Performance Evaluation of CMIP5 and CMIP6 Models on Heatwaves in Korea and Associated Teleconnection Patterns

Maeng-Ki Kim1, Dae-Geun Yu1, Ji-Seon Oh3, Young-Hwa Byun2, Kyung-On Boo2, Il-Ung Chung1, Jeong-Soo Park4, Doo-Sun R. Park5, Seung-Ki Min6, and Hyun Min Sung2

1Department of Atmospheric Science, Kongju National University, Gongju, South Korea, 2Innovative Meteorological Research Department, National Institute of Meteorological Sciences, Seogwipo, South Korea, 3Department of Atmospheric and Environmental Sciences, Gangneung-Wonju National University, Gangneung-si, South Korea, 4Department of Statistics, Chonnam National University, Gwangju, South Korea, 5Department of Earth Science Education, Kyungpook National University, Daegu, South Korea, 6Division of Environmental Sciences and Engineering, Pohang University of Science and Technology (POSTECH), Pohang, South Korea

Abstract This study assesses the performance of the Coupled Model Intercomparison Project (CMIP) models for simulating summer heatwaves in Korea during a historical simulation period (1979–2014) using four diagnostic indices that represent the teleconnection mechanism of summer heatwaves in Korea. Four skill metrics are used for the model evaluation, namely, relative error (RE), interannual variability skill-score (IVS), correlation coefficient (CC), and total ranking (TR) based on daily maximum temperature (TMAX) in Korea and the four diagnostic indices. The results show that the REs of CMIP5 models do not differ significantly from those of the CMIP6 models while the IVSs in the CMIP6 models are significantly improved compared with the CMIP5 models. Observations show that the heatwave circulation index (HWCI) contributes more to the interannual variability in TMAX in Korea than that of the Indian Monsoon Rainfall Index (IMRI), indicating that the teleconnection from the northwestern Pacific is more important than that from northwest India. Interestingly, the CMIP6 models simulate this property better than the CMIP5 ensemble. The higher TR of CMIP6 models than CMIP5 supports that CMIP6 models are better overall in simulating heatwaves in Korea and the associated diagnostic indices.

1. Introduction

In recent years, extreme climatic phenomena, such as heatwaves, floods, and typhoons, have occurred frequently around the world, and these phenomena greatly affect human society and ecosystems (IPCC, 2012, 2013; Sun et al., 2014). Thus, it is important to understand extreme phenomena in terms of mechanisms, and accurately predict them by global climate models to minimize the damage caused by extreme phenomena (Cherchi et al., 2014; Enomoto et al., 2009; H. J. Lee et al., 2016; Lee et al., 2019; Pezza et al., 2012).

Developed by various modeling groups around the world under the Coupled Model Intercomparison Project (CMIP) (Eyring et al., 2016; Taylor et al., 2012), global circulation models (GCMs) are collectively evaluated (e.g., Gleckler et al., 2008; Lee et al., 2019) and widely used to study the effects of past, present, and future global and regional climate change in terms of climate variables (Bozkurt et al., 2018; Chen et al., 2011; Gulizia & Camilloni, 2015; Gusain et al., 2020; Notz, 2015; Perkins et al., 2007), large-scale modes (Gillett et al., 2013; Lee et al., 2019; Lim et al., 2016), and mechanisms (Cai et al., 2018; Cherchi et al., 2014; Garreaud et al., 2020; Jiang et al., 2015; Purich et al., 2013; Zazuille et al., 2018).

Recent studies have shown that in the RCP4.5 and RCP8.5 scenarios of CMIP5, the frequency, intensity, and duration of heatwaves in the future will increase significantly in most regions of the world, and will increase more rapidly in RCP8.5 (Meehl & Tebaldi, 2004; Orlowsky & Seneviratne, 2012; Sun et al., 2014; Yao et al., 2012; Zhou et al., 2014). The frequency of heatwaves varies among regions but is not proportional to global warming (Barriopedro et al., 2011; Harrington et al., 2018; Kim et al., 2019; Knutti et al., 2016; D. Lee et al., 2016; Lee et al., 2014; Min et al., 2015; Penalba & Rivera, 2016; Shin et al., 2018).
An increase in anthropogenic greenhouse gases has increased heatwaves in East Asia including Korea (Im et al., 2019; D. Lee et al., 2016; Min et al., 2014; Sun et al., 2014). However, the simulated performance of heatwaves during historical experiments shows many differences between models (Guo et al., 2017; Seo et al., 2018). Therefore, it is necessary to assess whether climate models can reproduce actual heatwaves and the associated mechanisms found in the observations. There have been limited efforts to evaluate the mechanism of heatwaves in Korea in the CMIP model, especially their performances in CMIP5 and CMIP6. In this study we focused on the evaluation of the performance of both CMIP5 and CMIP6 in terms of heatwaves in Korea, as well as the associated mechanism. Thus, we first defined appropriate diagnostic indices that determine heatwaves in Korea and associated mechanism in section 2, and then, we examined how well the CMIP5 and CMIP6 models reproduce heatwaves in Korea, as well as the associated diagnostic indices in section 3. Section 4 consists of the summary and conclusions.

2. Data and Methods

2.1. Data
This study used the daily maximum temperature (TMAX) data of 59 stations from the Korea Meteorological Administration (KMA) for 36 years from 1979 to 2014. The TMAX in the observations was calculated by averaging the TMAX at 59 in situ observations in Korea for July–August (JA). High-resolution Infrared Radiation Sounder-Outgoing Longwave Radiation (HIRS-OLR) data and monthly precipitation data from the Global Precipitation Climatology Project (GPCP) were used to analyze convective activity and

Table 1
Properties of 29 CMIP5 Global Climate Models

| Institution ID | Modeling center | Model name | Atmospheric resolution (Lat × Lon) |
|----------------|-----------------|------------|----------------------------------|
| CSIRO–BoM      | Commonwealth Scientific and Industrial Research Organization and Bureau of Meteorology, Australia | ACCESS1.0 | 144 × 192                        |
| BCC            | Beijing Climate Center, China Meteorological Administration, China | bcc-csm1.1 | 64 × 128                          |
|                |                  | bcc-csm1.1(m) | 160 × 320                        |
| CCCma          | Canadian Centre for Climate Modeling and Analysis, Canada | CanESM2 | 64 × 128                          |
| NCAR           | National Center for Atmospheric Research, United States | CCSM4 | 192 × 288                         |
| CMCC           | Centro Euro-Mediterraneo per I Cambiamenti Climatici, Italy | CMCC-CESM | 48 × 96                           |
|                |                  | CMCC-CM | 240 × 480                         |
|                |                  | CMCC-CMS | 96 × 192                          |
| CNRM–CERFACS   | Centre National de Recherches Météorologiques–Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique, France | CNRM-CM5 | 128 × 256                         |
| QCCCE          | Australian Commonwealth Scientific and Industrial Research Organization, Australia | CSIRO-Mk3.6.0 | 96 × 192                          |
| NOAA/GFDL      | NOAA/Geophysical Fluid Dynamics Laboratory, United States | GFDL-CM3 | 90 × 144                          |
|                |                  | GFDL-ESM 2G | 90 × 144                          |
|                |                  | GFDL-ESM 2 M | 90 × 144                          |
| NASA           | Goddard Institute for Space Studies (NASA), United States | GISS-E2-H | 90 × 144                          |
|                |                  | GISS-E2-R | 90 × 144                          |
| NIMS           | National Institute of Meteorological Research, Republic of Korea | HadGEM2 AO | 144 × 192                         |
| MOHC           | Met Office Hadley Centre, United Kingdom | HadGEM2-CC | 144 × 192                         |
|                |                  | HadGEM2-ES |                                      |
| INM            | Institute for Numerical Mathematics (INM) | INM-CM4 | 120 × 180                         |
| IPSL           | L’Institut Pierre-Simon Laplace, France | IPSL-CM5A-LR | 96 × 96                           |
|                |                  | IPSL-CM5A-MR | 143 × 144                         |
|                |                  | IPSL-CM5B-LR | 96 × 96                           |
| MIROC          | National Institute for Environmental Studies, The University of Tokyo, Japan | MIROC5 | 128 × 256                         |
|                |                  | MIROC-ESM | 64 × 128                          |
|                |                  | MIROC-ESM-CHEM | 64 × 128                         |
| MPI-M          | Max Planck Institute for Meteorology, Germany | MPI-ESM-LR | 96 × 192                          |
| MRI            | Meteorological Research Institute, Japan | MPI-ESM-MR | 96 × 192                          |
| NCC            | Norwegian Climate Centre, Norway | NORESM1-M | 96 × 144                          |

Note. We used ensemble r1i1p1 for all the models.
precipitation associated with heatwaves in Korea. The horizontal resolutions of these data are both 2.5° × 2.5°. This study also used the geopotential height of the National Centers for Environmental Prediction/National Centers for Atmospheric Research (NCEP/NCAR) reanalysis 2 (R2) data to examine the association with atmospheric circulation patterns. The horizontal resolutions of these data are 2.5° × 2.5°.

CMIP data are available through the Earth System Grid Federation (ESGF) which is one of the most complex big data systems (Williams et al., 2016). The data from the historical simulation produced by 29 CMIP5 participating models and 25 CMIP6 participating models have been used in this study, as shown in Tables 1 and 2. The variables used are monthly TMAX, outgoing longwave radiation (OLR), precipitation (PRCP), and geopotential height (GPH) for 36 years from 1979 to 2014. Unfortunately, CMIP5 historical data are not available after 2005 because the historical simulation ends in 2005. Thus, the Representative Concentration Pathway (RCP) 8.5 scenario data were used for the period from 2006 to 2014. The CMIP models have different horizontal resolutions, as shown in Tables 1 and 2. We converted CMIP model data to a common 2.5° × 2.5° grid using a bilinear interpolation scheme to facilitate comparison between the models and observations.

Notably, the observation, reanalysis, and CMIP model data presented above were analyzed on average from JA to analyze summer heat waves.

2.2. Heatwave in Korea and Associated Diagnostic Indices

In this study we defined a heatwave as when TMAX exceeds 33°C more than two consecutive days, which corresponds to the KMA criteria. The 33°C threshold is the 90th percentile of the summer TMAX in Korea (Lee & Lee, 2016). We also have used the number of heatwave days (NHD) to represent the frequency of heatwaves. The NHD is calculated by the sum of the day of the TMAXs above 33°C for each station during
Table 3
Correlation in Korea Between Monthly TMAX and NHD During JA in CMIP5 and CMIP6 Participating Models

| CMIP5          | CMIP6          |
|---------------|---------------|
| ACCESS1-0     | ACCESS-CM2    | 0.731 |
| bcc-csm1-1    | ACCESS-ESM1-5 | 0.679 |
| bcc-csm1-1-m  | BCC-CSM2-MR   | 0.587 |
| CanESM2       | BCC-ESM1      | 0.841 |
| CCSM4         | CanESM5       | 0.821 |
| CMCC-CESM     | EC-Earth3-Veg | 0.876 |
| CMCC-CM       | EC-Earth3-Veg | 0.861 |
| CMCC-CMS      | FGOALS-g3     | 0.740 |
| CNRM-CM5      | GFDL-CM4      | 0.752 |
| CSIRO-Mk3-6-0 | GFDL-ESM 4    | 0.785 |
| GFDL-CM3      | GISS-E2-1-G   | 0.880 |
| GFDL-ESM 2G   | GISS-E2-1-H   | -    |
| GFDL-ESM 2M   | INM-CM4-8     | 0.808 |
| GISS-E2-H     | INM-CM5-0     | 0.732 |
| GISS-E2-R     | IPSL-CM6A-LR  | 0.796 |
| HadGEM2-AO    | KACE-1-0-G    | 0.754 |
| HadGEM2-CC    | MIROC6        | 0.871 |
| HadGEM2-ES    | MPI-ESM 1-2-HR| 0.836 |
| incm4         | MPI-ESM 1-2-LR| 0.913 |
| IPSL-CM5A-LR  | MRI-ESM 2-0   | 0.682 |
| IPSL-CM5A-MR  | NESM 3        | 0.805 |
| IPSL-CM5B-LR  | NorESM2-LM    | 0.847 |
| MIROC5        | NorESM2-MM    | 0.825 |
| MIROC-ESM     | SAM0-UNICON   | 0.813 |
| MIROC-ESM-CHEM| UKESM 1-0-LL  | 0.882 |
| MPI-ESM-LR    | 0.852         |
| MPI-ESM-MR    | 0.905         |
| MRI-CGCM3     | 0.828         |
| NorESM1-M     | 0.786         |

Note. Correlation in observation is 0.886. See Table 4 for the domain.

JA, divided by the total number of stations. The intensity of heatwave is defined as the sum of (TMAX minus 33°C) for each station, divided by the total number of stations. The monthly mean TMAX showed a high correlation with NHD, with a correlation coefficient (CC) of 0.886 during JA, as shown in Kim et al. (2019), indicating that the NHD increases as the intensity of summer heatwaves in Korea increases. Sun et al. (2014) also showed that the CC between two time series is approximately 0.74 in Eastern China. Thus, the magnitude of TMAX can explain both the intensity and frequency of heatwaves in Korea under KMA’s criteria. If such a relationship does not appear well in the model, it is difficult to say that the model has excellent performance. Although the CC is low in some CMIP models, most models show strong correlations similar to the observed (Table 3), indicating that monthly TMAX is appropriate to investigate the heatwave in Korea.

Figure 1a represents the correlation pattern between TMAX in the observation and the geopotential height at 500 hPa obtained from the reanalysis. The positive correlation pattern centering on the Korean Peninsula and the negative correlation pattern centering on the subtropics and equator show the zonal structure. Specifically, the wave pattern from the Northwest Pacific to the high latitudes through the Korean Peninsula is a typical pattern of the north-south teleconnection. Figure 1b shows the correlation pattern between TMAX in the observation and OLR. A strong positive correlation extends east-west from the Tibetan Plateau to the eastern part of Japan, centering on the Korean Peninsula, with strong negative correlations appearing in the Northwest Pacific and northwestern India. These results correspond with those of Lee and Lee (2016) and Kim et al. (2019). Furthermore, Figure 2 also shows that the correlation pattern with NHD, which represents the frequency of heatwaves in Korea, shows a similar pattern as that shown in Figure 1. The two figures indicate that the upward movement enhanced by the strong convection activity in the Northwest Pacific induces downward movement in East Asia through the north-south teleconnection, resulting in a high-pressure anomaly favorable for heatwaves in Korea. The convective activity near the South China Sea and the Philippine Sea (black box in Figures 1b and 2b) is specifically important in terms of the interannual variabilities in TMAX and NHD, which is consistent with the Pacific-Japan (PJ) teleconnection (Nitta, 1987) and previous studies (Kim et al., 2019, 2020; Kosaka & Nakamura, 2006; Lee & Lee, 2016; Yeh et al., 2018; Yeo et al., 2019), indicating that the OLR over the region can be one of diagnostic indices explaining heatwaves in Korea.

Notably, another strong convective activity related to heatwaves in Korea is located over northwestern India (Figures 1b and 2b). According to recent studies, the Circumglobal Teleconnection (CGT), a representative pattern of midlatitude waves, induces atmospheric circulation patterns and affects heatwaves in Korea through Rossby wave propagation along the jet stream guide (Enomoto et al., 2003; Kim et al., 2019, 2020; Min et al., 2020; Park & Schubert, 1997; Yeo et al., 2019). The increase in sensible heat in the Tibetan Plateau and the strengthening of convective activity in northwestern India specifically induce CGT-like patterns, leading to a high-pressure anomaly in favor of heatwaves over the Korean Peninsula (Ding & Wang, 2005; Kim et al., 2019; Lau et al., 2006; Wu et al., 2016). Thus, this result is consistent with previous studies, indicating that the OLR over northwestern India can be one of the indices explaining heatwaves in Korea.

To assess heatwaves in Korea and the associated mechanism, we defined four diagnostic indices explaining teleconnection with heatwaves in Korea (Table 4). Table 4 shows the definition of each index. When a heatwave occurs over the Korean Peninsula, a high-pressure anomaly over the Korean Peninsula and a low-pressure anomaly in the northwestern Pacific are evident in the middle and lower layers (Lee & Lee, 2016; Yeo et al., 2019). Thus, we have defined the heatwave circulation index (HWCI) using the 500 hPa geopotential height difference between the Korean Peninsula and the Northwest Pacific based on
the results of previous studies (Lee & Lee, 2016; Yeo et al., 2019) and Figures 1a and 2a. HWCI means that the greater the positive value is, the stronger the north-south teleconnection is. According to Kim et al. (2019), the interannual variability in precipitation in northwest India is closely related to heatwaves in Korea. Thus, we have defined the Indian monsoon rainfall index (IMRI) based on Kim et al. (2019) in Table 4. We have also defined OLRI1 and OLRI2 using OLRs in the Northwest Pacific region and northwestern India, which are closely related to heatwaves in Korea (Table 4), as shown in Figures 1b and 2b.

2.3. Model Evaluation Metrics

In this study we used some metrics to evaluate the model performance on TMAX in Korea and associated diagnostic indices for 36 years from 1979 to 2014. All indices each JA are calculated by the area average of grid points in the given domain (Table 4), and metrics are calculated using the JA mean value of the index. In situ observations in Korea and reanalysis data are used for model evaluation of TMAX and diagnostic indices, respectively.

First, relative error (RE) was used to check the mean error for each model (Equation 1). Furthermore, the interannual variability skill score (IVS) (Chen et al., 2011; Jiang et al., 2015) was used to verify that the models simulate the interannual variability in the observations well (Equation 2):
Figure 2. Correlation patterns of the number of heat wave days (NHD) in Korea with (a) GPH at 500 hPa and (b) OLR during JA for 36 years from 1979 to 2014. Small and large dots represent significance levels of 0.05 and 0.01 based on Student’s t tests, respectively. Boxes indicate the domain used for the definition of diagnostic indices in Table 4. TMAX, GPH, and OLR are obtained from the observations, reanalysis, and NOAA, respectively.

\[ \text{RE} = \frac{X_m - X_o}{X_o} \times 100 \, \% , \]  

(1)

\[ \text{IVS} = \left( \frac{\text{STD}_m}{\text{STD}_o} - 1 \right)^2 , \]  

(2)

where \( X \) and \( STD \) indicate the climatological mean of \( X \) and standard deviation of \( X \), respectively. Subscripts \( m \) and \( o \) represent the model and observation, respectively. The values of \( \text{RE} \) and \( \text{IVS} \) closer to zero indicate that the model simulates similarly to the observations.

Table 4
Indices Associated With Heatwaves in Korea

| Index name                                      | Variable                                            | Domain (longitude, latitude)                     |
|------------------------------------------------|-----------------------------------------------------|-------------------------------------------------|
| TMAX                                           | Daily maximum temperature                           | (125–130°E, 35–37.5°N)                          |
| Heatwave circulation index (HWCI)              | Geopotential height at 500 hPa                      | (115–145°E, 35–45°N) minus (110–135°E, 15–25°N) |
| Indian monsoon rainfall index (IMRI)            | Precipitation                                       | (70–80°E, 20–35°N)                              |
| Outgoing longwave radiation index 1 (OLRI1)     | OLR                                                 | (110–135°E, 15–25°N)                            |
| Outgoing longwave radiation index 2 (OLRI2)     | OLR                                                 | (65–77.5°E, 22.5–32.5°N)                        |

Note. All indices are calculated using the area averaged value in the domain. Note that TMAX in the observations was calculated from the average of 59 in-situ observations in Korea.
The RE and IVS can explain the similarity of the observed index to the index simulated by the model. However, they do not explain the association with heatwaves in Korea. Therefore, we used a multiple regression to find a combination of diagnostic indices that could explain the interannual variability in heatwaves in Korea using observational data.

In this process, the leave-one-out cross validation was performed using the 36 years of data for the verification of the multiple regression equation. The verification results showed that the combination of diagnostic indices that best explained the variability in TMAX in Korea was the linear combination of HWCI and IMRI, which accounted for approximately 61.4% of the total variability during the verification period. Furthermore, TMAX was highly correlated with HWCI and IMRI, with CCs of 0.70 and 0.45 in the verification, respectively. Therefore, we used the CC between the diagnostic indices and TMAX together with RE and IVS as a significant metric for model evaluation. In other words, even if the RE and IVS are similar to the observations, if the CC is low, the simulation performance of the Korean Peninsula heatwave is not considered to be good.

The RE, IVS, and CC provide us with model performance in terms of climatological mean, interannual variability, and teleconnection with diagnostic indices. Therefore, summarizing the rankings of TMAX and associated indices for each metric is also useful for model evaluation. The comprehensive rating metric (MR) of the models for each metrics was calculated by Equation 3 according to the method of Jiang et al. (2015).

$$ MR = 1 - \frac{1}{nm} \sum_{i=1}^{n} rank_i; $$

where \( n \) and \( m \) indicate the number of the index and the number of models, respectively. \( rank_i \) represents the ranking of the target model for index \( i \). From Equation 3, we can obtain the MRs of the CMIP models for RE, IVS, and CC. Here the MRs gives us the ranking information of the models for each metrics.

In addition, summarizing all the rankings should be useful in evaluating the CMIP models. Therefore, the total ranking (TR) metrics was defined in Equation 4 by applying Equation 3 to all ranking as follows:

$$ TR = 1 - \frac{1}{nmM} \sum_{j=1}^{M} \sum_{i=1}^{n} rank_{(i,j)}; $$

where \( M \) represents the number of metrics. \( rank_{(i,j)} \) indicates the ranking of index \( i \) for metric \( j \). Thus, TR provides us with the total ranking information of each model as only a single value regardless of index and metric.

Note that MR and TR values closer to 1 indicate better model performance. The TR can be calculated for each CMIP group. However, the objective way to compare TRs of two CMIP groups is to calculate TRs using all the models regardless of the CMIP phase, which will be shown later.

### 3. Model Evaluation

#### 3.1. Model REs and IVSs

Figure 3a shows the RE for the diagnostic indices as shading, with the number in each box representing the RE value. The MME located at the bottom represents the multimodel ensemble. Note that positive, zero, and negative values indicate overestimation, perfect, and underestimation, respectively.
underestimation, respectively. The MMEs show that the difference between CMIP groups is not large regardless of TMAX and associated indices. IMRI, TMAX, and HWCI in CMIP 5 (6) are underestimated with $-40.3\% (-44.0\%)$, $-16.1\% (-15.7\%)$, and $-7.8\% (-6.3\%)$ in MME, respectively. This result means that two teleconnections forced from the northwestern Pacific and Northwestern India in the CMIP models are weaker than the observed one. As a result, TMAX is also lower than that observed due to the impact of...
Figure 5. Scatter plot of CC between TMAX in Korea and associated diagnostic indices (HWCI on the y axis and IMRI on the x axis) for (a) CMIP5 and (b) CMIP6. The size of the circle indicates the coefficient of determination of the regression equation during the verification period. OBS indicates observation.
Figure 6. The portrait diagram for the rankings of RE (left), IVS (center), and CC (right) between TMAX in Korea and heatwave diagnostic indices for (a) CMIP5 and (b) CMIP6. The model names listed on the left in each panel follow their averaged rank.
weak teleconnections. In the case of OLR, OLRI1 simulates observations relatively well, while OLRI2 is overestimated in most models. This result means that convective activity in northwestern India is weaker than that observed, and the influence of east-west teleconnection on TMAX in Korea is weaker than that observed. The overestimation of OLRI2 is consistent with the underestimation of IMRI, which means weak convective activity. Note that while other indices are similar in the sign of RE between models, HWCI significantly differs in the sign of RE between models. The RE of HWCI in CMIP5 (6) is the least consistent between the models, with a minimum of −60.2% (−82.8%) and a maximum of 102.4% (58.8%). This means that the MME of the models simulates convective activities in the Northwest Pacific well on average, but there are significant differences between the models in the north-south teleconnection to heatwaves in Korea. In contrast, the RE of TMAX in CMIP 5 (6) is underestimated in all models compared with the observations with range from −34.4% (−30.9%) to −2.5% (−3.3%).

To summarize, the CMIP participating models simulate convective activity near the northwestern Pacific well but there is a limitation in simulating convective activity in the northwestern India. Thus, the impact of the Indian monsoon on heatwaves in Korea as a remote forcing may not work effectively in CMIP models compared with the observations. We will discuss this part in the next section.

IVS is a metrics for measuring model performance regarding interannual variability. Thus, the IVS is independent of the RE measuring mean bias. Figure 4 shows that the IVS of the CMIP6 models is smaller overall than that of the CMIP5 models, indicating that the interannual variability in TMAX and associated indices is improved in the CMIP6 models compared with the CMIP5 models. In particular, the IVS of IMRI in CMIP6 is significantly improved with ensemble mean value of 0.91 compared with 3.56 in CMIP5. The large IVS of IMRI in CMIP5 is partly due to IVS values over 10 in some models, as shown in Figure 4a.

### 3.2. Relationship Between TMAX in Korea and Associated Diagnostic Indices

As explained in section 2.3, we have used a multiple regression model in the observations to find a combination of diagnostic indices that best explains the interannual variability in TMAX. The results show that the HWCI and IMRI indices are major indices used to explain TMAX in Korea. Thus, the two indices were used to evaluate the simulation performance of the model (Figure 5). In Figure 5, the horizontal axis represents the CC between TMAX and HWCI, and the vertical axis represents the CC between TMAX and IMRI. Note that the size of a circle represents the coefficient of determination in the verification and model performance in CC is determined by the distance from the observed CC in Figure 5. Therefore, the models closest to the observed point have the best performance.

In the observations, a linear combination of HWCI and IMRI accounted for 61.4% of the interannual variability in the TMAX during the verification period. In CMIP5, the CC values are quite scattered (Figure 5a). However, CMCC-CM, CMCC-ESM, and IPSL-CM5A-LR are in a similar position as the observation. In the case of the INMCM4 and CSIRO-Mk3-6-0 models, the correlation between TMAX and HWCI was slightly higher than that in the observation, indicating that the two models simulate the link between heatwaves in Korea and convective activity over the northwestern Pacific. However, there is a limitation in simulating the teleconnection between heatwaves in Korea and the Indian summer monsoon in the CSIRO-Mk3-6-0 model.
The best heatwaves in Korea are more strongly related to teleconnection from the northwestern Pacific than from western India. Notably, in most models, the CC of the HWCI is greater than the CC of the IMRI, indicating that CCs between HWCI and TMAX in nine of the CMIP6 models are comparable with the observations. In CMIP6, the scattering of the CC points is greatly reduced compared with CMIP5 (Figure 5b). Interestingly, CCs between HWCI and TMAX in nine of the CMIP6 models are comparable with the observations. Notably, in most models, the CC of the HWCI is greater than the CC of the IMRI, indicating that heatwaves in Korea are more strongly related to teleconnection from the northwestern Pacific than from northwestern India. The best five models based on CC are GFDL-ESM 4, INM-CM5-0, EC-Earth3, INM-CM4-8, and KACE-1-0-G.

### 3.3. Overall Evaluation in TR

In this section we have assessed model performance using RE, IVS, and CC. For the overall evaluation, the ranking of the models for the three metrics was calculated for each index including TMAX. The overall ranking was determined using the average of the rankings and represented by the portrait diagram (Figure 6). Note that in Figure 6, the rankings of RE, IVS, and CC for each index are placed on the left, center, and right, respectively. Note that the red line is superior to the purple line.

In CMIP5, the CMCC-CM, MIROC5, HadGEM2-ES, HadGEM2-AO, and GFDL-ESM 2G models were ranked in the top five (Figure 6a). The IPSL-CM5B-LR, GISS-E2-R, and GISS-E2-H models, however, do not perform well, at least for the Korean Peninsula. Notably, the top five models do not have good rankings for all indices and metrics. For example, the rankings of the REs for three of the indices except for OLRI1 are lower than the IVSs and CCs in the CMCC-CM model. However, the purple line shows a lower ranking in most indices and metrics. Satisfactory model performance appears in the GFDL-ESM 2G and CanESM2 models for the HWCI, MPI-ESM-LR and MIROC5 models for OLRI1, the HadGEM2-CC and CMCC-CM models for OLRI2, and the MIROC-ESM and HadGEM2-ES models for IMRI. However, TMAX showed the highest ranking in MIROC-ESM-CHEM and MPI-ESM-MR.

In CMIP6, EC-Earth3, EC-Earth3-Veg, SAM0-UNICON, KACE-1-0-G, and UKESM-1-0-LL were ranked in the top five (Figure 6b). However, the GISS-E2-1-H, FGOALS-g3, and CanESM5 models do not show a
that the EC models satisfactorily simulate the Earth3-Veg index. By analyzing the model ranking for each diagnostic index, it was found that the EC models are more sensitive to OLR than in CMIP6 models. The best two models are EC Earth3-Veg and EC Earth3-Veg, respectively, with TR values of 0.687 and 0.687, respectively, which participate in CMIP6 (Table 5). The third ranking model is CMCC-CM with TR value of 0.627, which participates in CMIP5. The fourth and fifth models are also participating in CMIP6, indicating that four of the top five models are participating in CMIP6. Thus, the model performance in CMIP6 is superior overall to that in CMIP5. Refer to Table 5 for the detailed ranking.

As can be seen from observations (Figure 1), both OLRI1 and OLRI2 have a significant correlation with TMAX. This means that they are good diagnostic indices to represent remote forcing. In this respect, all diagnostic indices including OLRI1 and OLRI2 were included when calculating TR. In consideration of the readers' interests, we have tested the sensitivity of TR results to the exclusion of the two OLR indices, i.e., using only HWCI, IMRI, and TMAX (Table 6). There are some noticeable differences in TR of CMIP5 models, but CMIP6 model ranks remain largely unaffected. This supports that the TR obtained from four diagnostic indices reliably represent the model's overall performance in CMIP6 models. TR distributions in CMIP5 models are more sensitive to OLR than in CMIP6 models. The best two models are EC Earth3 and

Table 6
Total Ranking (TR) Values of CMIP Participating Models Using HWCI, IMRI, and TMAX

| No. | CMIP5        | TR  | Ranking | CMIP6        | TR  | Ranking |
|-----|--------------|-----|---------|--------------|-----|---------|
| 1   | GFDL-CM3     | 0.646 | 6       | EC-Earth3-Veg | 0.734 | 1       |
| 2   | MIROC-ESM-CHEM | 0.634 | 7       | EC-Earth3-Veg | 0.706 | 2       |
| 3   | GFDL-ESM 2G | 0.625 | 8       | SAM0-UNICON   | 0.660 | 3       |
| 4   | CanESM2      | 0.616 | 10      | KACE-1-0-G     | 0.655 | 4       |
| 5   | HadGEM2-AO   | 0.593 | 11      | GFDL-ESM 4     | 0.653 | 5       |
| 6   | IPSL-CM5A-LR | 0.590 | 12      | UKESM-1-0-LL   | 0.623 | 9       |
| 7   | MPI-ESM-MR   | 0.569 | 16      | MIROC6        | 0.583 | 13      |
| 8   | HadGEM2-ES   | 0.560 | 18      | BCC-CSM2-MR    | 0.576 | 14      |
| 9   | CMCC-CM      | 0.546 | 19      | GFDL-CM4       | 0.572 | 15      |
| 10  | MIROC5       | 0.537 | 20      | NorESM2-MM     | 0.569 | 16      |
| 11  | MIROC-ESM    | 0.525 | 21      | MRI-ESM 2-0    | 0.523 | 22      |
| 12  | CSM4         | 0.509 | 25      | BCC-CSM 1      | 0.519 | 23      |
| 13  | GFDL-ESM 2M  | 0.495 | 27      | IPSL-CM6A-LR   | 0.516 | 24      |
| 14  | NorESM1-M    | 0.488 | 29      | INM-CM4-8      | 0.509 | 26      |
| 15  | CMCC-CESM    | 0.475 | 31      | INM-CM5-0      | 0.493 | 28      |
| 16  | HadGEM2-CC   | 0.475 | 31      | ACCESS-ESM 1-5 | 0.486 | 30      |
| 17  | inmcm4       | 0.470 | 33      | MPI-ESM 1-2-HR | 0.447 | 35      |
| 18  | MPI-ESM-LR   | 0.461 | 34      | NESM3          | 0.442 | 37      |
| 19  | CMCC-CMS     | 0.444 | 36      | MPI-ESM 1-2-LR | 0.438 | 38      |
| 20  | CNRM-CM5     | 0.431 | 39      | NorESM2-LM     | 0.405 | 43      |
| 21  | IPSL-CM5A-MR | 0.424 | 40      | ACCESS-CM2     | 0.347 | 48      |
| 22  | ACCESS1-0    | 0.419 | 41      | CanESM5        | 0.336 | 49      |
| 23  | bcc-csm1-1-m | 0.410 | 42      | FGOLs-g3       | 0.329 | 50      |
| 24  | bcc-csm1-1   | 0.400 | 44      | GISS-E2-1-G    | 0.292 | 51      |
| 25  | IPSL-CM5B-LR | 0.396 | 45      | GISS-E2-1-H    | 0.255 | 53      |
| 26  | MRI-CGC3M    | 0.359 | 46      |                  |      |         |
| 27  | CSIRO-Mk3-6-0 | 0.354 | 47      |                  |      |         |
| 28  | GISS-E2-R    | 0.278 | 52      |                  |      |         |
| 29  | GISS-E2-H    | 0.102 | 54      |                  |      |         |
EC-Earth3-Veg, with TR values of 0.734 and 0.706, respectively, which participate in CMIP6. The SAM0-UNICON, KACE-1-0-G, and GFDL-ESM 4 also show high TR values over 0.650. Note that the top five models all belong to CMIP6 models.

Figure 8 shows the correlation pattern of TMAX in Korea with (a) GPH at 500 hPa and (b) OLR during JA for 36 years from 1979 to 2014. Small and large dots represent significance levels of 0.05 and 0.01 based on Student’s t tests, respectively. Boxes indicate the domain used for the definition of diagnostic indices in Table 4. The patterns are obtained from the ensemble mean of the best six models in CMIP6.

4. Summary and Conclusions

In this study we evaluated the performance of the CMIP participating models for heatwaves in Korea during a historical simulation (1979–2014) using four diagnostic indices that explain heatwaves in Korea. Two
indices, i.e., HWCI and OLRI1, represent the north-south teleconnection, while other indices, i.e., OLRI2 and IMRI, represent the east-west teleconnection. Four metrics of RE, IVS, CC, and TR were used for model evaluations of each index. The main results are summarized as follows:

1. By analyzing the REs of CMIP participating models, TMAX and IMRI tended to underestimate by 15–20% and 40–50%, respectively, and OLRI2 overestimated by 7–9% (Figure 3). This result means that the simulated convection over northwestern India is not enough to provide proper teleconnection effects for heatwaves in Korea. In addition, there is no clear difference in REs between CMIP5 and CMIP6.

2. By analyzing the IVs of the CMIP models, the CMIP6 models showed significantly improved IVS than the CMIP5 model (Figure 4). Especially in IMRI, the IVS is decreased by 74% in CMIP6 compared with CMIP5.

3. In the observations the interannual variability in TMAX in Korea can be explained best by the linear combination of HWCI and IMRI (Figure 5). Notably, HWCI mostly contributed to the interannual variability in TMAX, greater than that of IMRI, indicating that remote forcing from the northwestern Pacific is more important in modulating heatwaves in Korea than northwestern India. The CMIP6 models simulate this property better than CMIP5. In CMIP5, the CMCC-CM, CMCC-CESM, and IPSL-CM5A-LR models showed the best performances. In CMIP6, the best five models based on CCs are GFDL-ESM 4, INM-CM5-0, EC-Earth3, INM-CM4-8, and KACE-1-0-G.

4. To summarize, CMIP6 is better than CMIP5 in terms of TR, indicating that CMIP6 models have been improved overall in simulating heatwaves in Korea, as well as the associated diagnostic indices. The best five models are EC-Earth3, EC-Earth3-Veg, SAM0-UNICON, KACE-1-0-G, and GFDL-ESM 4 which participate in CMIP6 (Table 6). Thus, the model performance in CMIP6 is overall superior to that in CMIP5.

This study quantitatively evaluated the simulation performance of the CMIP model using four diagnostic indices to explain heatwaves in Korea at the interannual time scale. The long-term change and variability in heatwaves in Korea could be affected by other factors than those suggested in this study (Min et al., 2020). Further studies on the mechanism of heatwaves will require further development of indices, and further collection of CMIP6 participation models will lead to more comprehensive performance assessments. Notably, significant scientific progress has been made by CMIP, but there is still room for improvement in the GCMs on local, regional, and global scales, especially concerning extreme events.

**Data Availability Statement**

The CMIP data are available through ESGF’s website (https://esgf-node.llnl.gov/search/cmip5/ and https://esgf-node.llnl.gov/search/cmip6/) and DKRZ’s website (https://delos.dkrz.de/WDCUI/erasearch/q). Observation station data in Korea are provided by KMA from their website (https://data.kma.go.kr/data/grnd/selectAsosRltmList.do?pgmNo=36). The NCEP/NCAR reanalysis 2 data are available through NCEP’s website (https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis2.html). The HIRS OLR and GPCP precipitation data are provided by NOAA website (https://www.nci.noaa.gov/data/outgoing-longwave-radiation-monthly/access/ and https://www.esrl.noaa.gov/psd/data/gridded/data.gpcp.html, respectively).

**Acknowledgments**

This work was funded by the Korea Meteorological Administration Research and Development Program under grant KMI2018-03410. Work of NIMS/KMA authors was performed under the auspice of the Korea Meteorological Administration Research and Development Program “Development and Assessment of IPCC AR6 Climate Change Scenarios (KMA2018-09321).” We thank three reviewers for their constructive comments.

**References**

Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M., & Garcia-Herrera, R. (2011). The hot summer of 2010: Redrawing the temperature record map of Europe. *Science*, 320(6026), 220–224. https://doi.org/10.1126/science.1201224

Bozkurt, D., Rojas, M., Boisier, J. P., & Valdivieso, J. (2018). Projected hydroclimate changes over Andean basins in central Chile from downscaled CMIP5 models under the low and high emission scenarios. *Climatic Change*, 150(3–4), 131–147. https://doi.org/10.1007/s10584-018-2246-7

Cai, L., Alexeev, V. A., Walsh, J. E., & Bhatt, U. S. (2018). Patterns, impacts, and future projections of summer variability in the Arctic from CMIP5 models. *Journal of Climate*, 31(24), 9815–9833. https://doi.org/10.1175/JCLI-D-18-0119.1

Chen, W., Jiang, Z., & Li, L. (2011). Probabilistic projections of climate change over China under the SRES A1B scenario using 28 AOGCMs. *Journal of Climate*, 24(17), 4741–4756. https://doi.org/10.1175/2011JCLI4102.1

Cherchi, A., Annamalai, H., Masina, S., & Navarra, A. (2014). South Asian summer monsoon and the eastern Mediterranean climate: The monsoon–desert mechanism in CMIP5 simulations. *Journal of Climate*, 27(18), 6877–6903. https://doi.org/10.1175/JCLI-D-13-00530.1

Ding, Q., & Wang, B. (2005). Circumglobal teleconnection in the northern hemisphere summer. *Journal of Climate*, 18(17), 3483–3505. https://doi.org/10.1175/JCLI3473.1

Enomoto, T., Endo, H., Harada, Y., & Ohfuchi, W. (2009). Relationship between high-impact weather events in Japan and propagation of Rossby waves along the Asian jet in July 2004. *Journal of the Meteorological Society of Japan Series II*, 87(1), 139–156. https://doi.org/10.2151/jmsj.87.139
Enomoto, T., Hoskins, B. J., & Matsuda, Y. (2003). The formation mechanism of the Bonin high in August. Quarterly Journal of the Royal Meteorological Society, 129, 157–178. https://doi.org/10.1256/qj.01.211

Ilying, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geoscientific Model Development, 9, 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Garreau, R. D., Boisier, J. P., Rondonelli, R., Montecinos, A., Sepulveda, H. H., & Veloso-Aguila, D. (2020). The central Chile mega drought (2010–2018): A climate dynamics perspective. International Journal of Climatology, 40(1), 421–439. https://doi.org/10.1002/joc.6219

Gillett, N. P., Fyfe, J. C., & Parker, D. E. (2013). Attribution of observed sea level pressure trends to greenhouse gas, aerosol, and ozone changes. Geophysical Research Letters, 40, 2302–2306. https://doi.org/10.1002/2010GL04972

Gleckler, P. J., Taylor, K. E., & Doutriaux, C. (2008). Performance metrics for climate models. Journal of Geophysical Research, 113, D06104. https://doi.org/10.1029/2007JD008972

Gulizia, C., & Camilloni, I. (2015). Comparative analysis of the ability of a set of CMIP3 and CMIP5 global climate models to represent precipitation in South America. International Journal of Climatology, 35(4), 583–593. https://doi.org/10.1002/joc.4005

Guo, X., Huang, J., Liu, Y., Zhao, Z., & Xu, Y. (2017). Projection of heat waves over China for eight different global warming targets using 12 CMIP5 models. Theoretical and Applied Climatology, 128, 507–522. https://doi.org/10.1007/s00704-015-1718-1

Gusain, A., Ghosh, S., & Karmakar, S. (2020). Added value of CMIP over CMIP5 models in simulating Indian summer monsoon rainfall. Atmospheric Research, 232, 104680. https://doi.org/10.1016/j.atmosres.2019.104680

Harrington, L. J., Frame, D., King, A. D., & Otto, F. E. (2018). How uneven are changes to impact-relevant climate hazards in a 1.5°C world and beyond? Geophysical Research Letters, 45, 6672–6680. https://doi.org/10.1029/2018GL078888

Im, E. S., Thanu, N. X., Kim, Y. H., & Ahn, J. B. (2019). 2018 summer extreme temperatures in South Korea and their intensification under 3 °C global warming. Environmental Research Letters, 14, 094020. https://doi.org/10.1088/1748-9326/abf8ef

IPCC (2012). Managing the risks of extreme events and disasters to advance climate change adaptation. In C. B. Field, et al. (Eds.), Climate Change 2012: 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 (p. 197). Cambridge, UK, and New York, NY, USA: Cambridge University Press.

IPCC (2013). Climate change 2013: The physical science basis. In T. F. Stocker, et al. (Eds.), Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change on Climate and Beyond? Geophysical Research Letters, 25, 6672–6680. https://doi.org/10.1029/2010GL043973

Lee, D., Min, S. K., Park, C., Suh, M. S., Ahn, J. B., Cha, D. H., et al. (2016). Time of emergence of anthropogenic warming signals in the Indian subcontinent on heat waves in South Korea. International Journal of Climatology, 36(9), 3166–3180. https://doi.org/10.1002/joc.5869

Knutti, R., Rogelj, J., Sedláček, J., & Fischer, E. M. (2016). A scientific critique of the two-degree climate change target. Nature Geoscience, 9(1), 13–18. https://doi.org/10.1038/ngeo2595

Kosaka, Y., & Nakamura, H. (2006). Structure and dynamics of the summertime Pacific-Japan teleconnection pattern. Quarterly Journal of the Royal Meteorological Society, 132, 2009–2030. https://doi.org/10.1256/qj.05.204

Lau, K. M., Kim, M. K., & Kim, K. M. (2006). Asian summer monsoon anomalies induced by aerosol direct forcing: The role of the Tibetan Plateau. Climate Dynamics, 26, 855–864. https://doi.org/10.1007/s00382-006-0114-2

Lee, D., Min, S. K., Park, C., Suh, M. S., Ahn, J. B., Cha, D. H., et al. (2016). Time of emergence of anthropogenic warming signals in the Northeast Asia assessed from multi-regional climate models. Asia-Pacific Journal of Atmospheric Sciences, 52(2), 129–137. https://doi.org/10.1007/s13143-016-0014-2

Lee, J. H., Lee, W. S., & Yoo, J. H. (2016). Assessment of medium-range ensemble forecasts of heat waves. Atmospheric Science Letters, 17(1), 19–25. https://doi.org/10.1002/asl.593

Lee, J. W., Hong, S. Y., Chang, C. E., Suh, M. S., & Kang, H. S. (2014). Assessment of future climate change over East Asia due to the RCP scenarios downscaled by GRIMs-RMP. Climate Dynamics, 42. https://doi.org/10.1007/s00382-013-1841-6

Lee, J., Sperber, K. R., Gleckler, P. J., Bonfils, C. J., & Taylor, K. E. (2019). Quantifying the agreement between observed and simulated extratropical modes of interannual variability. Climate Dynamics, 52(7–8), 4057–4089. https://doi.org/10.1007/s00382-018-4355-4

Lee, W. S., & Lee, M. I. (2016). Interannual variability of heat waves in South Korea and their connection with large-scale atmospheric circulation patterns. International Journal of Climatology, 36(15), 4815–4830. https://doi.org/10.1002/joc.4671

Lim, E. P., Hendon, H. H., Arblaster, J. M., Delage, F., Nguyen, H., Min, S. K., & Wheeler, M. C. (2016). The impact of the southern annular mode on future changes in southern hemisphere rainfall. Geophysical Research Letters, 43, 7160–7167. https://doi.org/10.1002/2016GL069453

Meehl, G. A., & Tebaldi, C. (2004). More intense, more frequent, and longer lasting heat waves in the 21st century. Science, 305(5686), 994–997. https://doi.org/10.1126/science.1098704

Min, S. K., Kim, Y. H., Kim, M. K., & Park, C. (2014). Assessing human contribution to the summer 2013 Korean heat wave. Bulletin of the American Meteorological Society, 95(9), S48–S51.

Min, S. K., Kim, Y. H., Lee, S. M., Sparrow, S., Li, S., Lott, F. C., & Stott, P. A. (2020). Quantifying human impact on the 2018 summer longest heat wave in South Korea. Bulletin of the American Meteorological Society, 101, S103–S108. https://doi.org/10.1175/BAMS-D-19-0151.1

Min, S. K., Kim, Y. H., Park, S. M., Kim, M. K., & Boo, K. O. (2015). Anthropogenic influence on the 2014 record-hot spring in Korea. Bulletin of the American Meteorological Society, 96(12), S95–S99. https://doi.org/10.1175/BAMS-D-15-00079.1

Nitta, T. (1987). Convective activities in the tropical western Pacific and their impact on the northern hemisphere summer circulation. Journal of the Meteorological Society of Japan. Series II, 65(3), 373–390. https://doi.org/10.2151/jmsj1965.65.3_373

Notz, D. (2015). How well must climate models agree with observations? Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 373(2052), 20140164. https://doi.org/10.1098/rsta.2014.0164

Orlowsky, B., & Seneviratne, S. I. (2012). Global changes in extreme events: Regional and seasonal dimension. Climatic Change, 110(3–4), 669–696. https://doi.org/10.1007/s10584-011-0122-9
