T of 6-GCM-Ensemble is obviously lower than that of observations for the lower part of the mainland and is obviously higher than that of observations over archipelagoes to the south and east of the mainland.

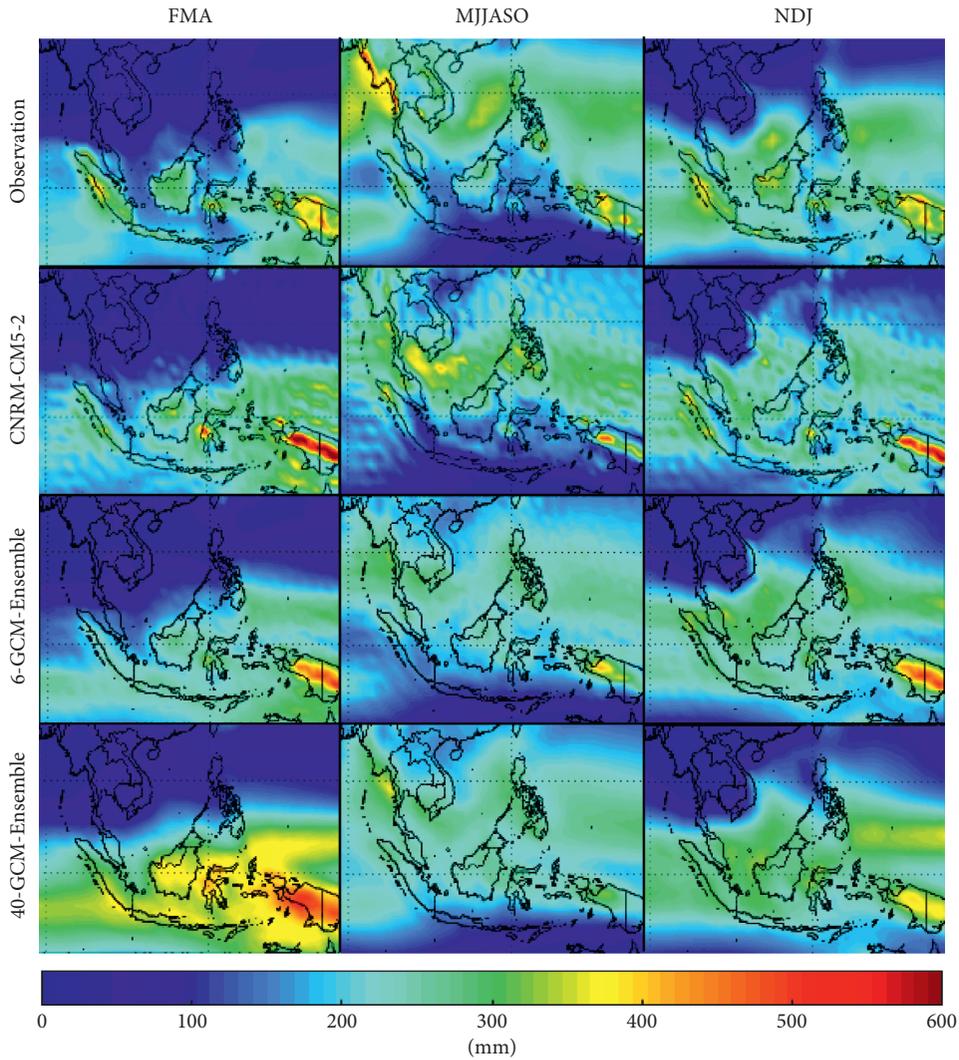


FIGURE 12: Comparisons of average seasonal precipitation (mm season^{-1}) for years 1960–1999 of observations, CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble, from top to bottom rows, respectively. Columns from left to right show hot (FMA), rainy (MJJASO), and cold (NDJ) seasons, respectively.

TABLE 6: Correlation coefficients (CCs) between observations and model simulations, standard deviations (STDs) normalized by the standard deviation of observations, and root mean squared errors (RMSEs) of average seasonal precipitation for hot (FMA), rainy (MJJASO), and cold (NDJ) seasons for years 1960–1999 of CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble.

Model	CC			STD			RMSE		
	Hot	Rainy	Cold	Hot	Rainy	Cold	Hot	Rainy	Cold
CNRM-CM5-2	0.85	0.73	0.88	1.10	0.95	0.90	46.04	60.44	48.76
6-GCM-Ensemble	0.92	0.84	0.90	1.07	0.75	0.97	37.36	46.61	37.98
40-GCM-Ensemble	0.91	0.81	0.90	1.23	0.74	1.05	47.72	47.25	38.97

Bold face represents the best value for each performance metric.

ENSO-P of 40-GCM-Ensemble is significantly higher than that of observations for all of the study area.

3.4. Overall Performances of CNRM-CM5-2 and GCM Ensembles. This section evaluates overall performances of CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble. When all performance metrics are considered, Figure 4 also shows relative errors of 6-GCM-Ensemble and 40-

GCM-Ensemble for each performance metric. It shows that 6-GCM-Ensemble performs best and is followed by CNRM-CM5-2, and 40-GCM-Ensemble, respectively. Although 40-GCM-Ensemble performs worst among the three models, it performs better than the rest of single GCMs.

The performances of CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble for temperature-only, precipitation-only, land-only, and sea-only categories are

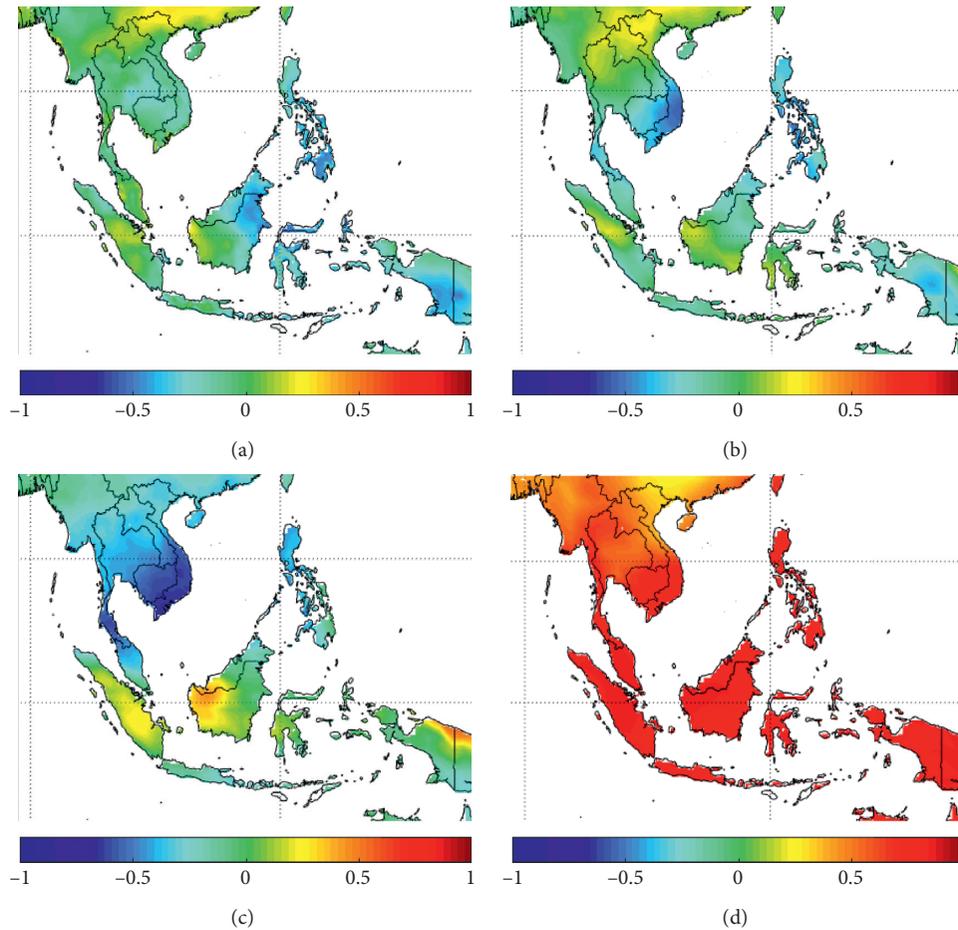


FIGURE 13: Correlation coefficients of cold-season precipitation and Niño3.4 index (ENSO-Ps) for years 1901–1999 of (a) observations, (b) CNRM-CM5-2, (c) 6-GCM-Ensemble, and (d) 40-GCM-Ensemble.

compared in Table 7. Numbers in the table are total errors for different categories. The best models for temperature-only, precipitation-only, land-only, and sea-only categories are 40-GCM-Ensemble, 6-GCM-Ensemble, CNRM-CM5-2, and 6-GCM-Ensemble, respectively. When all categories are considered, 6-GCM-Ensemble performs best, and overall total errors of CNRM-CM5-2 and 40-GCM-Ensemble are 15.63 and 19.00% higher than that of 6-GCM-Ensemble, respectively. The performance of 6-GCM-Ensemble for temperature simulations is close to that of 40-GCM-Ensemble, as the total error of 6-GCM-Ensemble is only 3.12% higher.

4. Summary and Conclusion

The performances for simulating climatological temperature and precipitation for Southeast Asia of 40 CMIP5 GCMs are evaluated using observation and reanalysis datasets for both land and sea for the 40-year period of 1960–1999 and for land for the 99-year period of 1901–1999. Nineteen different performance metrics are employed, where the sum of relative errors of all performance metrics is used to evaluate each GCM. Results are also subdivided into 4 different categories, including temperature only, precipitation only,

land only, and sea only. Since averaging different GCMs could improve the simulation performance, the performance of the best GCM is compared with those of the ensemble averages of the 6 best GCMs called 6-GCM-Ensemble and the ensemble averages of all 40 GCMs called 40-GCM-Ensemble.

The performances of the 40 GCMs are very different, where the total error of the worst GCM is ~ 3.25 times higher than that of the best GCM. This emphasizes the need of this study to find GCMs that can provide useful climate simulations for Southeast Asia. When all performance metrics are considered, CNRM-CM5-2 has the lowest total error among all 40 GCMs. Although there is no GCM performing best for all categories, CNRM-CM5-2 is the only GCM that is in the top three for all categories. The top two GCMs for each category are within the top six when all categories are considered.

Comparisons of CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble show that when all categories are combined, 6-GCM-Ensemble performs best and is followed by CNRM-CM5-2 and 40-GCM-Ensemble, respectively. The total errors of CNRM-CM5-2 and 40-GCM-Ensemble are 15.63 and 19.00% higher than that of 6-GCM-Ensemble, respectively. The 40-GCM-Ensemble, 6-GCM-Ensemble, CNRM-CM5-2, and 6-GCM-Ensemble perform best for

TABLE 7: Performance ranks of CNRM-CM5-2, 6-GCM-Ensemble, and 40-GCM-Ensemble for different categories, including temperature only, precipitation only, land only, and sea only.

Rank	Temperature only	Precipitation only	Land only	Sea only	All
1	40-GCM-Ensemble (7.69)	6-GCM-Ensemble (4.15)	CNRM-CM5-2 (5.90)	6-GCM-Ensemble (2.32)	6-GCM-Ensemble (11.84)
2	6-GCM-Ensemble (7.93)	CNRM-CM5-2 (5.38)	6-GCM-Ensemble (6.36)	40-GCM-Ensemble (3.31)	CNRM-CM5-2 (13.69)
3	CNRM-CM5-2 (8.31)	40-GCM-Ensemble (6.16)	40-GCM-Ensemble (8.03)	CNRM-CM5-2 (3.41)	40-GCM-Ensemble (14.09)

Numbers in parenthesis are total errors for different categories.

temperature-only, precipitation-only, land-only, and sea-only categories, respectively. Although 6-GCM-Ensemble performs second best for temperature simulations, its total error is only 3.12% higher than that of 40-GCM-Ensemble.

Detailed comparisons of 6-GCM-Ensemble and CNRM-CM5-2 simulations with observations for each performance metric show that their simulations agree well with observations. Results in this study lead to different conclusions from that found in the previous study [16], which only evaluates 10 GCMs out of the total of 40 CMIP5 GCMs and focuses only on precipitation for a relatively short term of 1986–2005. Although results from [16] show that no model performs well for climatological precipitation simulations in Southeast Asia, five out of six best GCMs found in this study are not evaluated in [16].

This study finds that 6-GCM-Ensemble and CNRM-CM5-2 can provide useful simulated climatological temperature and precipitation for Southeast Asia. This suggests the use of 6-GCM-Ensemble and CNRM-CM5-2 for climate simulations and projections for Southeast Asia. There is a tradeoff to be considered between using 6-GCM-Ensemble and CNRM-CM5-2. Although 6-GCM-Ensemble is 15.63% more accurate, using the averages of 6 GCMs will involve ~6 times more amount of data than using a single GCM and hence will require more time and computational resources, particularly for complex applications of these models, e.g., the use of GCM outputs as inputs for mesoscale models for dynamical downscaling in order to obtain climate simulations and projections at high resolution [28–31].

Data Availability

CMIP5 data employed in this study are available at the Program Climate Model Diagnosis and Intercomparison (PCMDI) website (<http://pcmdi9.llnl.gov/>). Observations and reanalysis are publicly available.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the Interdisciplinary Graduate School of Earth System Science and Andaman Natural Disaster Management of the Prince of Songkla University, Phuket Campus, Thailand.

References

- [1] T. F. Stocker, D. Qin, G.-K. Plattner et al., *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, IPCC, Cambridge University Press, Cambridge, UK, 2013.
- [2] C. B. Field, V. R. Barros, D. J. Dokken et al., *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, IPCC, Cambridge University Press, Cambridge, UK, 2014.
- [3] D. Eckstein, V. Künzel, and L. Schäfer, “Global climate risk index 2018. Who suffers most from extreme weather events?,” in *Weather-Related Loss Events in 2016 and 1997 to 2016*, Germanwatch, Bonn, Germany, 2017.
- [4] World Bank and GFDRL, *Vulnerability, Risk Reduction, and Adaptation to Climate Change—Vietnam*, World Bank, Washington, DC, USA, 2017.
- [5] D. A. Raitzer, F. Bosello, M. Tavoni et al., *Southeast Asia and the Economics of Global Climate Stabilization*, Asian Development Bank, Mandaluyong, Philippines, 2015.
- [6] M. C. Peel, B. L. Finlayson, and T. A. McMahon, “Updated world map of the Köppen-Geiger climate classification,” *Hydrology and Earth System Sciences*, vol. 11, no. 5, pp. 1633–1644, 2007.
- [7] C. Surussavadee, “Evaluation of high-resolution tropical weather forecasts using satellite passive millimeter-wave observations,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2780–2787, 2014.
- [8] M. K. Tippett, “Extreme weather and climate,” *npj Climate and Atmospheric Science*, vol. 1, no. 1, p. 45, 2018.
- [9] K. E. Taylor, R. J. Stouffer, and G. A. Meehl, “An overview of CMIP5 and the experiment design,” *Bulletin of the American Meteorological Society*, vol. 93, no. 4, pp. 485–498, 2012.
- [10] F. Su, X. Duan, D. Chen, Z. Hao, and L. Cuo, “Evaluation of the global climate models in the CMIP5 over the Tibetan plateau,” *Journal of Climate*, vol. 26, no. 10, pp. 3187–3208, 2013.
- [11] A. Moise, L. Wilson, M. Grose et al., “Evaluation of CMIP3 and CMIP5 models over the Australian region to inform confidence in projections,” *Australian Meteorological and Oceanographic Journal*, vol. 65, no. 1, pp. 19–53, 2015.
- [12] D. E. Rupp, J. T. Abatzoglou, K. C. Hegewisch, and P. W. Mote, “Evaluation of CMIP5 20th century climate simulations for the Pacific Northwest USA,” *Journal of Geophysical Research: Atmospheres*, vol. 118, no. 19, pp. 10884–10906, 2013.
- [13] M. A. Lovino, O. V. Müller, E. H. Berbery, and G. V. Müller, “Evaluation of CMIP5 retrospective simulations of

- temperature and precipitation in northeastern Argentina,” *International Journal of Climatology*, vol. 38, pp. e1158–e1175, 2018.
- [14] C. Miao, Q. Duan, Q. Sun et al., “Assessment of CMIP5 climate models and projected temperature changes over northern Eurasia,” *Environmental Research Letters*, vol. 9, no. 5, Article ID 055007, 2014.
- [15] S. Kumar, V. Merwade, J. L. Kinter, and D. Niyogi, “Evaluation of temperature and precipitation trends and long-term persistence in CMIP5 20th century climate simulations,” *Journal of Climate*, vol. 26, no. 12, pp. 4168–4185, 2013.
- [16] S. V. Raghavan, J. Liu, N. S. Nguyen, M. T. Vu, and S.-Y. Liong, “Assessment of CMIP5 historical simulations of rainfall over Southeast Asia,” *Theoretical and Applied Climatology*, vol. 132, no. 3–4, pp. 989–1002, 2018.
- [17] S. Kamworapan and C. Surussavadee, “Performance of cmip5 global climate models for climate simulation in southeast Asia,” in *Proceedings of the 2017 IEEE Region 10 Conference (TENCON)*, pp. 718–722, Institute of Electrical and Electronics Engineers, Penang, Malaysia, November 2017.
- [18] A. Jarvis, H. I. Reuter, A. Nelson, and E. Guevara, *Hole-filled Seamless SRTM Data V4*, International Centre for Tropical Agriculture (CIAT), Cali, Colombia, 2008, <http://srtm.csi.cgiar.org>.
- [19] Y. Y. Loo, L. Billa, and A. Singh, “Effect of climate change on seasonal monsoon in Asia and its impact on the variability of monsoon rainfall in southeast Asia,” *Geoscience Frontiers*, vol. 6, no. 6, pp. 817–823, 2015.
- [20] K. Matsuura and C. J. Willmot, “Terrestrial air temperature 1900–2010 gridded monthly time series (version 3.01),” 2016, http://climate.geog.udel.edu/Eclimate/html_pages/Global2011/README.GlobalTsT2011.html.
- [21] I. Harris, P. D. Jones, T. J. Osborn, and D. H. Lister, “Updated high-resolution grids of monthly climatic observations—the CRU TS3.10 Dataset,” *International Journal of Climatology*, vol. 34, no. 3, pp. 623–642, 2014.
- [22] E. Kalnay, M. Kanamitsu, R. Kistler et al., “The NCEP/NCAR 40-year reanalysis project,” *Bulletin of the American Meteorological Society*, vol. 77, no. 3, pp. 437–472, 1996.
- [23] S. M. Uppala, P. W. Kållberg, A. J. Simmons et al., “The ERA-40 re-analysis,” *Quarterly Journal of the Royal Meteorological Society*, vol. 131, no. 612, pp. 2961–3012, 2005.
- [24] C. F. McSweeney, R. G. Jones, R. W. Lee, and D. P. Rowell, “Selecting CMIP5 GCMs for downscaling over multiple regions,” *Climate Dynamics*, vol. 44, no. 11–12, pp. 3237–3260, 2015.
- [25] P. J. Gleckler, K. E. Taylor, and C. Doutriaux, “Performance metrics for climate models,” *Journal of Geophysical Research*, vol. 113, no. D6, p. D06104, 2008.
- [26] J. Xu, Y. Gao, D. Chen, L. Xiao, and T. Ou, “Evaluation of global climate models for downscaling applications centred over the Tibetan plateau,” *International Journal of Climatology*, vol. 37, no. 2, pp. 657–671, 2016.
- [27] L. D. Brekke, M. D. Dettinger, E. P. Maurer, and M. Anderson, “Significance of model credibility in estimating climate projection distributions for regional hydroclimatological risk assessments,” *Climatic Change*, vol. 89, no. 3–4, pp. 371–394, 2008.
- [28] Y. Gao, J. S. Fu, J. B. Drake, Y. Liu, and J.-F. Lamarque, “Projected changes in extreme events in the eastern United States based on a high-resolution climate modeling system,” *Environmental Research Letters*, vol. 7, no. 4, Article ID 044025, 2012.
- [29] J. Gula and W. R. Peltier, “Dynamical downscaling over the Great Lakes basin of North America using the WRF regional climate model: the impact of the Great Lakes system on regional greenhouse warming,” *Journal of Climate*, vol. 25, no. 21, pp. 7723–7742, 2012.
- [30] J. Ma, H. Wang, and K. Fan, “Dynamic downscaling of summer precipitation prediction over China in 1998 using WRF and CCSM4,” *Advances in Atmospheric Sciences*, vol. 32, no. 5, pp. 577–584, 2015.
- [31] M. Komurcu, K. A. Emanuel, M. Huber, and R. P. Acosta, “High-resolution climate projections for the northeastern United States using dynamical downscaling at convection-permitting scales,” *Earth and Space Science*, vol. 5, no. 11, pp. 801–826, 2018.