

# Future projections and uncertainty assessment of precipitation extremes in the Korean peninsula from the CMIP6 ensemble

Jeong-Soo Park<sup>1</sup>, Yonggwan Shin<sup>1</sup>, Yire Shin<sup>1</sup>, Juyoung Hong<sup>1</sup>, Maeong-Ki Kim<sup>2</sup>,  
Young-Hwa Byun<sup>3</sup>, Kyung-On Boo<sup>3</sup>, Il-Ung Chung<sup>4</sup>, and Doo-Sun R Park<sup>5</sup>

<sup>1</sup>Chonnam National University

<sup>2</sup>Kongju National University

<sup>3</sup>National Institute of Meteorological Sciences

<sup>4</sup>Dept of Atmospheric and Environmental Sciences

<sup>5</sup>Kyungpook National University

November 24, 2022

## Abstract

Projected changes in extreme climate are occasionally predicted through multi-model ensemble methods using a weighted averaging that combines predictions from individual simulation models. To predict future changes in precipitation extremes, observed data and 21 of the Coupled Model Inter-comparison Project Phase 6 (CMIP6) models are examined for 46 grids over the Korean peninsula. We apply the generalized extreme value distribution (GEVD) to the series of annual maximum daily precipitation (AMP1) data. Simulation data under three shared socioeconomic pathway (SSP) scenarios, namely, SSP2-4.5, SSP3-7.0, and SSP5-8.5, are used. A multivariate bias correction technique that considers the spatial dependency between nearby grids is applied to these simulation data. In addition, a model weighting approach that accounts for both performance and independence (PI-weighting) is employed. In this study, we estimate the future changes in precipitation extremes in the Korean peninsula using the multiple CMIP6 models and PI-weighting method. In applying the PI-weighting, we suggest simple ways for selecting two shape 1 parameters based on the chi-square statistic and entropy. Variance decomposition with the interaction term between the CMIP6 model and the SSP scenario is applied to quantify the uncertainty of projecting the future AMP1. Return levels spanning over 20 and 50 years, as well as the return periods relative to the reference years (1973-2014), are estimated for three future overlapping periods, namely, period 1 (2021-2050), period 2 (2046-2075), and period 3 (2071-2100). From these analyses, we estimate that relative increases in the observations for the spatial median 20-year (50-year) return level will be approximately 16.4% (16.5%) in the SSP2-4.5, 22.9% (22.8%) in the SSP3-7.0, and 37.6% (35.4%) in the SSP5-8.5 scenarios, respectively, by the end of the 21st century. The expected frequency of the reoccurring years, particularly for the AMP1 from 150 mm to 300 mm under the SSP5-8.5 scenario, are projected to increase by approximately 1.4 times that of the past 30 years for period 1, approximately 2.3 times that for period 2, and approximately 3.5 times that for period 3. From the analysis based on latitude, severe rainfall was found to be more prominent in the southern and central parts of the Korean peninsula.

# Future projections and uncertainty assessment of precipitation extremes in the Korean peninsula from the CMIP6 ensemble

Yonggwon Shin<sup>1</sup>, Yire Shin<sup>1</sup>, Juyoung Hong<sup>1</sup>, Maeong-Ki Kim<sup>2</sup>,  
Young-Hwa Byun<sup>3</sup>, Kyung-On Boo<sup>3</sup>, Il-Ung Chung<sup>4</sup>, Doo-Sun R. Park<sup>5</sup>,  
and Jeong-Soo Park<sup>1,\*</sup>

*1: Department of Statistics, Chonnam National University, Gwangju 500-757, Korea.*

*2: Department of Atmospheric Science, Kongju National University, Gongju, Korea.*

*3: Innovative Meteorological Research Department,  
National Institute of Meteorological Sciences, Seogwipo, Korea.*

*4: Dept of Atmospheric and Environmental Sciences,  
Gangneung-Wonju National University, Korea.*

*5: Dept. of Earth Sciences Education, Kyungpook National University, Daegu, Korea.*

*\*: Corresponding author, E-mail: jspark@jnu.ac.kr, Tel: +82-62-530-3445*

August 26, 2020

## Abstract

Projected changes in extreme climate are occasionally predicted through multi-model ensemble methods using a weighted averaging that combines predictions from individual simulation models. To predict future changes in precipitation extremes, observed data and 21 of the Coupled Model Inter-comparison Project Phase 6 (CMIP6) models are examined for 46 grids over the Korean peninsula. We apply the generalized extreme value distribution (GEVD) to the series of annual maximum daily precipitation (AMP1) data. Simulation data under three shared socioeconomic pathway (SSP) scenarios, namely, SSP2-4.5, SSP3-7.0, and SSP5-8.5, are used. A multivariate bias correction technique that considers the spatial dependency between nearby grids is applied to these simulation data. In addition, a model weighting approach that accounts for both performance and independence (PI-weighting) is employed. **In this study, we estimate the future changes in precipitation extremes in the Korean peninsula using the multiple CMIP6 models and PI-weighting method. In applying the PI-weighting, we suggest simple ways for selecting two shape**

parameters based on the chi-square statistic and entropy. Variance decomposition with the interaction term between the CMIP6 model and the SSP scenario is applied to quantify the uncertainty of projecting the future AMP1. Return levels spanning over 20 and 50 years, as well as the return periods relative to the reference years (1973–2014), are estimated for three future overlapping periods, namely, period 1 (2021–2050), period 2 (2046–2075), and period 3 (2071–2100). From these analyses, we estimate that relative increases in the observations for the spatial median 20-year (50-year) return level will be approximately 16.4% (16.5%) in the SSP2-4.5, 22.9% (22.8%) in the SSP3-7.0, and 37.6% (35.4%) in the SSP5-8.5 scenarios, respectively, by the end of the 21st century. The expected frequency of the reoccurring years, particularly for the AMP1 from 150 *mm* to 300 *mm* under the SSP5-8.5 scenario, are projected to increase by approximately 1.4 times that of the past 30 years for period 1, approximately 2.3 times that for period 2, and approximately 3.5 times that for period 3. From the analysis based on latitude, severe rainfall was found to be more prominent in the southern and central parts of the Korean peninsula.

**Keywords:** ANOVA; Climate change; Exceedance probability; Expected waiting time; Generalized extreme value distribution; Heavy rainfall; L-moments estimation; Return period;

## 1 Introduction

Heavy precipitation can have cascading effects on communities, infrastructure, agriculture, and livestock, as well as on economically and culturally important natural ecosystems. For example, extreme rain events can result in costly damage to wastewater treatment plants, culverts, and roads. Extreme precipitation can result in landslides and floods, accompanied with a loss of life and the deterioration of infrastructure. Thus, understanding and projecting heavy rainfall is of significant importance to climate change impact, adaptation, and vulnerability assessments.

Numerous studies have reported that extreme precipitation events have become more frequent during the past century, and even more often occurring over the 21st century [1, 2, 3, 4, 5, 6, for example]. A simplified and major reason for more frequent extreme rainfall is the following: Warming conditions mean more evaporation, which leads to more water vapor in the air. When rain-triggering conditions are favorable, more saturated air leads to heavier precipitation [7, 8]. This has been the case across some areas of the world during the past century, and is likely to be accelerated with increased global warming over the 21st century [9]. Some studies have projected that global warming leads to a higher intensity of precipitation and longer dry periods, for example, in Europe and Asia [10, 11, 12, 13].

Extreme rainfall occurs frequently over the Korean peninsula during the warm season, from June through September, in association with synoptic disturbances, typhoons, or convective

changes within the air masses over the region. Previous studies [14, 15, 16] have reported an increase in observed extreme precipitation in Korea. Lee et al.[17] predicted that the increasing changes in the future heavy rainfall across East Asia appear more distinctly in Korea at a local scale, which indicates a higher sensitivity of the Korean peninsula to global warming. It is thus crucial to project and assess the changes in extreme precipitation events in Korea under different scenarios.

Some authors [18, 19, 20, 21, 22, 23, 24] have predicted increasing changes in future extreme rainfall over Korea, using a single model, ensembles of regional climate models, or multiple Coupled Model Inter-comparison Project Phase 5 (CMIP5) models. In this study, we update the previous studies, based on the multiple CMIP6 models under the three shared socioeconomic pathway (SSP) scenarios, namely, SSP2-4.5, SSP3-7.0, and SSP5-8.5 [38].

Studies on the projection of future climate change have used ensembles of multiple climate simulations. Multi-model ensemble (MME) methods of climatic projection have been proven to improve upon the systematic bias and to have fewer general limitations that are typically associated with single simulation models. Among the many ensemble methods, model weighting or averaging is typically employed [25, 26, 27, for example]. Model averaging is a statistical method in which unequal or equal weights are assigned to those models. Despite some arguments, the equal weighting or “model democracy” [26] has been criticized because it does not take into account the performance, uncertainty, and independency of each model in constructing an MME [28, 29, 30, for example].

One typical unequal weighting scheme is giving more weights to those models that are more skillful and realistic for a specific process or application. This performance-based weighting method and its variants, including Bayesian model averaging (BMA), have been employed in many different studies. It has improved the accuracy of the projections and reduced the prediction uncertainty. However, it has been reported that only a few models often exhibit extremely high weights, and most others have very low weights [31, 24]. This phenomenon may be because some models are more “fit to” the observations for given applications than others, and thus, receive extremely high weights in a multi-model estimate of change [32]. Such an aggressive weighting based on performance will only be dangerous in the sense of an overfitting when observational uncertainty is large, and thus is not robust in quantifying the uncertainty.

In addition to the performance, some researchers have considered other criteria such as a model convergence [33], model independency [28, 34, 35, 36], and a semi-performance measure [37]. A weighting scheme that accounts for both the independence and performance simultaneously is called the PI-weighting. In this study, we employ PI-weighting to robustly quantify

uncertainty in MME. In applying the PI-weighting, one has to determine two shape parameters that control the strength of the weights. We suggest simple ways to determine these parameters based on the entropy and p-values of the chi-square statistic.

The remainder of this paper is structured as follows. The data construction and numerical models are described in Section 2. The statistical methods are briefly mentioned in Section 3. Section 4 describes the PI-weighting with computational details including simple ways for determining the shape parameters. The results of the model weights and projected future changes are presented in Sections 5 and 6, respectively. Section 7 describes the results of an uncertainty assessment and projection by latitude, based on an analysis of variance. In section 8, relative improvement of the PI-weighted method over the simple average is quantified using skill score and prediction variance. Discussions are then given in Section 9 followed by a summary of the paper in Section 10. Details including technical specifics, tables, and figures are provided in the accompanying supplementary material (hereafter referred to as the SM) file.

## 2 Data and Simulation models

Table S1 in the SM lists the 21 CMIP6 climate models used in this study. The considered scenarios are shared socioeconomic pathways SSP2-4.5, SSP3-7.0, and SSP5-8.5 [38]. Hereafter, we shorten the scenario names to SSP2, SSP3, and SSP5, although we use both versions interchangeably. To re-grid common grid points of  $1.5^\circ \times 1.5^\circ$ , the iterative Barnes interpolation scheme [39] was employed for the observations and simulation data from the 21 models. The Barnes technique produces a rainfall field on a regular grid from irregularly distributed rainfall observation stations.

The observations for the 42-year reference period (1973-2014) were obtained from the Korea Meteorological Administration. There are 64 and 27 observation stations in South and North Korea, respectively. Because the 27 stations (black circles in Figure 1) in North Korea are too sparse, re-gridding using Barnes interpolation may not capture the locality of severe rainfall there. For a better re-gridding in North Korea, we used the Asian Precipitation Highly Resolved Observational Data Integration Towards Evaluation (APHRODITE) reanalysis data [40] as auxiliary information. The APHRODITE data were used after a bias correction was applied by utilizing the data of nearby observation stations. Figure 1 depicts a map of the Korean peninsula, the spatial distribution of 91 rainfall observation stations, the 80 APHRODITE grid points in North Korea, and the final 46 grids used in this study.

The annual daily maximum precipitation (AMP1) in APHRODITE has serious bias in its mean and variance, as shown in Figure S1. We thus correct it using nearby observations using the quantile mapping technique [41]. Examples of time series plots of the observations, APHRODITE data, and the bias-corrected APHRODITE data near the observational stations are shown in Figure S1. The observations and bias-corrected APHRODITE data are used to construct a rainfall field over a  $1.5^\circ \times 1.5^\circ$  grid of 46 points using the Barnes method.

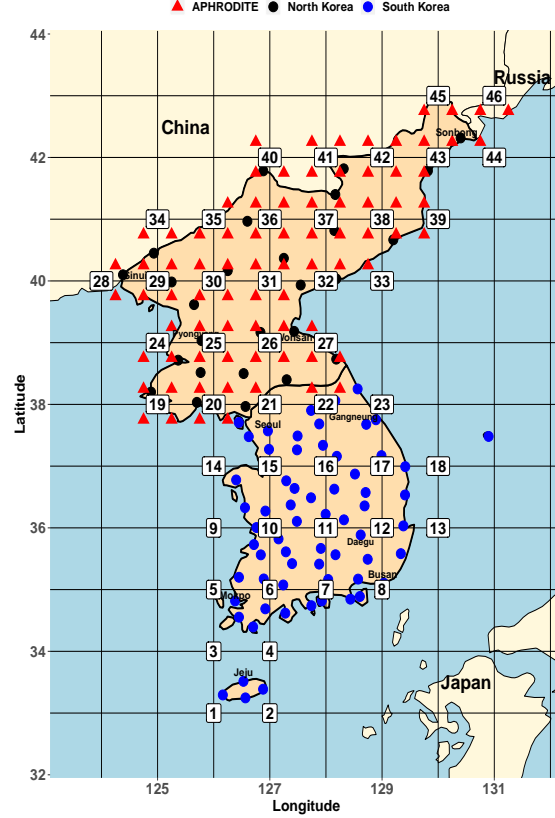


Figure 1: Map of the Korean peninsula from  $123^\circ$  to  $132^\circ$  longitude and  $33^\circ$  to  $43^\circ$  latitude including the sea and land, with 46 grid points of  $1.5^\circ \times 1.5^\circ$  for this study. The observations from 64 and 27 stations in South and North Korea, respectively, and the APHRODITE re-analysis data from over 80 grid points for North Korea are used to construct a rainfall field throughout 46 grids using the Barnes method.

## 3 Methods

### 3.1 Generalized extreme value distribution

The generalized extreme value distribution (GEVD) is widely used to analyze extreme univariate values. The three types of extreme value distributions are sub-classes of GEVD. The cumulative distribution function of the GEVD is as follows:

$$G(x) = \exp \left\{ - \left( 1 + \xi \frac{x - \mu}{\sigma} \right)^{-1/\xi} \right\}, \quad (1)$$

when  $1 + \xi(x - \mu)/\sigma > 0$ , where  $\mu$ ,  $\sigma$ , and  $\xi$  are the location, scale, and shape parameters, respectively. The particular case for  $\xi = 0$  in Eq.(1) is the Gumbel distribution, whereas the cases for  $\xi > 0$  and  $\xi < 0$  are known as the Fréchet and the negative Weibull distributions, respectively [42].

It can be helpful to describe the changes in extremes in terms of the changes in extreme quantiles. These are obtained by inverting the following (1):  $z_p = \mu - \frac{\sigma}{\xi} [1 - \{-\log(1 - p)\}^{-\xi}]$ , where  $G(z_p) = 1 - p$ . Here,  $z_p$  is known as the return level associated with the return period  $1/p$ , because level  $z_p$  is expected to be exceeded on average once every  $1/p$  years [42]. For example, a 20-year return level is computed as the 95th percentile of the fitted GEVD and a 50-year return level as the 98th percentile. Conversely to the above, the return period  $T(z) = 1/p$  for the given value  $z$  is obtained by calculating  $p = 1 - G(z)$ . For the given value  $z$ ,  $T(z)$  is sometimes called the expected waiting time, and the value  $p = 1 - G(z)$  is referred to as the exceedance probability of  $z$ .

### 3.2 Bias correction

Although simulations from climate or meteorological models provide significant information, the simulated data are associated with potential biases in that their statistical distribution differs from the distribution of the observations. This is partly because of unpredictable internal variability that differs from the observations, and because global climate models (GCMs) have a very low spatial resolution to be employed directly in most of the impact models [43, 44]. For example, in GCM precipitation fields, the bias may be due to errors in convective parameterizations and unresolved subgrid-scale orography [45]. Bias correction (BC) methods are commonly applied to transform the simulated data into new data with no or fewer statistical biases with respect to an observed time series. In this study, future simulation outputs are bias-corrected to compute the intensity of extreme rainfall.

To correct the model outputs more efficiently by taking account of the dependency among variables or nearby grids, several multivariate BC methods have recently been proposed. In this study, we employ the multivariate bias correction (MBC) method by Cannon [45] among the many available BC methods [41], which has an advantage of preserving the approximate trends of the model data. Because a regional frequency analysis is applied in this study, spatial dependency over nearby grids is considered. The MBC method is utilized by treating spatial data across nearby grids as  $k$ -dependent variables. More details are provided in the SM.

### 3.3 Performance and Independence weighting

Knutti et al.[35] argued that the growing number of models with different characteristics and considerable interdependence finally justifies abandoning a strict model democracy. They provided at least five reasons why PI-weighting is required. Brunner et al.[32] illustrated that PI-weighting leads to an increase in the investigated skill score for temperature and precipitation while minimizing the probability of an overfitting (or overconfidence).

As the basic idea of PI-weighting, models that agree poorly with observations for a selected set of diagnostics receive less weight, as do models that largely duplicate existing models [35]. Weights are calculated for each model based on a combination of the distance  $D_i$  (informing the performance) and the model similarity  $S_{ij}$  (informing the dependence):

$$w_i = \frac{\exp(-\frac{D_i}{\sigma_D})}{1 + \sum_{j \neq i}^M \exp(-\frac{S_{ij}}{\sigma_S})}, \quad (2)$$

with the total number of model runs  $M$  and the shape parameters  $\sigma_D$  and  $\sigma_S$ . The weights are normalized such that their sum equals 1.

The numerator represents the modeling skill when using a Gaussian weighting, where the weight decreases exponentially the farther away a model is from the observations. The denominator is the “effective repetition of a model” [28] and is intended to account for the model interdependency [35]. The details in computing the distance  $S_{ij}$  are given in the SM.

The shape parameters define the strength of the weighting and the relative importance of the performance and independence [32]. Large values will lead to an almost equal weighting, whereas small values will lead to aggressive weighting, giving a few models most of the weight. The shape parameters are often determined through a leave-one-out perfect model test using the continuous rank probability score [32, 36]. In the next subsections, we consider relatively simple ways to determine the shape parameters.



### 3.4 Selection of $\sigma_S$

To select an appropriate value of the shape parameter  $\sigma_S$  for the I-weights, we consider an entropy-based approach. Denote  $I_i(\sigma_S)$  as a normalized I-weight for model  $i$  and for the given  $\sigma_S$ , as defined in the following:

$$I_i(\sigma_S) = \frac{s_i(\sigma_S)}{\sum_{l=1}^M s_l(\sigma_S)}, \quad (3)$$

where  $s_i(\sigma_S) = \frac{1}{1 + \sum_{j \neq i}^M \exp(-\frac{S_{ij}}{\sigma_S})}$ .

The entropy of the I-weights as a measure of uncertainty [47] from these weights is defined by the following:

$$E(\sigma_S) = - \sum_{i=1}^M I_i(\sigma_S) \log I_i(\sigma_S) \quad (4)$$

as a function of  $\sigma_S$ . When all  $I_i(\sigma_S)$ s are almost equal, the entropy has a high value. We thus expect the entropy to increase because  $\sigma_S$  has a large value. Note that the calculation of  $S_{ij}$  does not depend on  $\sigma_S$ , and thus the  $S_{ij}$  values obtained are fixed for the entropy computation. The entropy is computed as  $\sigma_S$  changes from 0.1 to 1.0 by increments of 0.01.

Figure 2 presents the entropy function of  $\sigma_S$  computed from the data used for this study, which indicates that it is minimum at  $\sigma_S = 0.4$ . It is interesting to note that the entropy function increases as  $\sigma_S$  decreases from 0.4 to zero. This is explained by looking into the similarity measure  $1 + \sum_{j \neq i}^M \exp(-\frac{S_{ij}}{\sigma_S})$ . As  $\sigma_S$  moves toward zero, this measure converges at one for all  $i$ . Thus,  $s_i$  moves toward one, and  $I_i$  is close to  $1/M$  for all  $i$ . Because we want to have a shape parameter  $\sigma_S$  that can differentiate the I-weights most distinctly with minimum uncertainty, the value  $\sigma_S = 0.4$  minimizing the entropy is chosen in this study.

### 3.5 Selection of $\sigma_D$

To select an appropriate value of  $\sigma_D$  for the P-weights, we attempted to use the entropy criteria again, but were not fortunate enough to obtain the optimal result, as in  $\sigma_S$ . Thus, a technique based on the p-value of the chi-square statistic [48] is considered in this study.

Denote  $P_i(\sigma_D)$  as a normalized P-weight for model  $i$  and for the given  $\sigma_D$ , which is defined as follows:

$$P_i(\sigma_D) = \frac{\exp(-\frac{D_i}{\sigma_D})}{\sum_{l=1}^M \exp(-\frac{D_l}{\sigma_D})}, \quad (5)$$

where  $D_i$  is the performance measure of the  $i$ -th model.

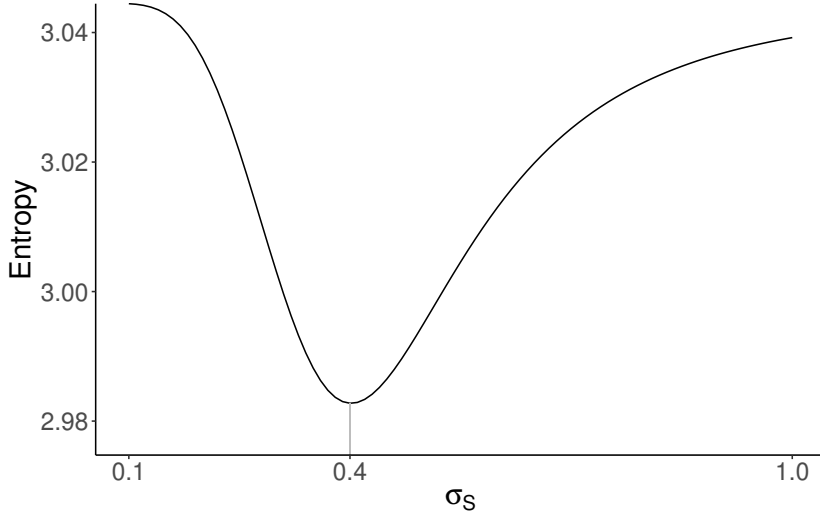


Figure 2: Plot of the entropy as  $\sigma_S$  changes from 0.1 to 1.0, and the selected  $\sigma_S = 0.4$ .

For testing the hypothesis frame, the null hypothesis and the alternative hypothesis for  $i = 1, \dots, M$  are as follows:

$$\begin{aligned} H_0 : \text{all weights are equal} &\Leftrightarrow P_i = \frac{1}{M} \text{ for all } i \\ H_1 : \text{some weights are not equal} &\Leftrightarrow P_i \neq \frac{1}{M} \text{ for some } i. \end{aligned} \quad (6)$$

For the given  $P_i$ , the chi-square statistic used to test the above hypothesis is as follows:

$$\chi_0^2(\sigma_D) = \sum_{i=1}^M \frac{(\frac{1}{M} - P_i(\sigma_D))^2}{\frac{1}{M}}. \quad (7)$$

Because we do not want to accept equal weights,  $\sigma_D$  should be selected to reject the null hypothesis. That is, the p-value [48] obtained from the chi-square statistic should be less than a preassigned value  $\alpha$  (significance level), e.g.,  $\alpha = 0.05$ . In addition, because we also do not want aggressive weights, a  $\sigma_D$  can be selected as the maximum value of  $\sigma_D$  in which we still reject  $H_0$  with  $\alpha$  level. That is, our selection is

$$\sigma_D^* = \max \{ \sigma_D : p\text{-value}(\sigma_D) < \alpha \}, \quad (8)$$

where  $p\text{-value}(\sigma_D) = Pr[\chi^2 > \chi_0^2(\sigma_D) | H_0]$ . Here,  $\chi^2$  indicates a random variable of (7) under the equal  $P_i$  weights. Although this selection assures the use of the least aggressive weights, it is still statistically significantly different from the equal weights. The details of computing  $D_i$  and the p-values are provided in the SM.

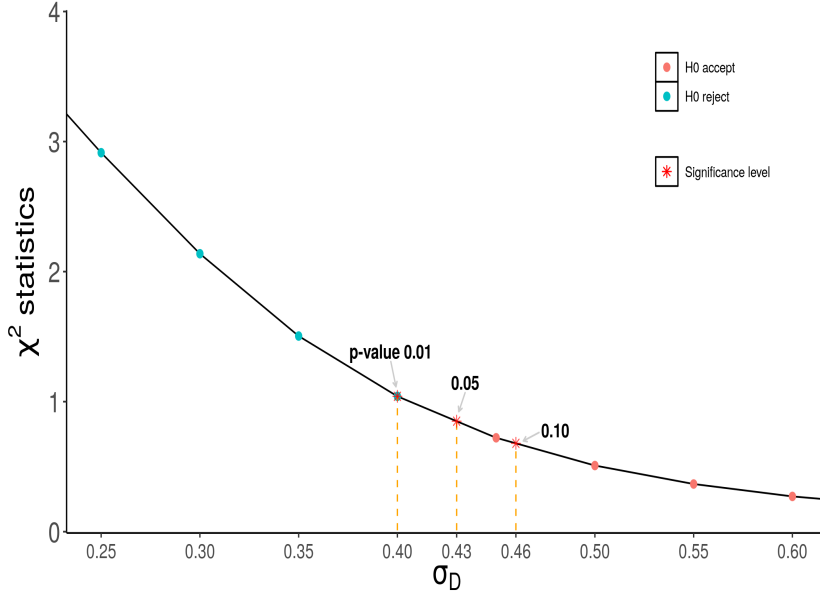


Figure 3: Plot of the chi-square statistic values as  $\sigma_D$  changes from 0.25 to 0.6, and the selected  $\sigma_D$ s corresponding to the p-values (0.01, 0.05, and 0.1).

Figure 3 depicts the chi-square statistic values computed from the data used for this study with some p-values as  $\sigma_S$  changes from 0.2 to 0.6 by increments of 0.05. When  $\alpha = 0.05$  as is usually applied in testing a hypothesis in statistics, the selected  $\sigma_D^*$  is 0.43. When  $\alpha = 0.01$  or 0.1,  $\sigma_D^* = 0.40$  or 0.46, respectively. We use  $\sigma_D^* = 0.43$  in this study.

## 4 Results: model weights

### 4.1 Model similarity

Table S2 provides the similarity values  $S_{ij}$  for certain models. Figure 4 shows the intermodel distance matrix for the 21 CMIP6 models considered in this study. Each box represents a pairwise combination, where red indicates a greater distance. According to Figure 4, FGOALS-g3, MRI-ESM2-0, INM-CM5-0, and MIROC6 were found to be the most independent, whereas CanESM5, ACCESS-CM2, UKESM, and KACE-1-0-G were found to be the most dependent. This result is confined only to the AMP1 over the Korean peninsula, and will differ for other variables and other regions.

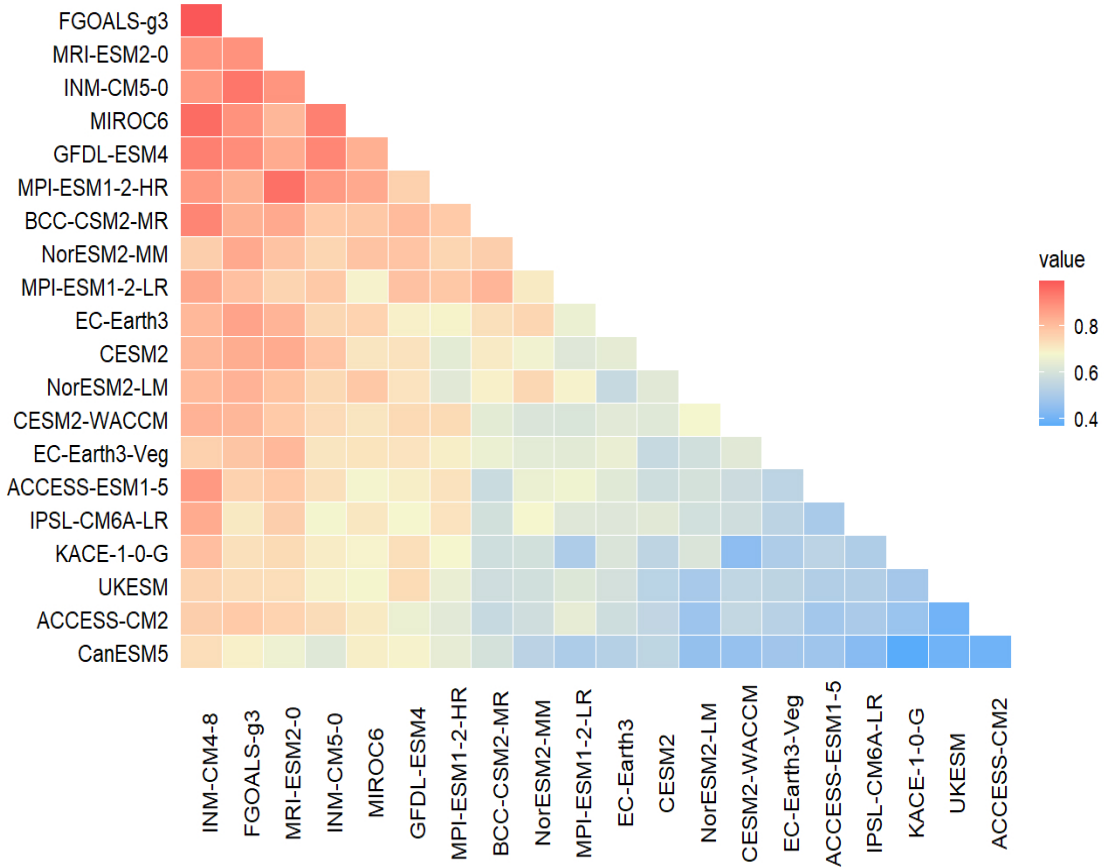


Figure 4: A graphical representation of the inter-model distance matrix for 21 CMIP6 models. Each box represents a pairwise combination, where red colors indicate a greater independence and blue colors indicate a greater similarity.

## 4.2 PI-weights

The normalized PI-weights are obtained using Eq.(2), with  $\sigma_S = 0.4$  and  $\sigma_D = 0.43$ . Figure 5 demonstrates the distributions of the P-, I-, and PI-weights. The variability of the I-weights is smaller than that of the P-weights.

The high P-weights of the KACE-1-0-G and ACCESS-CM2 models decrease in PI-weights owing to the low I-weights. The medium-low P-weights of FGOALS-g3, MRI-ESM2-0, and MIROC6 models increase in PI-weights owing to a high independency. The PI-weight is not located in the middle of the P- and I-weights but is close to the P-weight, except in a few cases. When the P-weight is almost the same as the equal weight in the CanESM5 model,

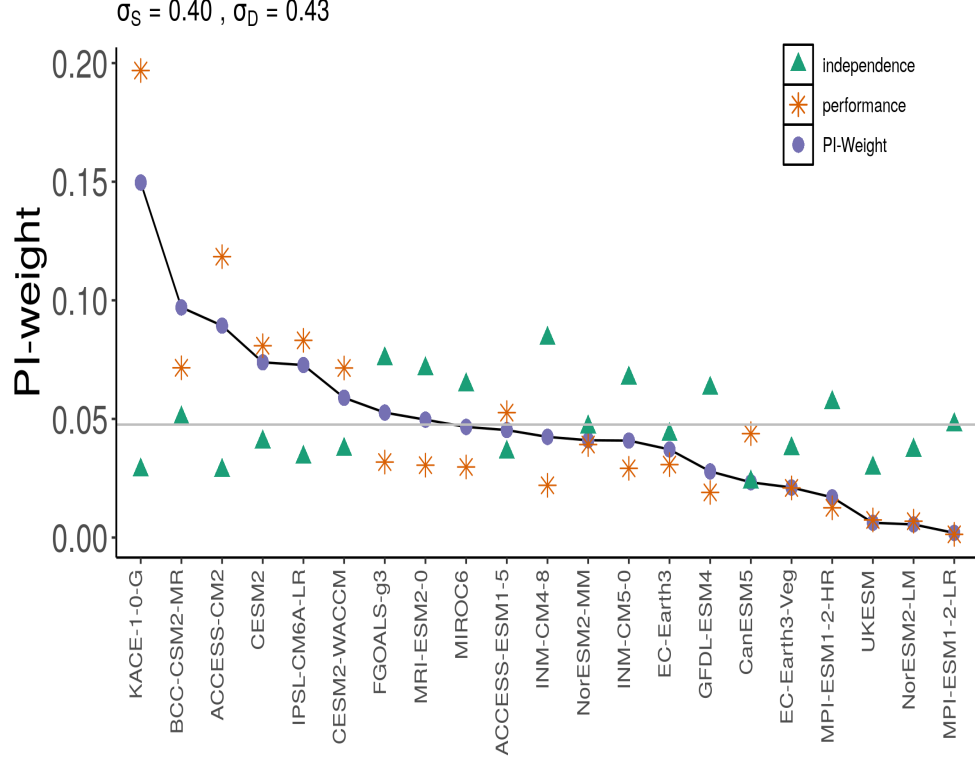


Figure 5: Spread of the weights for 21 CMIP6 models obtained based on the performance only, the independence only, and by both the performance and independence. The weights calculated for each grid are averaged over the Korean peninsula.

it seems that the PI-weight is wholly influenced by the I-weight. The performance for some of the models (EC-Earth3-Veg, MPI-ESM1-2-HR, UKESM, NorESM2-LM, and MPI-ESM1-2-LR) is so low that their I-weights do not affect the final weights. Based on these aspects, the performance is more influential than the independency of the PI-weights. Some of these observations may change if different  $\sigma_S$  and  $\sigma_D$  are used.

## 5 Results: Future projection of extreme precipitation

Using the PI-weights obtained in the above section, the future extreme precipitations are projected by the MME. Note that the future climate data are used after the bias correction

by the MBC method [45].

## 5.1 Return levels

Figure 6 displays boxplots of the 20-year return levels of the AMP1. The boxplot of the historical data after the BC is similar to that from the observations, whereas the boxplot before the BC is much smaller than that from the observations. The increasing trends from P1 to P3 are evident in every scenario. Summary statistics of the corresponding values of these boxplots are provided in Table S3.

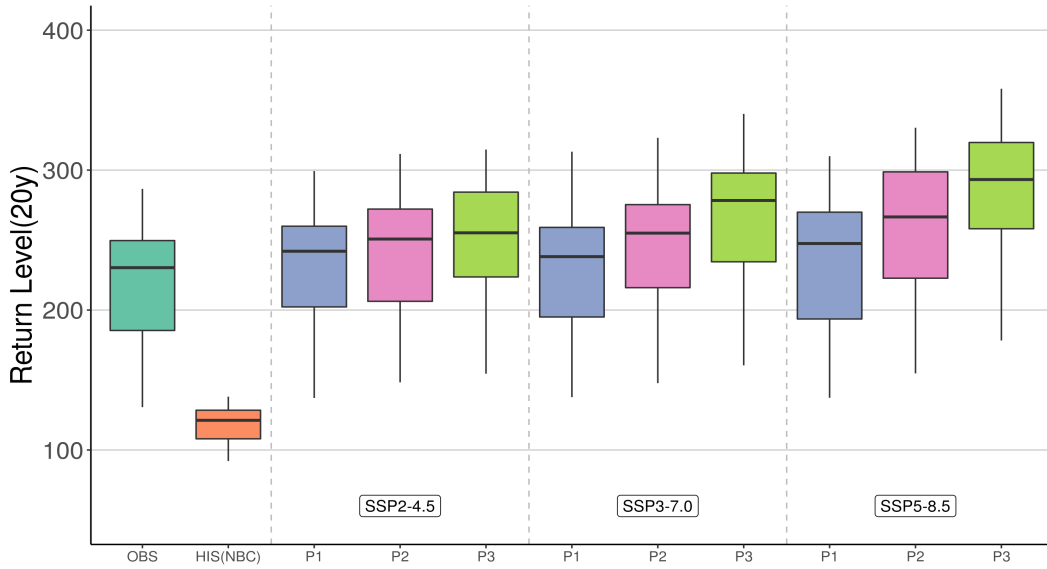


Figure 6: Schematic box-plots of 20-year return levels (unit: mm) of the annual maximum daily precipitation averaged over 46 grids in the Korean peninsula for the future periods, i.e., P1 (2021-2050), P2 (2046-2075), and p3 (2071-2100), under the SSP2-4.5, SSP3-7.0, and SSP5-8.5 scenarios. OBS and HIST(NBC) indicate the observations and the historical data without a bias correction. The box-plot for p0 represents the bias-corrected historical data.

These values are lower (similar or higher) in the 20-year (50-year) return levels than the results by Lee et al.[24], who used the BMA method with the CMIP5 models. The median values for P3 in Table S3 are higher than the results by Shin et al.[37], who used the CMIP6 models and a hybrid weighting method between the BMA and equal weighting. It seems that the hybrid weighting by Shin et al.[37] and the PI-weighting result in some of the CMIP6

models with heavy future precipitation receiving lower weights than those models receive from a weighting method based on the performance only.

Figure 7 shows isopluvial maps of 20-year return levels of the AMP1 for the three future periods under the three scenarios. The difference by latitude is more evident than that by longitude. Jeju island, the southern coast, and the central region receive more downpours than the northern parts.

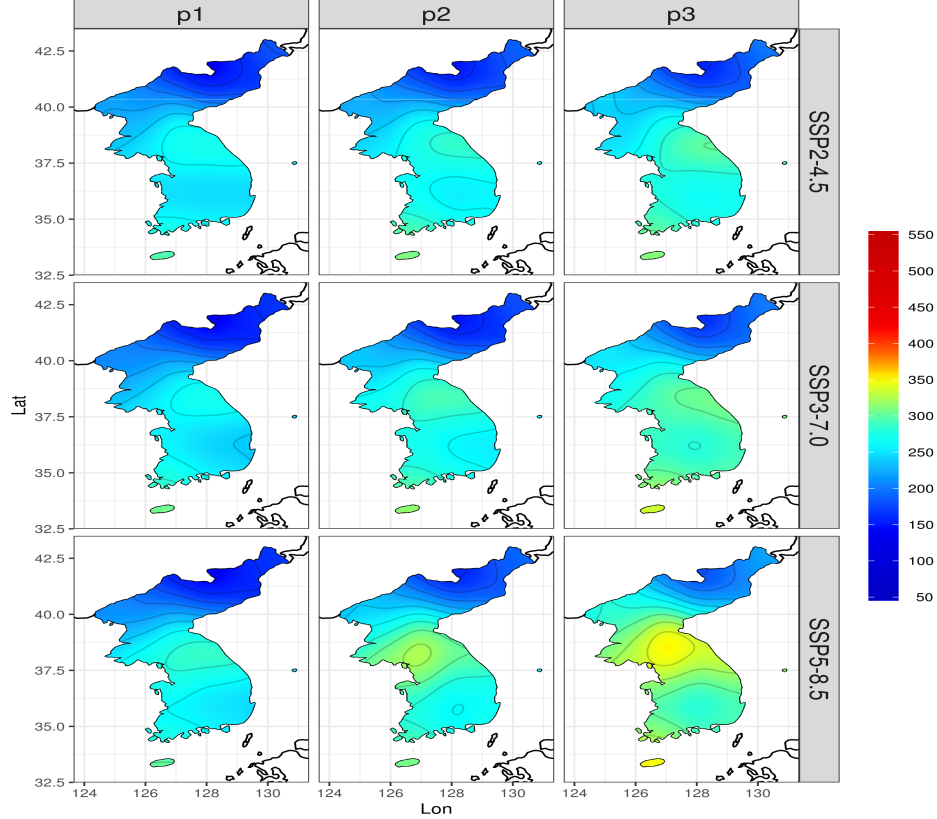


Figure 7: Isopluvial maps of 20-year return levels (unit: mm) of the annual maximum daily precipitation for 46 grids over the Korean peninsula for the future periods; P1 (2021-2050), P2 (2046-2075), and P3 (2071-2100) under the SSP2-4.5, SSP3-7.0, and SSP5-8.5 scenarios.

## 5.2 Changes in return levels

Figure S5 exhibits relative changes (unit: %) of the 20- and 50-year return levels of extreme precipitation over the Korean peninsula relative to the years 1973–2014. The increasing changes

are more evident in the northern part than in the southern area. Summary statistics of relative changes are presented in Table S4 in the SM. The relative increase in the observations for spatially averaged 20-year (50-year) return level was approximately 16.4% (16.5%) in the SSP2, 22.9% (22.8%) in the SSP3, and 37.6% (35.4%) in the SSP5 scenario by the end of the 21st century.

These rates of change are lower than the results by Lee et al.[24], who used the BMA method with the CMIP5 models. This is perhaps due to the differences in the reference period and the research methods, and the difference between the CMIP5 and CMIP6. For the 20-year return level, our result is approximately 1.6 times faster than the changes in the globally averaged value (10% in the SSP2-4.5 and 20% in the SSP5-8.5) reported by Kharin et al.[49].

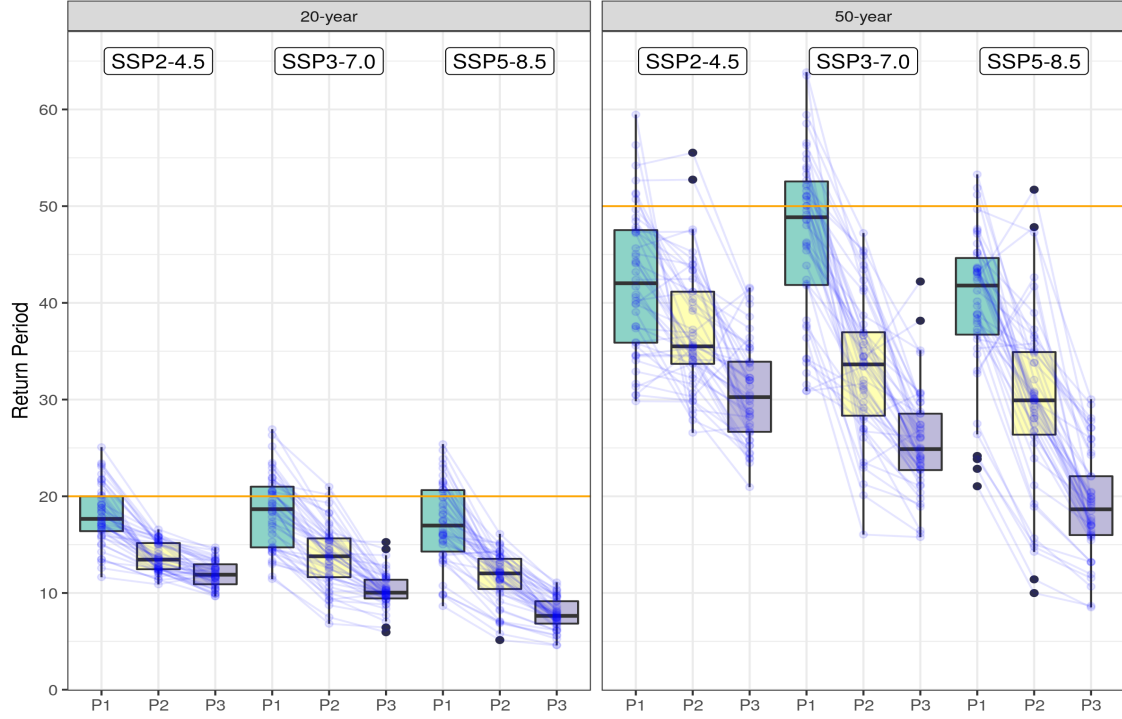


Figure 8: Parallel coordinated box-plots, similar to those in Figure 6 but for 20- and 50-year return periods relative to the observations from 1973 to 2014.



### 5.3 Change in return periods

Figure 8 displays boxplots for the 20-year and 50-year return periods, as compared to the reference years (1973–2014) for the three future periods under the three scenarios. The corresponding statistics are presented in Table S5. We realize that a 1-in-20 year (1-in-50 year) AMP1 in the Korean peninsula will likely become 1-in-12 (1-in-30) year, 1-in-10 (1-in-25) year, and 1-in-8 (1-in-19) year events in the median by the end of the 21st century based on the SSP2, SSP3, and SSP5 scenarios, respectively, as compared to the observations from 1973 to 2014. These findings indicate that both 20-year and 50-year return periods are likely to decrease by approximately 40% under the SSP2 and 62% under the SSP5 by the end of the 21st century.

These projections of extreme rainfall are less frequent than in the results obtained by previous studies [20, 24], i.e., approximately 1-in-11 (1-in-21) year and 1-in-7 (1-in-13) year events under SSP2 and SSP5, respectively. Shin et al.[37] realized the occurrence as approximately 1-in-10 (1-in-30) and 1-in-8 (1-in-17) year events under SSP2 and SSP5, respectively, which are similar or more frequent than our projection.

### 5.4 Exceedance probability and waiting time

Because of computational issues and defects of the return period [50], the exceedance probability as an alternative to the return period is often used [53]. This is defined as  $Pr[Y(\hat{\theta}) > z]$ , where  $z$  is a specified precipitation value, and  $Y(\hat{\theta})$  is a random variable following a GEVD with a parameter estimate  $\hat{\theta}$ . Here,  $Y(\hat{\theta})$  depends on the models, periods, and scenarios.

The spatially averaged estimates of the exceedance probability are presented in Figure S6 and Table S6. There are relatively large differences in the exceedance probability of a downpour of 100 *mm* to 250 *mm* compared with that for over 250 *mm*, as shown in Figure S6. The differences between the past and future scenarios are distinct during the period P3.

From Table S6, the return period or expected waiting time ( $T(z)$ ) until the reoccurrence of a specific AMP1 value ( $z$ ) is computed by  $T(z) = 1/p(z)$ , where  $p(z)$  is the exceedance probability of the AMP1  $z$ . These values are listed in Table S7. For  $z = 200$  *mm* of rainfall as an example, the expected waiting times until a reoccurrence are 20.8 years in the past, 15.9 years in the future period P1, 9.1 years in P2, and 6.2 years in P3 based on the SSP5 scenario. For the case of a  $z = 300$  *mm* downpour, the expected waiting times are 118.8 years in the past, 99.4 years in P1, 53.2 years in P2, and 30.5 years in P3 based on the SSP5 scenario.

## 5.5 Expected number of reoccurring years

Another quantity we can obtain is the expected number of reoccurrences during a certain period. By multiplying 30 years to the exceedance probability  $p(z)$ , one can estimate the expected frequency of such years, during 30 years, that we have more than  $z$  amount of AMP1 for a year. These values are given in Table S8. For  $z = 150 \text{ mm}$  ( $200 \text{ mm}$ ) as a specific example, during the past 30 years, we have experienced 4.5 (1.4) years in which AMP1 was greater than  $150\text{mm}$  ( $200\text{mm}$ ). In addition, we are likely to have an expected number of years of 7.1 (1.9) for the future period P1, 9.4 (3.3) for P2, and 11.5 (4.8) for P3 under the SSP5 scenario.

From this comparison, particularly for AMP1 of  $150 \text{ mm}$  to  $300 \text{ mm}$ , the expected number of years of occurrence for the future periods under the SSP5 scenario increase by approximately 1.4 times that over the past 30 years for P1, 2.3 times that for P2, and 3.5 times that for P3, by the end of 21st century. These results are based on the spatially averaged values. When the exceedance probability is considered for each grid, locally different results will be obtained.

## 6 Results: Quantifying uncertainty and projection by latitude

Three or four major sources of climate projection uncertainty might be considered when trying to understand uncertainties in projected metrics: (1) climate model, (2) emission scenario, and (3) internal variation or randomness unexplained by other sources [51]. We apply the analysis of variance (ANOVA) technique [52] in this study. Other methods considered in [53] can also be applied.

### 6.1 Variance decomposition with interaction

Hawkins and Sutton [51] applied a method of variance decomposition to quantify uncertainty when assuming no interactions between the sources. In addition to their assumption, we added the interaction term between the model and scenario. Other factors, i.e., (4) future period and (5) location, such as the latitude applied in this study, are also considered in ANOVA modeling. Although the period and latitude may not be real sources of uncertainty, they can be included as independent variables in the ANOVA model because they may affect the response variable. Here, the response variable is a 20-year return level, for example.

Figure 9 presents the proportions of variances contributed by each variable for each period. In P1, the variance by the model is the largest, whereas that by the scenario is the smallest. The proportion of internal variation (residual) decreases from P1 to P3. The variance proportion by

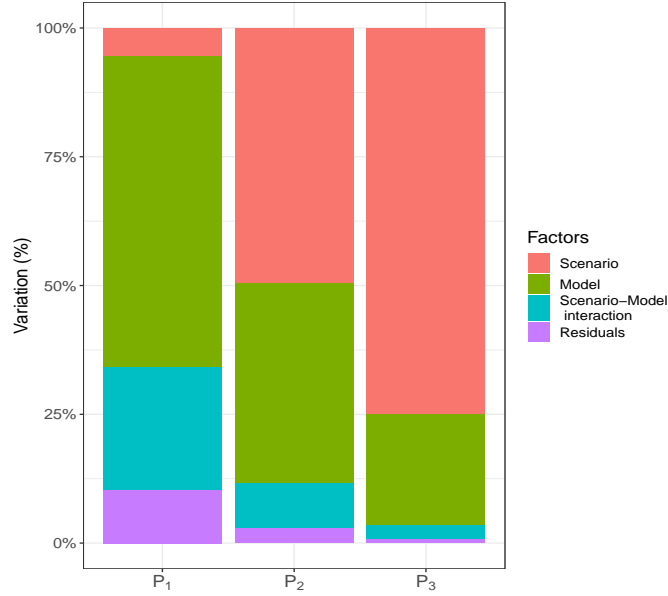


Figure 9: Bar plots (unit: %) representing uncertainties in projecting future climate or the variations owing to the models, scenarios, model and scenario interaction, and residuals for 20-year return levels over the Korean peninsula, for each period.

model decreases from P1 to P3, whereas the variation due to the scenario increases significantly from P1 to P3.

It is notable that the variation contributed to by the interaction between the model and scenario is relatively large during period 1 but becomes smaller during period 3. The details of this variation are presented in Figure S7, where the return levels from the models generally increase in mean and variance as the scenario changes from SSP2 to SSP5. However, the return levels of some models such as CESM2-WACOM, ACCESS-ESM1-5, NorESM2-MM, MiCRO6, MRI-ESM2-0, and GFDL-ESM4 decrease from SSP2 to SSP3, contrary to the expectation. Moreover, the return values of some models such as MPI-ESM1-2-HR, MPI-ESM1-2-LR, and NorESM2-LM decrease from SSP3 to SSP5. The latter two ‘unexpected’ situations make the interaction variance large in P1, as shown in Figure 10. Despite such ‘unexpected’ situations still appearing in P3, the variations owing to the scenario and model themselves are relatively too large, and thus the proportion of the interaction variance becomes relatively small in P3.

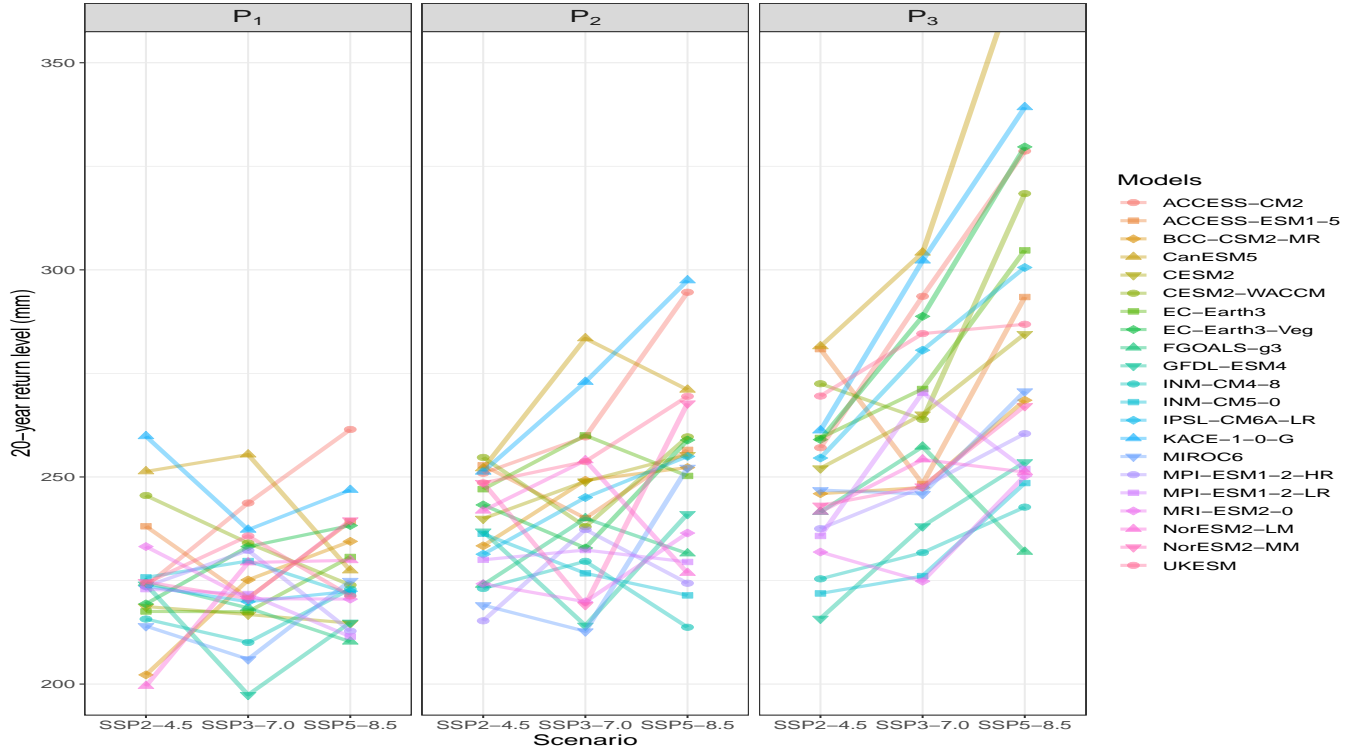


Figure 10: Interaction plots of 20-year return levels from 21 CMIP6 for three SSP scenarios for each period (P<sub>1</sub>, P<sub>2</sub>, and P<sub>3</sub>).

## 6.2 Return levels by latitude

Figure 11 depicts boxplots of the 20-year return levels (unit: mm) from 21 CMIP6 as the latitude changes from south to north. It is evident that the return levels decrease initially from  $33^\circ$  to  $36^\circ$  but increase from  $36^\circ$  to  $38^\circ$ . However, they decrease from  $38^\circ$  to  $42^\circ$  more rapidly. In Figure 11, the solid lines around the boxes across the latitude are the spatial medians of the 20-year return levels for each latitude and four periods (P<sub>0</sub>, P<sub>1</sub>, P<sub>2</sub>, and P<sub>3</sub>). Here, P<sub>0</sub> indicates the reference period, and thus the result is obtained from the observations. The dashed lines depict those values for three scenarios (SSP2-4.5, SSP3-7.0, and SSP5-8.5), in which three dashed lines locate the inside range of the solid lines. The spatial medians for the scenario (period) are obtained over all values across the periods (scenarios) over 21 models. It is notable and interesting that the variation owing to the scenario is smaller than that from the period. Future projections are less fluctuating than the observations.

The detailed values of the 20- and 50-year return levels by latitude are presented in Table

Table 1: 20- and 50-year return levels (unit: mm) averaged over each latitude for the observations (OBS) during 1973–2014 and for future periods, namely, P1 (2021–2050), P2 (2046–2075), and P3 (2071–2100). For each future period, the values were averaged over the three scenarios.

	20 year return level				50 year return level			
Latitude	OBS	$P_1$	$P_2$	$P_3$	OBS	$P_1$	$P_2$	$P_3$
33	292	303	319	345	352	361	379	407
34	283	288	303	329	338	340	358	388
35	254	262	279	304	306	316	336	369
36	218	246	268	290	278	292	320	345
37	218	258	276	304	253	315	338	374
38	263	262	281	307	326	317	337	380
39	229	245	267	287	278	298	328	354
40	170	213	232	252	212	268	296	325
41	161	174	190	211	210	215	239	269
42	138	159	175	193	176	199	219	242
43	171	171	190	207	215	215	237	254

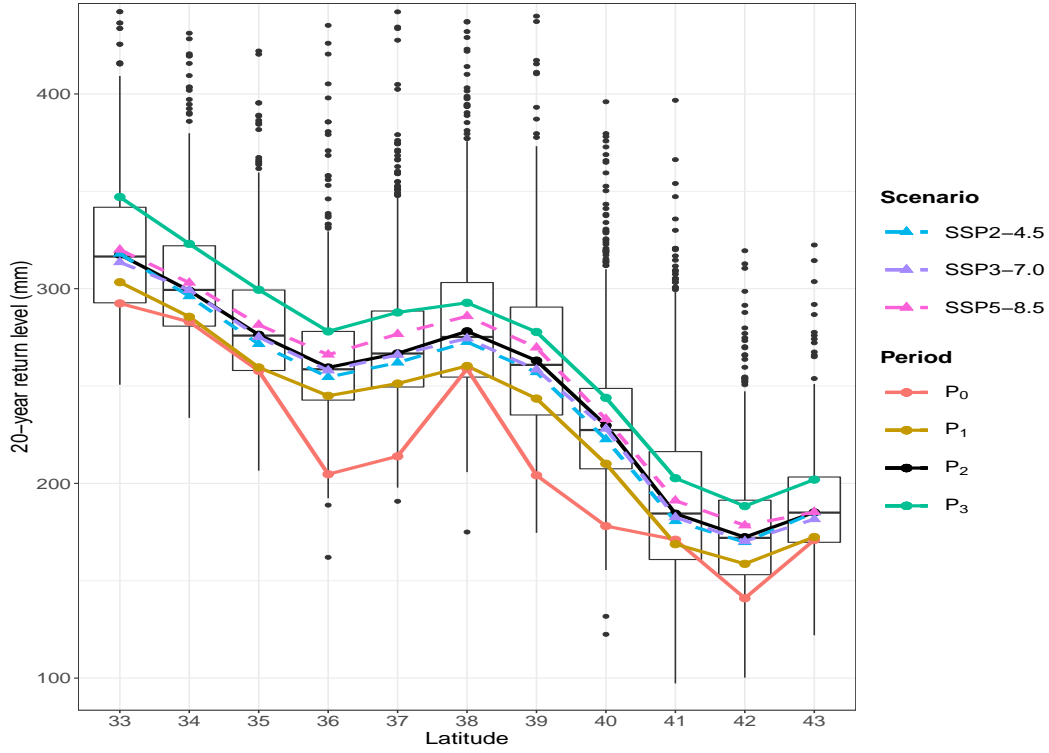


Figure 11: Box-plots of 20-year return levels (unit: mm) from 21 CMIP6 as the latitude changes from  $35^{\circ}$  to  $42^{\circ}$ . The solid lines around the boxes across the latitude are the spatial medians of 20-year return levels for each latitude and for P0 (1973–2014, i.e., the observation period), P1 (2021–2050), P2 (2046–2075), and P3 (2071–2100). The dashed lines depict those values for three scenarios (SSP2-4.5, SSP3-7.0, and SSP5-8.5).

1. These localized values provide different information from the previous results based on the spatial medians and averages over the Korean peninsula (Table S3). Table 1 indicates that a much higher intensity occurs in the southern part of the Korean peninsula than in the northern part. When the areas below and above  $38^{\circ}$  are considered the south and north, respectively, the south is likely to receive approximately 1.4 (1.18) times heavier rainfall on average than the north (the spatial median), based on a rough calculation using the above tables. Although we did not obtain these values, localized values of the return periods, relative changes in the return levels, the exceedance probabilities, and the expected number of occurring years can be obtained for each latitude. Those values in the south would show higher intensity during extreme precipitation with shortened expected waiting times compared to the spatial medians

throughout the Korean peninsula.

## 7 Comparison of PI-weighted ensemble to simple average

For comparison between the PI-weighting scheme and the simple average for multimodel ensemble, we consider two measures. The first one is the error index  $I^2$  based on Baker and Taylor [54] for the reference (historical) period. The second is the weighted variance of return level prediction [55, 56] for the future periods.

### 7.1 Error index for historical period

The error index can measure combined errors, that is, over multiple variables, compared to the observed climate [36]. However, we only evaluate our target variable AMP1 over the reference period 1973-2014. In a first step, the normalized error is calculated for T-year return level as the difference between PI-weighted multimodel and the observations;

$$e_w^2 = \sum_{n=1}^N \frac{(S_n - o_n)^2}{\sigma_n^2}, \quad (9)$$

where  $S_n$  is the weighted multimodel T-year return level per grid point  $n$ .  $S_n$  is calculated from the historical data which are not bias corrected.  $o_n$  is the T-year return level obtained from the observations, and  $\sigma_n$  is the standard deviation of  $o_n$ . The corresponding  $e_{eq}^2$  is calculated for the nonweighted multimodel T-year return level. Note that, for each  $n$ ,  $o_n$  and  $\sigma_n$  are unchanged between  $e_w^2$  and  $e_{eq}^2$  because those are computed from the observations on grid point  $n$ . Then the error index is obtained by:

$$I^2 = \frac{e_w^2}{e_{eq}^2}. \quad (10)$$

This index can be useful to evaluate if the weighted mean is improved compared to the simple average [36]. The smaller  $I^2$  the larger in the improvement in the weighted average compared to the simple average for the target variable (AMP1) for the reference period.

The computed values of  $I^2$  in this study are 0.535 for PI-weights and 0.498 for P-weights for 20-year return level. Thus, there are 46.5% and 50.2% improvement in the PI-weighted and the P-weighted averages, respectively, compared to nonweighted average. The gain of error index by adding I-weights from P-weights is 0.037, which is 7.43% gain relative to that of the P-weights. The results for 50-year return level are similar to these.

## 7.2 Prediction variance for the future

The second measure to compare is from the weighted variance formula [55, 56]:

$$Var(r(T)) = \sum_{k=1}^K [r_k(T) - \bar{r}(T)]^2 w_k + \sum_{k=1}^K Var(r_k(T)) w_k, \quad (11)$$

where  $r(T)$  is the T-year return level,  $r_k(T)$  is the T-year return level from  $k$ -th model, and  $\bar{r}(T) = \frac{1}{K} \sum_{k=1}^K r_k(T) w_k$ . The first term of (11) is the among-model variance, and the second one is within-model variance.  $Var(r_k(T))$  is estimated using a bootstrap technique in this study. Based on this variance, we quantify the skill of PI-weighted method by the Brier skill score [48]:

$$BSS = 100 \times \left\{ 1 - \frac{Var_{PI}}{Var_{Eq}} \right\}, \quad (12)$$

where  $Var_{PI}$  is the variance of T-year return level calculated by (11) from the PI-weights, and  $Var_{Eq}$  is the variance calculated by (11) from the equal weights, for each grid point. BSS is a relative quantity and shows a percentage improvement of the PI-weighted method over the simple average.

Figure 12 shows the differences of variances of 20-year return levels calculated from the PI-weights and the equal weights, plotted for 46 grid points for each future period and for each scenario.  $Var_{PI}$  for Y-axes is the variance calculated by (11) from the PI-weights.  $Var_{Eq}$  for X-axes is the corresponding variance calculated from the equal weights. In the top panel, more points locate below the diagonal lines, which mean the variances calculated from the PI-weights are more often smaller than those from the simple average. The middle (bottom) panel show the differences of between-model (within-model) variances.  $Var_{BM,PI}$  and  $Var_{WM,PI}$  for Y-axis are between-model and within-model variances from the PI-weights, respectively, and  $Var_{BM,Eq}$  and  $Var_{WM,Eq}$  for X-axis are the corresponding variances calculated from the equal weights. Within-model variances from both weights are similar, as seen in the bottom panel. Thus, we know that the differences in total variances in the top panel are mainly because of the differences of between-model variances.

Table 2 reads the averaged values of BSS over 46 grid points, the averaged differences of variances (unit:  $mm^2$ ) of 20-year return level, and relative improvements (RI) for three periods and for three scenarios. The BSS values range from 1.7% to 24.6%. The averages of BSS over all periods are 7.5% for SSP2, 13.1% for SSP3, and 15.9% for SSP5 (12.2% overall). These averages of BSS also increase as the period moves from P1 to P3. BSS due to between-model are bigger than those due to within-model. Watching three columns from 6th to 8th of this table, we know that the positive differences of two variances ( $Var_{Eq} - Var_{PI}$ ) are mainly



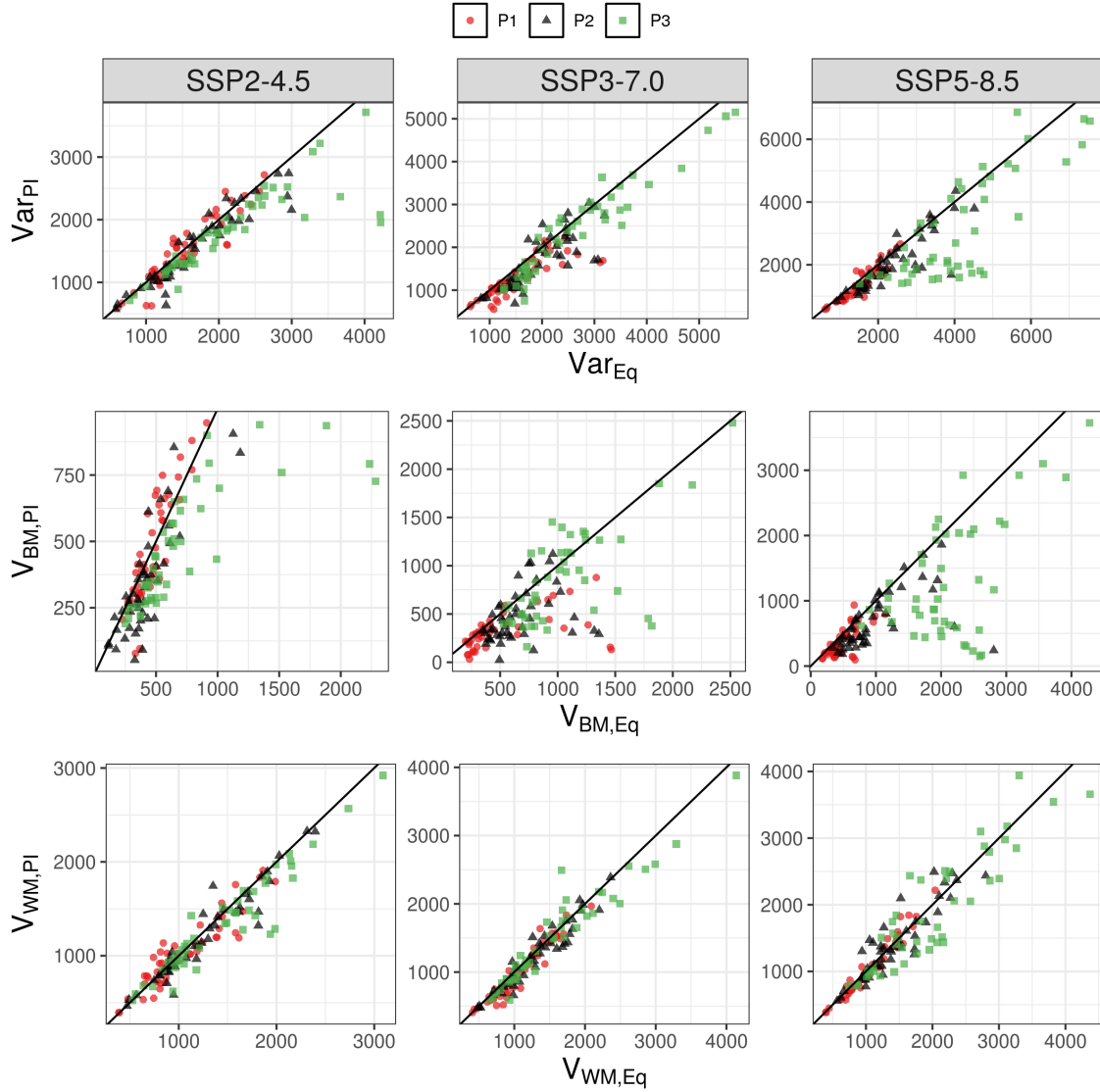


Figure 12: Scatter plot showing the differences of variances of 20-year return levels calculated from the PI-weights and the equal weights, plotted for 46 grid points for each future period and for each scenario.  $Var_{PI}$  for Y-axes (top panel) is the variance calculated by (11) from the PI-weights, and  $Var_{BM,PI}$  (middle panel) and  $Var_{WM,PI}$  (bottom) are between-model and within-model variances, respectively.  $Var_{Eq}$ ,  $Var_{BM,Eq}$ , and  $Var_{WM,Eq}$  for X-axis are the corresponding variances calculated from the equal weights. The value on each point in the top panel is a sum of those in the middle and in the bottom panels.

Table 2: The averaged values of the Brier skill scores (BSS), averaged differences of variances (unit:  $mm^2$ ) of 20-year return level over 46 grid points, and relative improvements (RI) of the PI-weighted method over the simple average (last column) for three periods and for three scenarios.  $RI = \frac{V_{Eq} - V_{PI}}{V_{PI}} \times 100$ .  $DV = Var_{Eq} - Var_{PI}$ ,  $DV_{BM} = Var_{BM,Eq} - Var_{BM,PI}$ , and  $DV_{WM} = Var_{WM,Eq} - Var_{WM,PI}$ . Other acronyms are same as Figure 12.

ssp	p	BSS(%)	$BSS_{BM}$	$BSS_{WM}$	DV	$DV_{BM}$	$DV_{WM}$	RI(%)
SSP2-4.5	P1	1.7	2.3	0.1	21	2	19	1.4
	P2	7.2	18.1	3.7	127	76	51	8.2
	P3	13.6	27.9	6.1	343	235	108	18.9
SSP3-7.0	P1	15.2	28.2	8.1	273	195	78	20.6
	P2	13.9	27.3	5.4	266	197	69	15.8
	P3	10.3	16.9	3.3	246	184	62	10.3
SSP5-8.5	P1	9.5	26.0	2.2	135	124	11	9.7
	P2	13.7	29.0	0.7	321	304	17	15.6
	P3	24.6	38.8	5.3	965	848	117	30.5

because of the differences of two between-model variances ( $Var_{BM,PI} - Var_{BM,Eq}$ ). These patterns are prominent for SSP5-8.5 scenario with increasing values from P1 to P3. The last column reads the RI of the PI-weighted method over the simple average in the variances. The RI values range from 1.4% to 30.5%. The averages of RI over all periods are 9.5% for SSP2, 15.6% for SSP3, and 18.6% for SSP5 (14.6% overall). These averages of RI also increase as the period moves from P1 to P3.

## 8 Discussion

It is generally accepted that increasing greenhouse gases induces atmospheric temperature warming, which results in increasing equivalent potential temperatures and specific humidity according to the Clausius-Clapeyron relationship [57]. The increase in atmospheric water vapor is the main factor in generating convective instability. In view of this insight, Kim et al.[23] obtained an indication that increasing the extreme precipitation over South Korea in the past and future scenarios is related more to a change in convective instability rather than to synoptic conditions. Another plausible explanation for the increase in the maximum precipitation over the Korean peninsula is the increase in both the frequency and strength of typhoons in the

region. Typhoons have a greater influence on the southern area than the northern part. The occurrence of extreme rainfall in the southern part is related to more sources or variables than in the northern area [22, 23].

We observed a scale discrepancy between the model grids and the observation stations, which may compromise the credibility of our results. Regional projections require a fine resolution of the simulation models, whereas some of the CMIP6 models have a coarse resolution. It is known that the model-spread is one of the major sources of uncertainty in regional predictions (Hawkins and Sutton, 2009). However, large numbers of simulation models can reduce the uncertainty and provide a robust projection [58, 59]. Twenty-one CMIP6 models with different scales used in this study can cover the study region with a finer spatial structure to increase the reliability of the projection.

The daily precipitation data consist of measurements from 00h to 24h throughout the day. In daily observations, the rainfall does not accumulate between 22 and 02 hours. In the data used in this study, such precipitation is divided and recorded into two separate days. The actual serious daily risk due to heavy rainfall does not exactly depend on the precipitation over the time duration from 00h to 24h exclusively. It is therefore recommended to consider the AMP1 data based on the maximum precipitation during the 24 hour movement. In this sense, the results presented in this study underestimate the actual intensity and frequency of AMP1. More realistic daily data such those as obtained after moving for 24 hour and the annual maximum 2 (and several) days of precipitation should be used in a future study for a risk assessment owing to extreme rainfall.

Some studies [36, 29, 32] considered several variables and multiple diagnostics to determining the weights, whereas only one variable (AMP1) and several return levels are applied in this study. Thus, the results presented here may be overconfident. We may need to employ more variables and multiple diagnostics in the future study.

Brunner et al.[32, 60] considered multiple observational (or reanalysis) datasets to include an estimate of the observational uncertainty. They proposed a novel approach to account for the observational spread and uncertainty in a multi-model weighting study, which can lead to robust results and a more precise uncertainty quantification. In addition, considering multiple observational datasets may address the problem in which the BC and performance-based weighting scheme utilizes an excessive number of observations. We believe that using the observations twice in the BC and weight calculation is unadvisable. Xu et al.[31] considered a Bayesian weighting method that removes observations during the initial phase of the down-scaling and adds them in the estimation of the posterior distribution. However, if the series

of observations is sufficiently long to divide into two parts, we may use one part for the BC and the other part for the weight calculation. Although we did not apply these methods, this would be a good approach in a future study.

In calculating the performance weights, we used the standardized return levels as described in the SM and determined through  $\sigma_D$ . However, as a different approach, the BMA method described in [24, 37, 56] can be employed. The error index  $I^2$ , considered in the above subsection, for the BMA method is 0.33 which is smaller than those for the P-weights and the PI-weights methods. The BMA weighting does not need to standardize the return levels or determine the shape parameter  $\sigma_D$ . It simply requires a bootstrap to estimate the variances of the return levels, which is straightforward in current computing facilities. Thus calculating the P-weights based on the BMA, with an effort to minimize the probability of overconfidence, is recommendable.

## 9 Summary

We estimated the future changes in precipitation extremes within the Korean peninsula using observations, 21 multiple CMIP6 models, the generalized extreme value distribution, the multivariate bias correction technique, and the model weighting method (PI-weighting) which account for both the performance and independence of the models.

In applying the PI-weighting method, we suggest two ways of selecting two shape parameters, based on the p-value of the chi-square statistic and entropy. The suggested methods are simple and intuitively appealing, although they may need more justification for use in other studies.

From the analysis described this study, we realized that a 1-in-20 year (1-in-50 year) annual maximum daily precipitation within the Korean peninsula will likely become a 1-in-12 (1-in-30) year, a 1-in-10 (1-in-25) year, and a 1-in-8 (1-in-19) year event in terms of the median by the end of the 21st century under the SSP2-4.5, SSP3-7.0, and SSP5-8.5 scenarios, respectively, as compared to the observations from 1973 through 2014. These results are similar or less frequent than those obtained by previous studies [20, 24, 37], but still, predict more frequent and intensified extreme precipitation events by the end of the 21st century as compared to 1973 through 2014.

The expected frequency of the reoccurring years, particularly for AMP1 from 150 *mm* to 300 *mm* under the SSP5 scenario, is projected to increase by approximately 1.4 times that of the past 30 years for period 1 (2021–2050), approximately 2.3 times that for period 2 (2046–

2075), and approximately 3.5 times that for period 3 (2071–2100).

From the analysis based on latitude, we found that extreme rainfall is more prominent in the southern and central parts of the peninsula. The downpour in the southern part is approximately 1.4 times heavier than that of the northern part and approximately 1.18 times that of the spatial median of the Korean peninsula. For 200 *mm* of rainfall as an example, the expected waiting times until reoccurrence in the southern part (spatial median of the peninsula) are 17.6 (20.8) years during the reference period, 13.5 (15.9) years during P1, 7.7 (9.1) years during P2, and 5.3 (6.2) years during P3 based on the SSP5-8.5 scenario.

The relative improvement of the PI-weighted method over the simple average is turned out to be approximately 46.5% in the error index for the reference period. The improvements for the future are approximately 12.2% and 14.6% in the Brier skill score and in the variance of rainfall intensity (20-year return level) prediction, respectively.

Heavy rainfall can have a significant effect on human life, communities, infrastructure, agriculture and livestock, and natural ecosystems. Thus, in addressing the impact of climate change owing to more frequent extreme precipitation events, governments and communities should prepare the proper infrastructure and systems more carefully and securely to prevent critical damage such as a loss of life from landslides and flooding.

## Acknowledgment

We acknowledge the World Climate Research Programme Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups for producing and making available their model output. For CMIP, the U.S. Department of Energy Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. The CMIP data is available through ESGF's website at <https://esgf-node.llnl.gov/projects/> and DKRZ's website at <https://cera-www.dkrz.de/WDCC/ui/cersearch/>. Observational station data in the Korean peninsula are provided by KMA at <http://data.kma.go.kr/>. The APHRODITE data are available at <http://aphrodite.st.hirosaki-u.ac.jp/download/>. We thank all contributors to the numerical R packages which were crucial for this work. The authors are grateful to Prof. Seung-Ki Min (Division of Environmental Sciences and Engineering, POSTECH) who provided valuable comments to improve this paper. This study was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (No.2020R1I1A3069260), and funded by the Korea Meteorological Administration Research and Development Program under Grant KMI2018-03414.

## References

- [1] IPCC (2012) Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. Special Report of the Intergovernmental Panel on Climate Change, Cambridge University Press. <http://ipcc-wg2.gov/SREX/report/>

- [2] Easterling, D.R., K.E. Kunkel, J.R. Arnold, T. Knutson, A.N. LeGrande, L.R. Leung, R.S. Vose, D.E. Waliser, M.F. Wehner (2017) Precipitation change in the United States. In: Climate Science Special Report: Fourth National Climate Assessment, Volume I [Wuebbles, D.J., D.W. Fahey, K.A. Hibbard, D.J. Dokken, B.C. Stewart, and T.K. Maycock (eds.)]. U.S. Global Change Research Program, Washington, DC, USA, pp.207–230, doi:10.7930/J0H993CC.
- [3] Westra S, Alexander LV, Zwiers FW (2013) Global increasing trends in annual maximum daily precipitation. *J Climate* 26:3904–3918.
- [4] Freychet N, Hsu H, Chou C, Wu C (2015) Asian summer monsoon in CMIP5 projections: a link between the change in extreme precipitation and monsoon dynamics. *J Clim* doi:10.1175/JCLI-D-14-00449.1
- [5] Alexander LV (2016) Global observed long-term changes in temperature and precipitation extremes: a review of progress and limitations in IPCC assessments and beyond. *Weather and Climate Extremes* 11:4–16.
- [6] Park C, Min SK, Lee D, Cha DH, Suh MS et al (2016) Evaluation of multiple regional climate models for summer climate extremes over East Asia. *Clim Dynam* 46:2469—2486.
- [7] Lenderink G, van Meijgaard E (2008) Increase in hourly precipitation extremes beyond expectations from temperature changes. *Nature Geosci* 1:511–514.
- [8] Berg P, Moseley C, Haerter JO (2013) Strong increase in convective precipitation in response to higher temperatures. *Nature Geosci* 6:181–185.
- [9] Scott M. (2019) Prepare for more downpours: Heavy rain has increased across most of the United States, and is likely to increase further. *ClimateWatch Magazine*, NOAA Climate.gov
- [10] Hov Ø, Ulrich Cubasch U, Fischer E, Höppe P, Iversen T et al.(2013) *Extreme Weather Events in Europe: Preparing for Climate Change Adaptation*. Report produced by Norwegian Meteorological Institute in cooperation with EASAC, Oslo, Norway.
- [11] Ho C-H, Park T-W, Jun S-Y et al (2011) A projection of extreme climate events in the 21st century over East Asia using the community climate system model 3. *Asia Pac J Atmos Sci* 47:329–344
- [12] Kwon SH, Kim J, Boo KO, Shim S et al (2019) Performance-based projection of the climate change effects on precipitation extremes in East Asia using two metrics. *Intern J Climatol* 39(4):2324–2335.
- [13] Mukherjee S, Aadhar S, Stone D, Mishra V (2018) Increase in extreme precipitation events under anthropogenic warming in India. *Weather and Climate Extremes* 20:45–53.
- [14] Jung HS, Choi YE, Lim GH (2002) Recent trends in temperature and precipitation over South Korea. *Int J Climatol* 22:1327–1337.
- [15] Choi KS, Moon JY, Kim DW (2010) The significant increase of summer rainfall occurring in Korea from 1998. *Theor Appl Climatol* 102:275–286. doi:10.1007/s00704-010-0256-0
- [16] Park J-S, Kang H-S, Lee Y, Kim M-K (2011) Changes in the extreme daily rainfall in South Korea. *Intern J Climatol* 31:2290–2299.
- [17] Lee Y, Paek J, Park JS, Boo K-O (2020) Changes in temperature and rainfall extremes across East Asia in the CMIP5 ensemble. *Theoretical and Applied Climatology* 141:143–155. doi:10.1007/s00704-020-03180-w
- [18] Boo KO, Kwon WT, Baek HJ (2006) Change of extreme events of temperature and precipitation over Korea using regional projection of future climate change. *Geophy Res Lett* 33:L01701 doi:10.1029/2005GL023378

- [19] Im ES, Jung IW, Bae DH (2011) The temporal and spatial structures of recent and future trends in extreme indices over Korea from a regional climate projection. *Int J Climatol* 31(1):72–86.
- [20] Seo YA, Lee Y, Park J-S, Kim M-K, Cho C, Baek H-J (2015) Assessing changes in observed and future projected precipitation extremes in South Korea. *Intern J Climatol* 35:1069–1078.
- [21] Ahn J-B, Jo S, Suh M-S, Cha D-H, Lee D-K et al.(2016) Changes of precipitation extremes over South Korea projected by the 5 RCMs under RCP scenarios. *Asia Pac J Atmos Sci* 52:223–236.
- [22] Cha DH, Lee DK, Jin CS, Kim G et al. (2016) Future changes in summer precipitation in regional climate simulations over the Korean Peninsula forced by multi-RCP scenarios of HadGEM2-AO. *Asia Pac J Atmos Sci* 52:139–149.
- [23] Kim G, Cha D-H, Park C, Lee G, Jin C-S et al (2019). Future changes in extreme precipitation indices over Korea. *Int J Climatol* 38(6):862–874 DOI:10.1002/joc.5414
- [24] Lee Y, Shin YG, Park JS, Boo KO (2020) Future projections and uncertainty assessment of precipitation extremes in the Korean peninsula from the CMIP5 ensemble. *Atmos Sci Lett* e954. <https://doi.org/10.1002/asl.954>
- [25] Tebaldi C, Hayhoe K, Arblaster JM, Meehl GA (2006) Going to the extremes: An intercomparison of model-simulated historical and future changes in extreme events. *Clim Change* 79:185–211.
- [26] Knutti, R.(2010) The end of model democracy? *Clim Change* 102:394-404.
- [27] Suh MS, Oh SG, Lee DK, Cha DH et al (2012) Development of new ensemble methods based on the performance skills of regional climate models over South Korea. *J Clim* 25:7067–7082.
- [28] Sanderson BM, Knutti R, Caldwell P (2015b) A representative democracy to reduce interdependency in a multimodel ensemble. *J Clim* 28:5171-94.
- [29] Massoud EC, Espinoza V, Guan B, Waliser DE (2019) Global Climate Model Ensemble Approaches for Future Projections of Atmospheric Rivers. *Earth’s Future* 7(10):1136-1151 <http://dx.doi.org/10.1029/2019EF001249>
- [30] Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., et al. (2019). Taking climate model evaluation to the next level. *Nature Climate Change*, 9(2), 102–110
- [31] Xu D, Ivanov V, Kim J, Fatichi S (2019) On the use of observations in assessment of multi-model climate ensemble. *Stoch Env Res Risk Assess* 33:1923–1937.
- [32] Brunner L, Lorenz R, Zumwald M, Knutti R (2019) Quantifying uncertainty in European climate projections using combined performance-independence weighting. *Environ Res Lett* 14:124010
- [33] Georgi F, Mearns LO (2002) Calculation of average, uncertainty range and reliability of regional climate changes from AOGCM simulations via the ‘Reliability Ensemble Averaging (REA)’ method. *J Clim* 15:1141-1158.
- [34] Abramowitz G, Gupta H (2008) Toward a model space and model independence metric. *Geophys Res Lett* 35:L05705
- [35] Knutti R, Sedlacek J, Sanderson BM, Lorenz R, Fischer EM, Eyring V (2017) A climate model projection weighting scheme accounting for performance and independence. *Geophys Res Lett* 44:1909-18.
- [36] Lorenz R, Herger N, Sedlacek J, Eyring V, Fischer EM, Knutti R. (2018) Prospects and caveats of weighting climate models for summer maximum temperature projections over North America. *J Geophys Res Atmos* 123:4509-26.
- [37] Shin Y, Lee Y, Park JS (2020) A Weighting Scheme in A Multi-Model Ensemble for Bias-Corrected Climate Simulation. *Atmosphere* 11:775 doi:10.3390/atmos11080775

- [38] O'Neill BC, Kriegler E, Riahi K, Ebi KL, Hallegatte S, Carter TR, Mathur R, van Vuuren DP (2014) A new scenario framework for climate change research: the concept of Shared Socioeconomic Pathways. *Clim. Change* 122:387-400
- [39] Koch SE, DesJardins M, Kocin PJ (1983). An interactive Barnes objective map analysis scheme for use with satellite and conventional data, *J. Climate Appl. Meteo.* 22:1487-1503.
- [40] Yatagai A, Kamiguchi K, Arakawa O, Hamada A, Yasutomi N, Kitoh A (2012) APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Bull Amer Meteorol Soc* 93:1401-1415.
- [41] Maraun D, Widmann M (2018) *Statistical downscaling and bias correction for climate research*. Cambridge Univ Press.
- [42] Coles S (2001) *An Introduction to Statistical Modelling of Extreme Values*. Springer, New York, pp 224.
- [43] Christensen JH, Boberg F, Christensen OB, Lucas-Picher P (2008) On the need for bias correction of regional climate change projections of temperature and precipitation. *Geophys Res Lett.* 35:L20709, doi:10.1029/2008GL035694.
- [44] Vrac M, Friederichs P (2015) Multivariate-intervariable, spatial, and temporal-bias correction. *J Clim* 28:218-237.
- [45] Cannon AJ. Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables. *Clim Dyn* 2018;50:31-49.
- [46] Sanderson BM, Knutti R, Caldwell P (2015a) Addressing interdependency in a multimodel ensemble by interpolation of model properties. *J Clim* 28:5150-70.
- [47] Ross S (2010) *A first course in probability*. 8th ed., Pearson Prentice Hall
- [48] Wilks D (2011) *Statistical Methods in the Atmospheric Sciences*, 3rd Ed., Academic Press, New York.
- [49] Kharin VV, Zwiers FW, Zhang X, Wehner M (2013) Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Clim Change* 119:345-357.
- [50] Serinaldi F (2015) Dismissing return periods!. *Stoch Environ Res Risk Assess* 29, 1179-1189 <https://doi.org/10.1007/s00477-014-0916-1>
- [51] Hawkins E, Sutton RT (2009) The potential to narrow uncertainty in regional climate predictions. *Bull Amer Meteorol Soc* 90(8):1095-1107.
- [52] Yip S, Ferro CAT, Stephenson DB (2011) A simple, coherent framework for partitioning uncertainty in climate predictions. *Jour of Climate* 24:4634-4643.
- [53] Paciorek CJ, Stone DA, Wehner MF (2018) Quantifying statistical uncertainty in the attribution of human influence on severe weather. *Weather and Climate Extremes* 20:69-80. <https://doi.org/10.1016/j.wace.2018.01.002>
- [54] Baker NC, Taylor PC (2016) A framework for evaluating climate model performance metrics. *Jour of Climate* 29:1773-1782.
- [55] Draper D (1995) Assessment and propagation of model uncertainty. *Jour Royal Statist Soc: Series B* 57(1):45-97
- [56] Zhu J, Forsee W, Schumer R, Gautam M (2013) Future projections and uncertainty assessment of extreme rainfall intensity in the United States from an ensemble of climate models. *Climatic Change* 118(2):469-485. doi:10.1007/s10584-012-0639-6



- [57] Ruckstuhl C, Philipona R, Morland J, Ohmura A (2007) Observed relationship between surface specific humidity, integrated water vapor, and longwave downward radiation at different altitudes. *J Geophys Res Atmos* 112:1–7.
- [58] Kendon EJ, Rowell DP, Jones RG, Buonomo E (2008) Robustness of future changes in local precipitation extremes. *J Clim* 21:4280–4297.
- [59] Sillmann J, Kharin VV, Zhang X, Zwiers FW, Branaugh D (2013) Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *J Geophys Res Atmos* 118:1–18.
- [60] Brunner L, Pendergrass AG, Lehner F, Merrifield AL, Lorenz R, Knutti R (2020) Reduced global warming from CMIP6 projections when weighting models by performance and independence. *Earth System Dynamics Discussion*. Preprint Doi:10.5194/esd-2020-23

Supplementary Material:  
Future projections and uncertainty assessment of precipitation  
extremes in the Korean peninsula from the CMIP6 ensemble

Yonggwon Shin<sup>1</sup>, Yire Shin<sup>1</sup>, Juyoung Hong<sup>1</sup>, Maeong-Ki Kim<sup>2</sup>,  
Young-Hwa Byun<sup>3</sup>, Kyung-On Boo<sup>3</sup>, Il-Ung Chung<sup>4</sup>, Doo-Sun R. Park<sup>5</sup>,  
and Jeong-Soo Park<sup>1,\*</sup>

*1: Department of Statistics, Chonnam National University, Gwangju 500-757, Korea.*

*2: Department of Atmospheric Science, Kongju National University, Gongju, Korea.*

*3: Innovative Meteorological Research Department,  
National Institute of Meteorological Sciences, Seogwipo, Korea.*

*4: Dept of Atmospheric and Environmental Sciences,  
Gangneung-Wonju National University, Korea.*

*5: Dept. of Earth Sciences Education, Kyungpook National University, Daegu, Korea.*

*\*: Corresponding author, E-mail: jspark@jnu.ac.kr, Tel: +82-62-530-3445*

August 26, 2020

## 1 Data and simulation models

Figure S1 shows examples of time series plots of the observations, APHRODITE data [1], and the bias-corrected data, for the AMP1. The APHRODITE values of the AMP1 are smaller than the observations when comparing those data at near stations. Because of this difference, we applied a bias correction technique to the APHRODITE data, based on the observations of nearest neighbor stations.

Table S 1: The list of 21 CMIP6 (Coupled Model Intercomparison Project Phase 6) models analyzed in this study. The detailed information on each model are available at ESGF-node <https://esgf-node.llnl.gov/projects/cmip6/>.

Model Name	Institution	Country	Resolution (Lon $\times$ Lat Level)
MIROC6	JAMSTEC, AORI, NIES, R-CCS (MIROC)	Japan	256 $\times$ 128 L81(T85)
BCC-CSM2-MR	Beijing Clim Center	China	320 $\times$ 160 L46(T106)
CanESM5	Canadian Centre Clim Model & Analysis, Enviro & Clim Change (CCCma)	Canada	128 $\times$ 64 L49(T63)
MRI-ESM2.0	Meteoro Research Institute (MRI)	Japan	320 $\times$ 160 L80(TL159)
CESM2-WACCM	Nat Center for Atmos Res, Clim & Global Dynamics Lab (NCAR)	USA	288 $\times$ 192 L70
CESM2	Nat Center for Atmos Res, Clim & Global Dynamics Lab (NCAR)	USA	288 $\times$ 192 L32
KACE1.0-GLOMAP	National Inst of Meteo Sci/Meteo Admin, Clim Res Div (NIMS-KMA)	Korea	192 $\times$ 144 L85
UKESM1-0-N96ORCA1	MOHC & NERC, NIMS-KMA, NIWA	UK, Korea New Zealand	192 $\times$ 144 L85
MPI-ESM1.2-LR	Max Planck Inst for Meteo (MPI-M)	Germany	192 $\times$ 96 L47(T63)
MPI-ESM1.2-HR	Max Planck Inst for Meteo (MPI-M)	Germany	384 $\times$ 192 L95(T127)
INM-CM5-0	Inst for Numerical Math, Russian Acad of Sci (INM)	Russia	180 $\times$ 120 L73
INM-CM4-8	Inst for Numerical Math, Russian Acad of Sci (INM)	Russia	180 $\times$ 120 L21
IPSL-CM6A-LR	Institut Pierre Simon Laplace (IPSL)	France	144 $\times$ 143 L79
NorESM2-LM	NorESM Consortium of CICERO, MET-Norway, NERSC, NILU, UiB, UiO, UNI	Norway	144 $\times$ 96 L32
NorESM2-MM	NorESM Consortium of CICERO, MET-Norway, NERSC, NILU, UiB, UiO, UNI	Norway	288 $\times$ 192 L32
EC-Earth3-Veg	EC-Earth consortium, Swedish Meteo & Hydro Inst/SMHI, Sweden	EU	512 $\times$ 256 L91(TL255)
EC Earth 3.3	EC-Earth consortium, Swedish Meteo & Hydro Inst/SMHI, Sweden	EU	512 $\times$ 256 L91(TL255)
ACCESS-CM2	CSIRO, ARCCSS (Australian Res Council Centre of Excellence for Clim System Sci)	Australia	192 $\times$ 144 L85
ACCESS-ESM1-5	Commonwealth Scientific & Industrial Res Organ (CSIRO)	Australia	192 $\times$ 145 L38
GFDL-ESM4	National Oceanic & Atmos Admi, Geophy Fluid Dynamics Lab	USA	360 $\times$ 180 L49
FGOALS-g3	Chinese Academy of Sciences (CAS)	China	180 $\times$ 80 L26

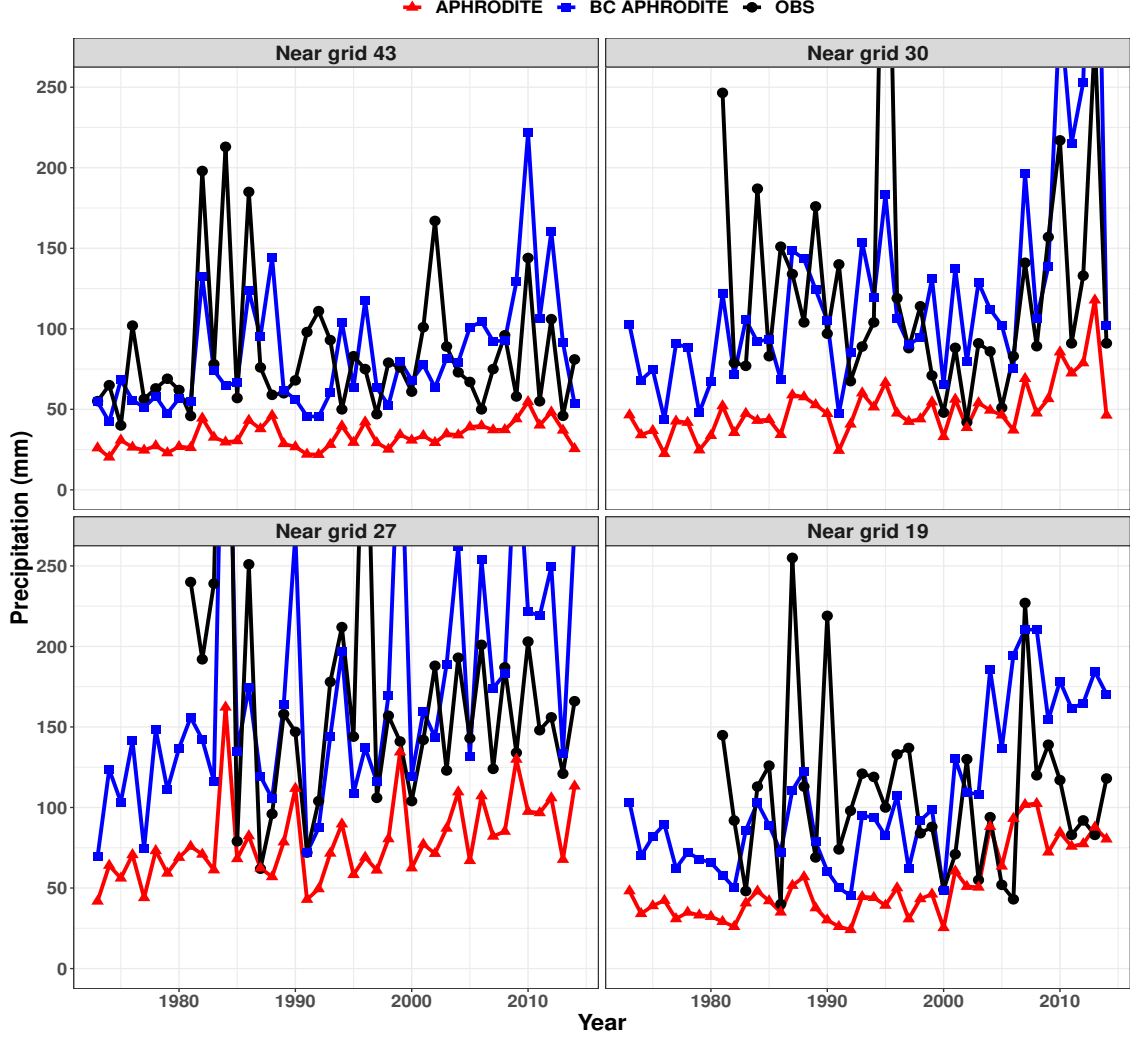


Figure S 1: Examples of time series plots of the observations, APHRODITE data, and the bias-corrected data.

## 2 Methods

### 2.1 Generalized extreme value distribution

Assuming the data approximately follow a GEV distribution, the parameters can be estimated by the maximum likelihood method[2, 3] or the method of L-moments estimation. The L-moments estimator is more efficient than the maximum likelihood estimator in small samples for typical shape parameter values [4]. The L-moments method is employed in this study

13 using the "lmom" package in R [5] because a relatively small number of samples are analyzed  
 14 for each comparison period. Moreover, the formulae used to obtain the L-moments estimator  
 15 are simple compared to that of obtaining the maximum likelihood estimator which needs an  
 16 iterative optimization until convergence.

## 17 **2.2 Bias correction: Multivariate generalization of quantile mapping**

18 Some BC methods such as quantile mapping or delta change [6] make a perfect matching in  
 19 the sense that the quantiles of the observations and the historical data are same. When the  
 20 BC such as quantile mapping is used, most the model weights based on performance become  
 21 equal because of a perfect matching, and consequently, the prediction is the simple average  
 22 of bias-corrected model outputs. This is approximately true for the MBC [8] employed in  
 23 this study because the MBC is a multivariate generalization of quantile mapping. Thus the  
 24 historical data is not bias corrected. No-bias-corrected historical data are utilized to calculate  
 25 the performance weight of a model.

26 Chen et al.[7] found that the joint BC of precipitation and air temperature led to a much  
 27 better performance than univariate BC, in terms of hydrological modelling for all their studying  
 28 basins located in various climates except for the coldest Canadian basin. Cannon [8] proposes a  
 29 multivariate generalization of quantile mapping (QM). It is an iterative method which concep-  
 30 tually lays between univariate bias correction (BC) methods and the empirical copula-based  
 31 correction (EC-BC) [9]. For a univariate BC, the quantile delta mapping (QDM) [10] is used,  
 32 which preserves trends of model data

33 It approximately preserves the multivariate dependence of the driving climate model. Here,  
 34 an image processing technique—the N-dimensional probability density function transform (N-  
 35 pdft)—designed to transfer color information from one image to another is adapted. In each  
 36 iteration, univariate QM is first applied separately to each variable. Then a linear multivariate  
 37 BC is applied by re-scaling the multivariate anomalies based on Cholesky decomposition of the  
 38 covariance matrix. The algorithm ends when both the corrected marginals and the dependence  
 39 structure are sufficiently close to their observed counter parts. A variant is based on ranks  
 40 rather than on the actual values [6]. It provides a multivariate quantile delta mapping, referred  
 41 to as MBCn (multivariate bias correction with N-pdft) algorithm. It consists, in each iteration,  
 42 of a random orthogonal rotation of multivariate input data, a univariate quantile delta mapping  
 43 on the rotated fields and the inverse rotation. This algorithm approximately preserves trends  
 44 of model data. We used 'MBC' package [11] in R for computation. The details are found in  
 45 Cannon [8].

### 3 PI-weights

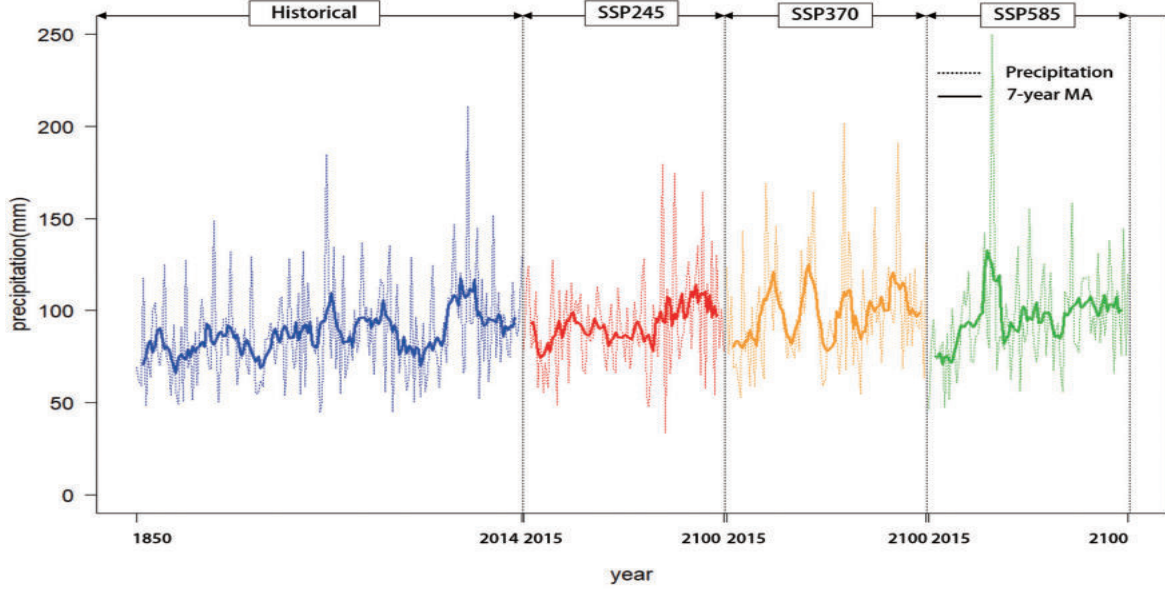


Figure S 2: Arrangement of data and 7-year moving averages composed of the historical data from 1850 to 2014 and the future data from 2015 to 2100 under SSP2, SSP3, and SSP5 scenarios for computing the Spearman correlation coefficient between models.

#### 3.1 Computing independence weights

If a model has no close neighbors, then all  $S_{ij} (i \neq j)$  are large, the denominator of the PI-weight is approximately one and has no effect. If two models  $i$  and  $j$  are identical, then  $S_{ij} = 0$ , the denominator equals two, so each model gets half the weight.

To calculate the model similarity  $S_{ij}$ , we follow a technique among several methods proposed by Sanderson et al.[12]. A method employed in this study is based on empirical orthogonal function (EOF) or principal component analysis. The following process is done for each grid: First, for each model, the historical data from 1850 to 2015 and the future simulation data from three scenarios are lined as one row as in Figure S2. The bias correction is not applied to all data for this process. One can choose the historical data only as did by Brunner et al.[13], but we deploy all simulation data for a maximum use of available information. For each time series induced from each model, 7-year moving averages are obtained. Then, a correlation matrix  $R$  among all  $M$  models is constructed by applying the Spearman correlation coefficients

to those  $M$  number of series of 7-year moving averages. That is,  $R$  is the correlation matrix of  $M$  models, with size  $M \times M$ .

A singular value decomposition (SVD) is performed on  $R^{1/2}$  and truncated to  $t$  modes to obtain the dominant modes of multivariate ensemble variability such that

$$R^{1/2} = U\lambda V^T, \quad (1)$$

where  $U$  is an orthogonal matrix of model loadings (size  $M$  by  $t$ ) whose columns are the eigenvectors of the model correlation matrix  $R$ ,  $\lambda$  (size  $t$  by  $t$ ) are the eigenvalues of  $R$ , and  $V$  (size  $M$  by  $t$ ) are the eigenvectors of  $R$ . The dimensions are sorted by decreasing eigenvalue, such that the basis set can be truncated to a smaller number of modes  $t$  [12]. Note  $t$  is often determined by selecting number of the eigenvalues greater than 1.

The model loadings  $U$  now define a  $t$ -dimensional space (where  $t$  is the truncation length of the SVD) in which intermodel and observation-model Euclidean distances may be defined. The intermodel distances can then be measured in a Euclidean sense in the loadings matrix, such that the distances  $S_{ij}$  between two models  $i$  and  $j$  can be expressed as [12]

$$S_{ij} = \left\{ \sum_{l=1}^t [U(i, l) - U(j, l)]^2 \right\}^{1/2}. \quad (2)$$

$U(i, l)$  is interpreted as a correlation or a dependency of the model  $i$  to the  $l$ -th principal component. Thus small  $S_{ij}$  value means high dependency or similarity between models  $i$  and  $j$ .

An example of the distances  $S_{ij}$  between two models  $i$  and  $j$  calculated from Eq.(2) for some models is given in Table S2. Small value indicates high dependency or similarity between two models. In Table S2, the first 4 models show highest similarity whereas the last 4 models show lowest similarity.

### 3.2 Computing performance weights

To compute the performance of each model,  $T$ -year return levels are compared based on the GEVD fitting on the historical data and the observations. Let us denote  $r_T^i$  and  $r_T^0$  as  $T$ -year return level obtained from the historical data of  $i$ -th model and the observations, respectively. These values are normalized as follows to make it scale-free, for  $i = 0, 1, \dots, M$ :

$$\tilde{r}_T^i = \frac{r_T^i - med_i}{R_i}, \quad (3)$$

Table S 2: The similarity distance metric  $S_{ij}$  between model  $i$  and model  $j$  calculated from Eq.(2) for some models. Small value indicates high dependency or similarity between two models.

	CanESM5	ACCESS-CM2	UKESM	KACE	NorESM2-LM	EC-Earth3	FGOALS	INM-CM4
CanESM5	0	0.40	0.40	0.37	0.46	0.52	0.69	0.73
ACCESS-CM2	0.40	0	0.40	0.47	0.47	0.58	0.77	0.76
KACE-1-0-G	0.37	0.47	0.49	0	0.61	0.61	0.72	0.80
UKESM	0.40	0.40	0	0.49	0.50	0.59	0.73	0.75
IPSL-CM6A-LR	0.43	0.50	0.52	0.51	0.59	0.62	0.71	0.84
ACCESS-ESM1-5	0.48	0.48	0.51	0.54	0.59	0.63	0.76	0.87
NorESM2-LM	0.46	0.47	0.50	0.61	0	0.56	0.82	0.81
CESM2-WACCM	0.46	0.55	0.55	0.44	0.68	0.63	0.81	0.82
EC-Earth3-Veg	0.48	0.53	0.54	0.51	0.59	0.65	0.78	0.76
CESM2	0.54	0.55	0.53	0.54	0.63	0.64	0.83	0.81
EC-Earth3	0.52	0.58	0.59	0.61	0.56	0	0.85	0.81
NorESM2-MM	0.53	0.58	0.59	0.59	0.74	0.75	0.84	0.76
MPI-ESM1-2-LR	0.50	0.64	0.62	0.51	0.68	0.65	0.79	0.84
BCC-CSM2-MR	0.59	0.56	0.58	0.58	0.69	0.73	0.82	0.91
MPI-ESM1-2-HR	0.64	0.63	0.65	0.68	0.63	0.69	0.82	0.87
GFDL-ESM4	0.69	0.65	0.73	0.73	0.72	0.69	0.90	0.92
MIROC6	0.70	0.70	0.68	0.68	0.78	0.75	0.89	0.96
INM-CM5-0	0.62	0.73	0.69	0.70	0.74	0.74	0.94	0.87
MRI-ESM2-0	0.66	0.75	0.73	0.74	0.79	0.82	0.89	0.88
FGOALS-g3	0.69	0.77	0.73	0.72	0.82	0.85	0	0.99
INM-CM4-8	0.73	0.76	0.75	0.80	0.81	0.81	0.99	0
SUM	10.89	11.70	11.78	11.82	13.08	13.51	16.35	16.76

where

$$R_i = \begin{cases} \max_i - \text{med}_i & \text{if } r_T^i \geq \text{med}_i, \\ \text{med}_i - \min_i & \text{if } r_T^i < \text{med}_i, \end{cases} \quad (4)$$

and  $\max_i$ ,  $\min_i$ , and  $\text{med}_i$  are the maximum, the minimum, and the median of  $i$ -th model data. Other ways of standardizations are also possible.

The distance for performance measure is obtained by

$$D_i^2 = \sum_T (\tilde{r}_T^i - \tilde{r}_T^0)^2. \quad (5)$$

We set  $T = 2, 5, 10, 20, 30, 50$ , and  $100$ . Note that  $D_i$  does not depend on the shape parameter  $\sigma_D$ , and so obtained  $D_i$ s are fixed for the next computation.



### 3.3 Computing the p-value in selecting $\sigma_D$

The p-values in the manuscript are computed by a Monte-Carlo simulation in which random numbers of weights are generated from the Dirichlet distribution [14]. When the parameters are all equal to 1, the Dirichlet distribution is same as the multivariate uniform distribution with values between 0 and 1, which represents the null hypothesis. We used ‘MCMCpack’ package [15] in R to generate the random weights satisfying  $H_0$ .

The detailed steps of computing the p-value for given  $\sigma_D$  and  $\chi_0^2(\sigma_D)$  are:

**Step 1:** Generate random weights  $P_i^{(k)}$  from the Dirichlet distribution with all parameters equal to 1 (under  $H_0$ ), for  $i = 1, \dots, M$

**Step 2:** Compute  $\chi^2 = \sum_{i=1}^M \frac{(\frac{1}{M} - P_i^{(k)})^2}{1/M}$ , and denote it  $\chi_{(k)}^2$

**Step 3:** Iterate the above two Steps K (=1,000 for example) times

**Step 4:** Calculate  $p - value(\sigma_D) = \sum_{k=1}^K \frac{I[\chi_{(k)}^2 > \chi_0^2(\sigma_D)]}{K}$ ,

where  $I[A]$  denotes the identity function which takes 1 or 0 according that the condition  $A$  is satisfied or not. Note that  $P_i^{(k)}$ s generated in Step 1 do not depend on  $\sigma_D$ .

## 4 Result: Future projection of extreme rainfall

The relative change of 20-year return level in the period P1 relative to the reference period P0 is defined as:

$$\delta R_{20}(P1) = \frac{R_{20}(P1) - R_{20}(P0)}{R_{20}(P0)} \times 100, \quad (6)$$

where  $R_{20}(P)$  is the 20-year return level in the period P.

### 4.1 Return period and exceedance probability

We have experienced some technical flows in computing the waiting time or the return periods corresponding to a return value. For example, the resulting return period sometimes turns out to be greater than 500 years even though it is expected to correspond to 50 years. It may be due to the cumulation of truncation or rounding errors in computer, related to inverting the quantile function of the GEVD. A trouble caused by this flow does not vanish even applied to unequally weighted regional frequency analysis (RFA). In this study, we thus adopted the trimmed mean [3] in RFA in which unfairly very high estimates of return periods are deleted

Table S 3: Statistics of 20-year and 50-year return levels of the annual maximum daily precipitation (unit: mm) averaged over 46 grids in the Korean peninsula for the observations (OBS) and the future periods; P1 (2021-2050), P2 (2046-2075), and P3 (2071-2100) under the SSP2, SSP3, and SSP5 scenarios.

			SSP2-4.5			SSP3-7.0			SSP5-8.5		
	Statistic	OBS	P1	P2	P3	P1	P2	P3	P1	P2	P3
20-year	Mean	216	230	240	249	226	244	263	232	256	285
	Q1	185	203	206	224	195	217	236	194	224	259
	Median	230	243	251	254	238	255	281	249	269	295
	Q3	250	260	273	285	259	276	300	271	301	322
50-year	Mean	267	279	293	305	273	297	322	280	313	351
	Q1	231	254	257	280	235	273	296	235	276	319
	Median	276	285	306	319	285	308	331	291	327	366
	Q3	303	309	332	345	306	334	363	319	357	392

in computing the weighted average. The defects of return periods are described in Serinaldi [16].

The spatially averaged estimates of exceedance probability are presented in Figure S6 and in Table S6.

## 5 Results: Quantifying uncertainty

From the analysis of variance, Figure S7 shows the interaction plots of 20-year return levels from 21 CMIP6 for 3 SSP scenarios over 3 periods. In this figure, the return levels of some models such as CESM2-WACOM, ACCESS-ESM1-5, NorESM2-MM, MiCRO6, MRI-ESM2-0, and GFDL-ESM4 decrease from the SSP2 to the SSP3, contrary to the expectation. Moreover, the return values of some models such as MPI-ESM1-2-HR, MPI-ESM1-2-LR, and NorESM2-LM decrease from the SSP3 to the SSP5. Figure S8 shows the interaction plots between 21 CMIP6 and the latitude in which the latitude changes from  $33^\circ$  to  $43^\circ$ , for 20-year return levels (unit: mm).

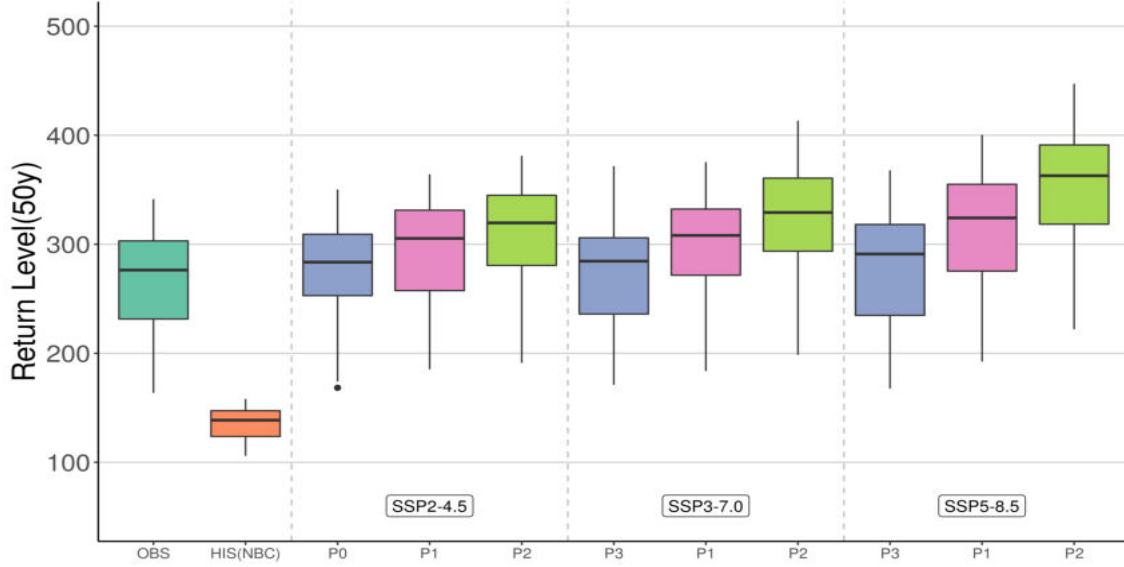


Figure S 3: Schematic box-plots of 50-year return levels of the annual maximum daily precipitation (unit: mm) averaged over 46 grids in the Korean peninsula for the future periods P1 (2021-2050), P2 (2046-2075), and p3 (2071-2100) under the SSP2, SSP3, and SSP5 scenarios. OBS and HIST(NBC) stand for the observations and the historical data without bias correction. The box-plot for p0 represents the bias-corrected historical data.

## References

- [1] Yatagai A, Kamiguchi K, Arakawa O, Hamada A, Yasutomi N, Kitoh A (2012) APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Bull Amer Meteorol Soc* 93:1401–1415.
- [2] Coles S (2001) *An Introduction to Statistical Modelling of Extreme Values*. Springer, New York, pp 224.
- [3] Wilks D (2011) *Statistical Methods in the Atmospheric Sciences*, 3rd Ed., Academic Press, New York.
- [4] Hosking JRM, Wallis JR (1997) *Regional frequency analysis: An approach based on L-moments*, Cambridge University Press, Cambridge, 244pp.
- [5] Hosking JRM (2019) L-Moments. R package, version 2.8. <https://CRAN.R-project.org/package=lmom>.
- [6] Maraun D, Widmann M (2018) *Statistical downscaling and bias correction for climate research*. Cambridge Univ Press.

Table S 4: Relative change (unit: %) in 20-year and 50-year return levels of the annual maximum daily precipitation averaged over the Korean peninsula relative to 1973–2014.

		SSP2-4.5			SSP3-7.0			SSP5-8.5		
		p1	p2	p3	p1	p2	p3	p1	p2	p3
20-year	Mean	7.1	11.7	16.4	4.9	13.3	22.5	7.2	19.2	33.4
	Q1	3.2	8.2	12.0	4.2	10.4	18.7	4.7	11.6	21.3
	Median	6.9	10.7	16.4	5.1	13.0	22.9	7.3	21.4	37.6
	Q3	10.7	15.4	21.6	6.3	16.6	26.9	10.4	24.4	42.0
50-year	Mean	6.3	11.3	16.4	3.4	12.5	22.4	5.9	19.1	34.0
	Q1	1.2	7.5	11.8	1.3	9.8	17.3	1.8	9.8	19.6
	Median	5.2	10.6	16.5	3.9	12.0	22.8	5.7	20.6	35.4
	Q3	10.3	15.5	22.6	5.6	15.8	26.1	10.3	25.0	45.8

Table S 5: Statistics of 20-year and 50-year return periods (unit: year) of the annual maximum daily precipitation averaged over 46 grids in the Korean peninsula for the future periods P1 (2021-2050), P2 (2046-2075), and p3 (2071-2100) under the SSP2, SSP3, and SSP5 scenarios.

		SSP2-4.5			SSP3-7.0			SSP5-8.5		
		P1	P2	P2	P1	P2	P3	P1	P2	P3
20-year	Mean	18.0	13.7	12.0	18.4	13.7	10.3	17.2	11.5	7.8
	Q1	16.4	12.5	10.9	14.7	11.6	9.5	14.2	10.4	6.8
	Median	17.6	13.4	11.9	18.6	13.8	10.1	17.0	12.1	7.6
	Q3	20.0	15.2	13.0	21.0	15.6	11.4	20.6	13.5	9.2
50-year	Mean	42.2	37.2	30.8	47.0	33.2	25.6	39.8	29.6	19.1
	Q1	35.9	33.7	26.6	41.8	28.4	22.7	36.7	26.4	16.0
	Median	42.0	35.5	30.2	48.9	33.6	24.9	41.8	29.9	18.6
	Q3	47.6	41.2	33.9	52.6	37.0	28.6	44.7	34.9	22.1

- 150 [7] Chen J, Li C, Brissette FP, Chen H, Wang M, Essou GRC (2018) Impacts of correcting  
151 the inter-variable correlation of climate model outputs on hydrological modeling. J Hydrol  
152 560:326–341.
- 153 [8] Cannon AJ (2018) Multivariate quantile mapping bias correction: an N-dimensional prob-  
154 ability density function transform for climate model simulations of multiple variables. Clim  
155 Dyn 50:31–49.

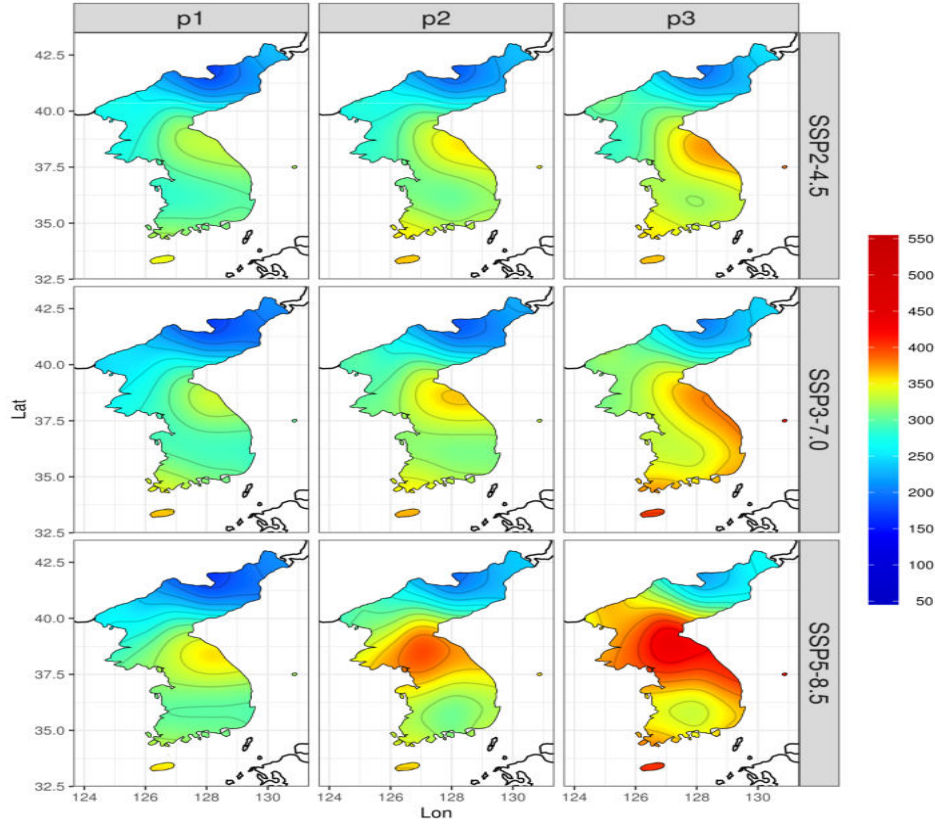


Figure S 4: Isopluvial maps of 50-year return levels of the annual maximum daily precipitation for 46 grids over the Korean peninsula for the future periods P1 (2021-2050), P2 (2046-2075), and P3 (2071-2100) under the SSP2, SSP3, and SSP5 scenarios.

- 156 [9] Vrac M, Friederichs P (2015) Multivariate-intervariable, spatial, and temporal-bias correc-  
157 tion. J Clim 28:218–237.
- 158 [10] Cannon AJ, Sobie SR, Murdock TQ (2015) Bias correction of GCM precipitation by  
159 quantile mapping: How well do methods preserve changes in quantiles and extremes? J  
160 Climate 28:6938–6959.
- 161 [11] Cannon AJ (2018) Package ‘MBC’ in R, version 0.10-4. [https://cran.r-](https://cran.r-project.org/web/packages/MBC/MBC.pdf)  
162 [project.org/web/packages/MBC/MBC.pdf](https://cran.r-project.org/web/packages/MBC/MBC.pdf)
- 163 [12] Sanderson BM, Knutti R, Caldwell P (2015a) Addressing interdependency in a multimodel  
164 ensemble by interpolation of model properties. J Clim 28:5150-70.
- 165 [13] Brunner L, Pendergrass AG, Lehner F, Merrifield AL, Lorenz R, Knutti R (2020) Re-  
166 duced global warming from CMIP6 projections when weighting models by performance and  
167 independence. Earth System Dynamics Discussion. Preprint Doi:10.5194/esd-2020-23

Table S 6: spatially averaged the exceedance probability for the annual maximum daily precipitation (AMP1) from 50mm to 300mm, obtained from the observations (OBS) and the CMIP6 models under the three scenarios for three future periods.

	AMP1	OBS	SSP2-4.5	SSP3-7.0	SSP5-8.5
Period 1	50 mm	0.989	0.989	0.990	0.991
	100 mm	0.531	0.704	0.687	0.705
	150 mm	0.149	0.240	0.222	0.238
	200 mm	0.048	0.069	0.069	0.063
	250 mm	0.019	0.024	0.021	0.022
	300 mm	0.008	0.009	0.008	0.010
Period 2	50 mm	0.989	0.988	0.990	0.992
	100 mm	0.531	0.712	0.748	0.789
	150 mm	0.149	0.238	0.300	0.312
	200 mm	0.048	0.080	0.097	0.110
	250 mm	0.019	0.027	0.034	0.038
	300 mm	0.008	0.013	0.018	0.019
Period 3	50 mm	0.989	0.990	0.992	0.993
	100 mm	0.531	0.755	0.793	0.822
	150 mm	0.149	0.262	0.343	0.384
	200 mm	0.048	0.086	0.127	0.161
	250 mm	0.019	0.032	0.048	0.062
	300 mm	0.008	0.015	0.019	0.033

- 168 [14] Everitt BS, Skrondal A (2010) *The Cambridge Dictionary of Statistics*, Cambridge Uni-  
169 versity Press.
- 170 [15] Martin AD, Quinn KM, Park JH (2011) MCMCpack: Markov Chain Monte Carlo in R.  
171 Journal of Statistical Software 42(9)
- 172 [16] Serinaldi F (2015) Dismissing return periods! Stoch Environ Res Risk Assess  
173 29:1179–1189. <https://doi.org/10.1007/s00477-014-0916-1>

Table S 7: The expected waiting time (unit: year) until reoccurrence or the return period of specific the annual maximum daily precipitation (AMP1) values from 50mm to 300mm, obtained from the observations (OBS) and the CMIP6 models under the 3 scenarios for 3 future periods.

		SSP2-4.5			SSP3-7.0			SSP5-8.5		
AMP1	OBS	P1	P2	P3	P1	P2	P3	P1	P2	P3
50 mm	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
100 mm	1.9	1.4	1.4	1.3	1.5	1.3	1.3	1.4	1.3	1.2
150 mm	6.7	4.2	4.2	3.8	4.5	3.3	2.9	4.2	3.2	2.6
200 mm	20.8	14.5	12.5	11.6	14.6	10.4	7.8	15.9	9.1	6.2
250 mm	52.6	41.2	37.3	31.6	47.8	29.5	20.8	46.1	26.4	16.1
300 mm	118.8	107.0	79.4	65.0	123.5	57.4	52.1	99.4	53.2	30.5

Table S 8: The expected frequency of reoccurring years during 30 years for specific the annual maximum daily precipitation (AMP1) values from 50mm to 300mm, obtained from the observations (OBS) and the CMIP6 models under the 3 scenarios for 3 future periods

		SSP2-4.5			SSP3-7.0			SSP5-8.5		
AMP1	OBS	P1	P2	P3	P1	P2	P3	P1	P2	P3
50 mm	29.7	29.7	29.6	29.7	29.7	29.7	29.8	29.7	29.8	29.8
100 mm	15.9	21.1	21.4	22.6	20.6	22.4	23.8	21.1	23.7	24.7
150 mm	4.5	7.2	7.1	7.9	6.7	9.0	10.3	7.1	9.4	11.5
200 mm	1.4	2.1	2.4	2.6	2.1	2.9	3.8	1.9	3.3	4.8
250 mm	0.6	0.7	0.8	1.0	0.6	1.0	1.4	0.7	1.1	1.9
300 mm	0.2	0.3	0.4	0.4	0.2	0.5	0.6	0.3	0.6	1.0

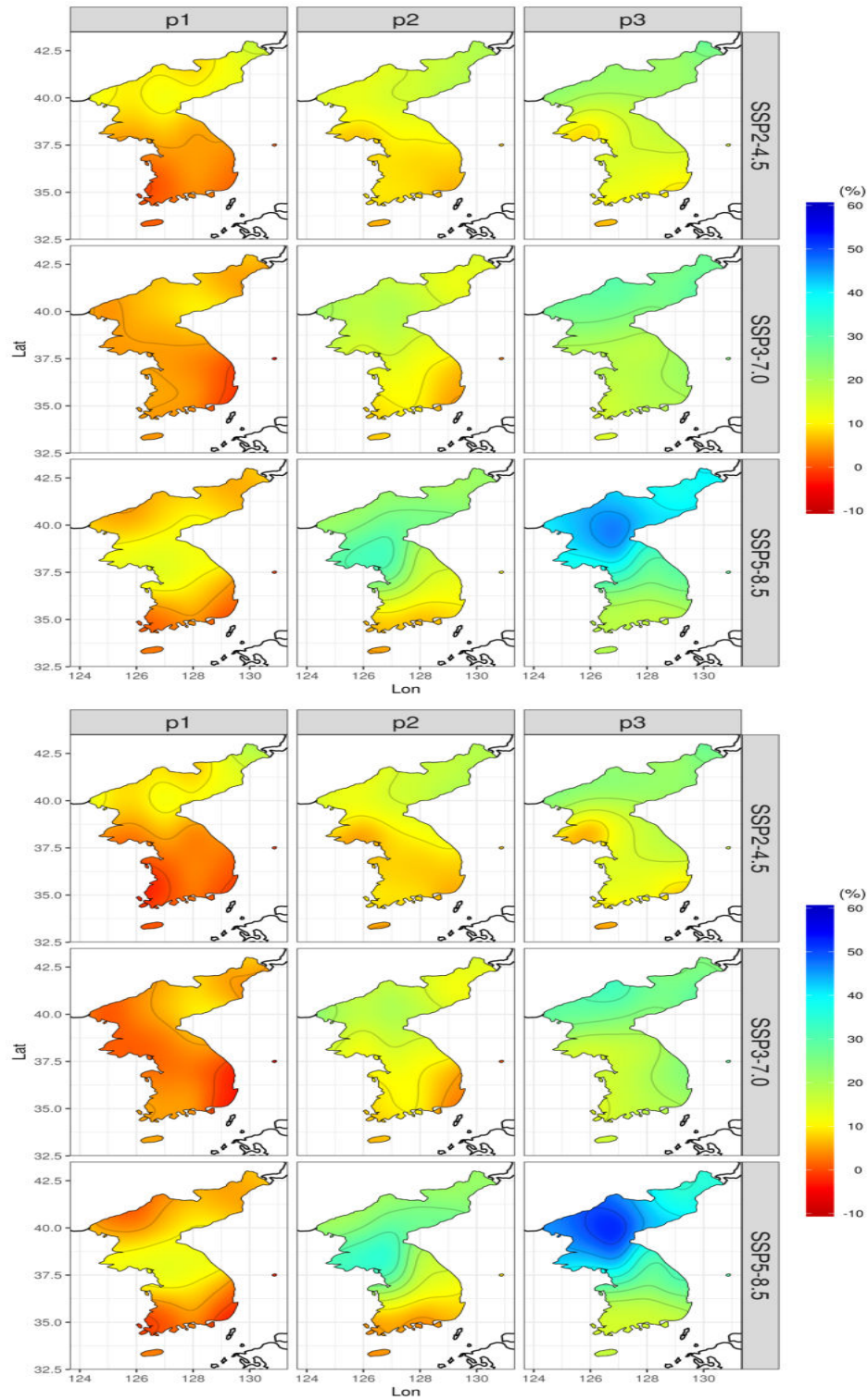


Figure S 5: Isopluvial maps of for the relative changes (unit: %) of 20-year and 50 return levels relative to 1973–2014 for the annual maximum daily precipitation for 46 grids over the Korean peninsula for the future periods P1 (2021-2050), P2 (2046-2075), and p3 (2071-2100) under the SSP2, SSP3, and SSP5 scenarios.



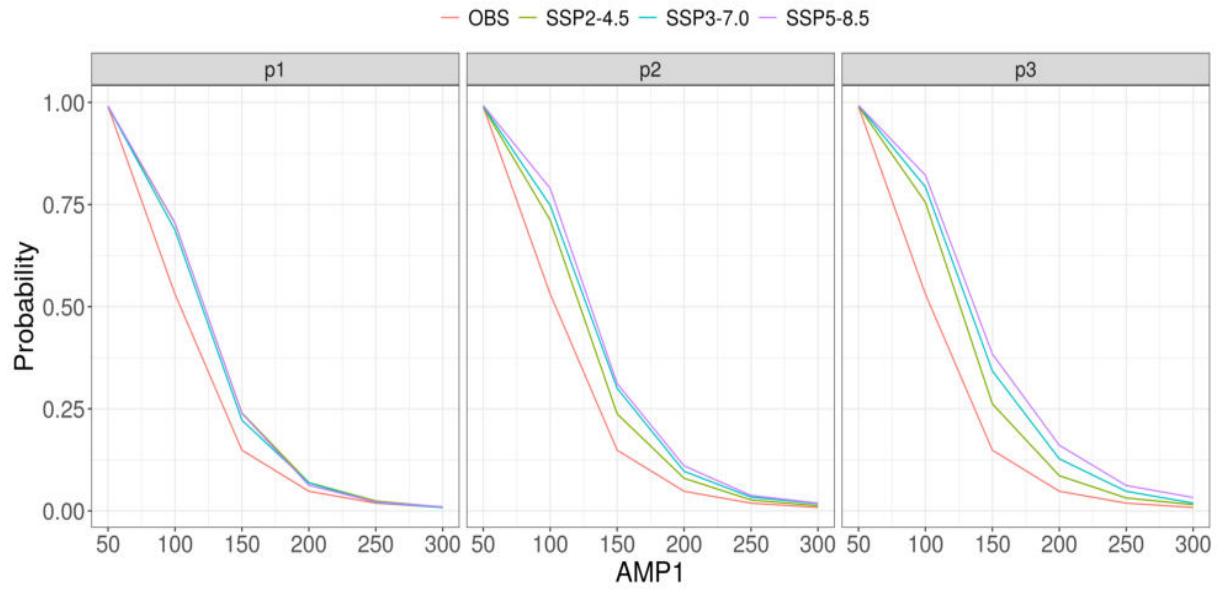


Figure S 6: The exceedance probability plots for the annual maximum daily precipitation (AMP1) from 50mm to 300mm, obtained from the observations (OBS) and the CMIP6 models under the three scenarios for three future periods.

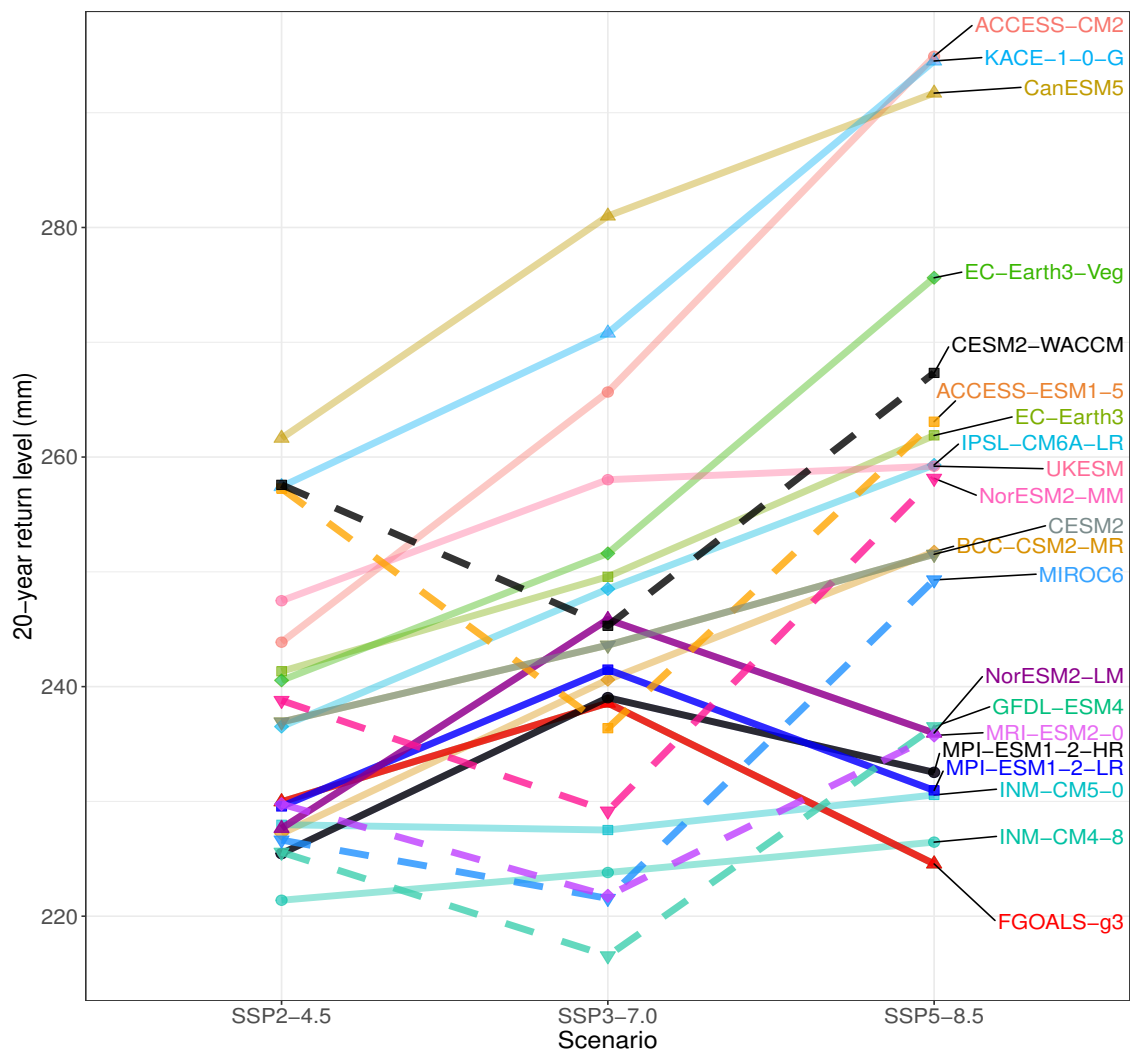


Figure S 7: Interaction plots of 20-year return levels from 21 CMIP6 for 3 SSP scenarios, averaged over 3 periods and 46 grids.

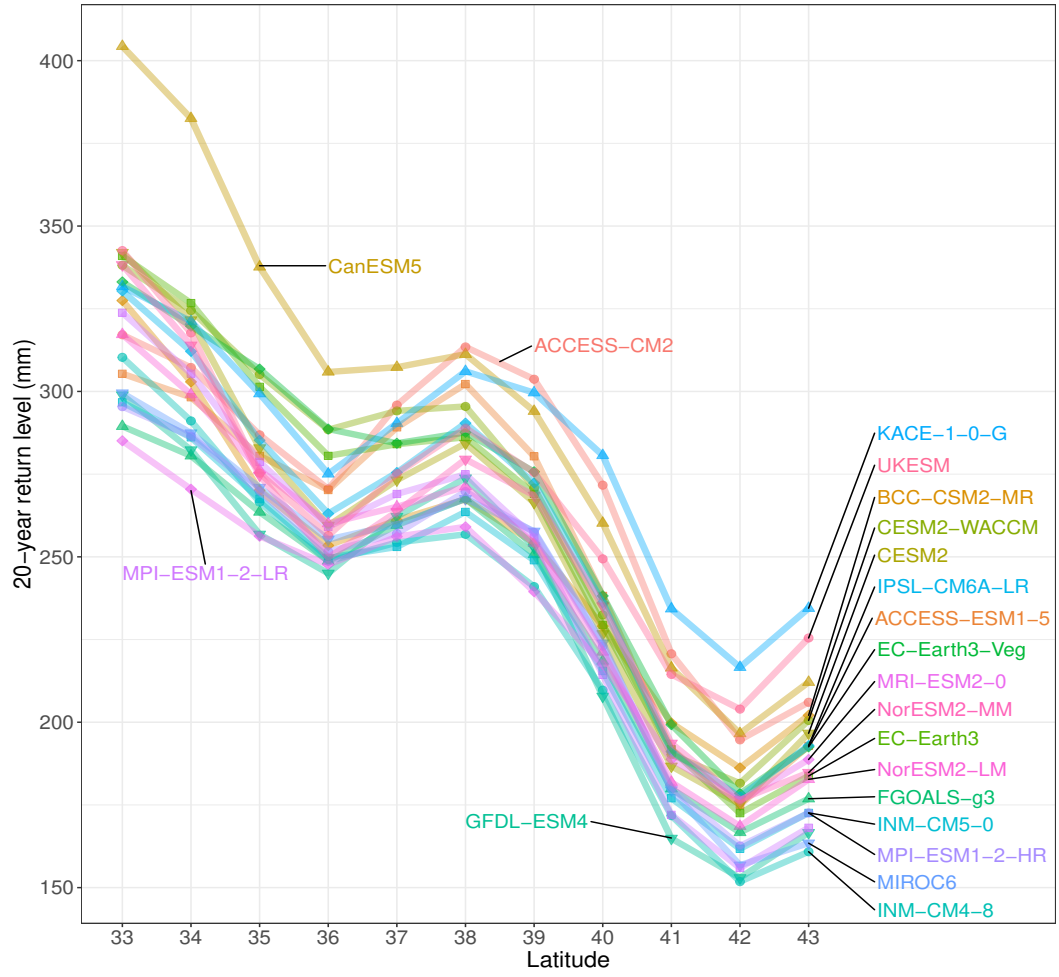


Figure S 8: Interaction plots between 21 CMIP6 and the latitude in which the latitude changes from  $33^{\circ}$  to  $43^{\circ}$ , for 20-year return levels (unit: mm).