



## Article

# Evaluation of Present-Day CMIP6 Model Simulations of Extreme Precipitation and Temperature over the Australian Continent

Nidhi Nishant <sup>1,2,\*</sup> , Giovanni Di Virgilio <sup>1,2,3</sup>, Fei Ji <sup>2,3</sup>, Eugene Tam <sup>3</sup>, Kathleen Beyer <sup>3,4</sup> and Matthew L. Riley <sup>3</sup> 

<sup>1</sup> Climate Change Research Centre, University of New South Wales, Sydney, NSW 2052, Australia

<sup>2</sup> Australian Research Council Centre of Excellence for Climate Extremes, University of New South Wales, Sydney, NSW 2052, Australia

<sup>3</sup> Science, Economics and Insights Division, NSW Department of Planning and Environment, Sydney, NSW 2141, Australia

<sup>4</sup> School of Geography, Planning and Spatial Sciences, University of Tasmania, Hobart, TAS 7001, Australia

\* Correspondence: n.nidhi@unsw.edu.au

**Abstract:** Australia experiences a variety of climate extremes that result in loss of life and economic and environmental damage. This paper provides a first evaluation of the performance of state-of-the-art Coupled Model Intercomparison Project Phase 6 (CMIP6) global climate models (GCMs) in simulating climate extremes over Australia. Here, we evaluate how well 37 individual CMIP6 GCMs simulate the spatiotemporal patterns of 12 climate extremes over Australia by comparing the GCMs against gridded observations (Australian Gridded Climate Dataset). This evaluation is crucial for informing, interpreting, and constructing multimodel ensemble future projections of climate extremes over Australia, climate-resilience planning, and GCM selection while conducting exercises like dynamical downscaling via GCMs. We find that temperature extremes (maximum-maximum temperature -TXx, number of summer days -SU, and number of days when maximum temperature is greater than 35 °C -Txge35) are reasonably well-simulated in comparison to precipitation extremes. However, GCMs tend to overestimate (underestimate) minimum (maximum) temperature extremes. GCMs also typically struggle to capture both extremely dry (consecutive dry days -CDD) and wet (99th percentile of precipitation -R99p) precipitation extremes, thus highlighting the underlying uncertainty of GCMs in capturing regional drought and flood conditions. Typically for both precipitation and temperature extremes, UKESM1-0-LL, FGOALS-g3, and GCMs from Met office Hadley Centre (HadGEM3-GC31-MM and HadGEM3-GC31-LL) and NOAA (GFDL-ESM4 and GFDL-CM4) consistently tend to show good performance. Our results also show that GCMs from the same modelling group and GCMs sharing key modelling components tend to have similar biases and thus are not highly independent.

**Keywords:** climate change adaptation; climate extremes; CMIP6; ET-SCI indices; extreme weather; regional climate modelling



**Citation:** Nishant, N.; Di Virgilio, G.; Ji, F.; Tam, E.; Beyer, K.; Riley, M.L. Evaluation of Present-Day CMIP6 Model Simulations of Extreme Precipitation and Temperature over the Australian Continent. *Atmosphere* **2022**, *13*, 1478. <https://doi.org/10.3390/atmos13091478>

Academic Editor: Paolo Stocchi

Received: 9 August 2022

Accepted: 7 September 2022

Published: 12 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Climate extremes threaten human health, economic stability, and the well-being of natural and built environments. As the world continues to warm, climate hazards are expected to increase in frequency and intensity [1]. Climate extremes have increased on all continents since 1980, most notably in Australia, and this has dramatically impacted human life, economy and environment [2–5]. Zander et al. (2015) [6] showed that heat-related extremes during 2013–2014 led to an annual economic burden of around US\$6.2 billion (95% CI: 5.2–7.3 billion) for the Australian workforce. This amounts to 0.33 to 0.47% of Australia's GDP (Gross Domestic Product).

Australia is exposed to various climates due to its large size and meridional extent. With such a wide range of climates, Australia also experiences a great diversity of climatic extremes [2,5,7]. Australia regularly experiences extremes in precipitation and temperature: from severe droughts to intense precipitation [8–13] and from increases in hot extremes and decreases in cold extremes [14–16]. These extremes have a substantial impact on Australia's unique flora and fauna [17], agriculture [18], urban infrastructure [19], and human health [20]. Thus, it is crucial to examine the simulated extremes in Australia's present-day and future climate, as it is for other regions. This interpretation is an important aspect of adaptation planning, as changes to rare but high-impact climate extremes are likely to be a greater challenge to communities compared to changes in the mean climate state.

Global climate models (GCMs) have been used as a primary tool for examining the past and future changes in climate extremes at global and continental scales [21–25]. Several studies have conducted comprehensive reviews of projected changes in climate extremes globally and regionally based on the different phases of the Coupled Modelling Intercomparison Project (CMIP) [26–28]. Recently, the latest iteration of CMIP, phase 6 (CMIP6) [29], has been released, which provides new opportunities to examine the climate system and make regional projections via downscaling. Compared to CMIP5 [30], the models in CMIP6 generally have finer model resolution and improved physical processes [31]. One of the focuses of CMIP6 is assessing climate extremes changes and understanding associated physical processes [29].

Several studies have already evaluated the performance of CMIP6 GCMs in simulating extremes globally and over different regions of the world. However, most of these studies have evaluated how the multimodel ensemble of CMIP6 performs in comparison to the multimodel ensemble of CMIP5 (for example, Grose et al. 2020 [32], Di Virgilio et al. 2022 [33] and Deng et al. 2021 [34] over Australia; Masud et al. 2021 [35] over Canada; Gusain et al. 2020 [36] over India; Akinsanola et al. 2021 [37] over Eastern Africa; Ukkola et al. 2020 [38] and Seneviratne & Hauser 2020 [39] globally). Most of these studies have concluded that the CMIP6 GCM ensemble offers marginal improvements versus the CMIP5 ensemble in terms of capturing observed climate extremes. These studies assist in understanding how GCM ensembles perform in simulating climate extremes; however, to understand the physical processes driving modelled climate extremes, knowledge of the performances of individual GCMs is needed. Assessment of the performance of individual GCMs in resolving climate extremes also assists with the selection of GCMs for regional dynamical downscaling [40,41] and for assessing the potential impacts of extreme events on biophysical and socio-economic systems [42].

There have been a few studies that have looked at the performance of individual CMIP6 GCMs in capturing observed climate extremes over different parts of the globe: for example, the United States [43], southeast Asia [44], China and Western North Pacific and East Asia [45]. Over Australia, Di Virgilio et al. 2022 [33] evaluated CMIP6 GCMs for climate extremes; however, their study was only limited to percentile-based temperature and precipitation extremes. Understanding the relative performances of CMIP6 GCMs in simulating climate extremes over Australia is of direct relevance to multiple stakeholders, such as researchers and climate impact resilience planners, as GCMs can show a wide variety of inter-model differences in performance.

Therefore, the aims of this study are: 1. To evaluate how well individual CMIP6 GCMs perform in terms of capturing precipitation and temperature extremes over Australia. 2. To find the subset of best performing CMIP6 GCMs for capturing these climate extremes. 3. To assess if the GCM performance varies for continental Australia and sub-regions across Australia with different topography and climate.

## 2. Data and Methods

### 2.1. CMIP6 GCMs

This study uses one simulation each from a group of 37 CMIP6 GCMs that are available at the time of writing. Out of 37 CMIP6 GCMs, some GCMs are from the same parent institute and share modelling components such as GCMs from Australian Community, whereas some GCMs just use different resolutions such as MPI-ESM1-2-HR and MPI-ESM1-2-LR. The spatial resolution of the GCMs varies (Table 1). To compare the GCMs with the observations, we re-grid all 37 GCMs to the most common resolution of  $1.5^\circ \times 1.5^\circ$  on a regular lat/lon grid using conservative re-gridding. We acknowledge that re-gridding coarser resolution models onto a higher resolution has limitations [46]; however, going to any coarser resolution than  $1.5^\circ \times 1.5^\circ$  would lead to fewer grid points in the domain and thus affect the robustness of the analysis. In addition, past studies have also used similar re-gridding models to the most common resolution when analysing datasets like CMIP5 and CMIP6 [32,33]. We chose the most common present-day period between GCMs, i.e., 1951–2014, for the evaluation. The reason for choosing a longer period (64 years) is to minimize inter-annual variations.

**Table 1.** CMIP6 Models and simulations used in this study. Here colours other than black denote GCMs from the same modelling institution.

GCM Name	Institution/Centre	Run	Atmosphere Lat/lon Grid ( $^\circ$ )
1. ACCESS-CM2	Australian Community	r1ilp1f1	$1.2 \times 1.8$
2. ACCESS-ESM1-5	Australian Community	r1ilp1f1	$1.2 \times 1.8$
3. AWI-ESM-1-1-LR	Alfred Wegener Institute	r1ilp1f1	$0.9 \times 0.9$
4. BCC-CSM2-MR	Beijing Climate Centre	r1ilp1f1	$1.1 \times 1.1$
5. BCC-ESM1	Beijing Climate Centre	r1ilp1f1	$2.8 \times 2.8$
6. CMCC-CM2-SR5	Euro-Mediterranean Centre	r1ilp1f1	$\sim 0.9$
7. CNRM-CM6-1-HR	National Centre of Meteorological Research (NCMR), France	r1ilp1f2	$\sim 0.5$
8. CNRM-CM6-1	National Centre of Meteorological Research (NCMR), France	r1ilp1f2	$1.4 \times 1.4$
9. CNRM-ESM2-1	National Centre of Meteorological Research (NCMR), France	r1ilp1f2	$1.4 \times 1.4$
10. CanESM5	Canadian Centre for Climate Modelling and Analysis	r1ilp1f1	$2.8 \times 2.8$
11. EC-Earth3-Veg-LR	EC-EARTH consortium, The Netherlands/Ireland	r1ilp1f1	$0.7 \times 0.7$
12. EC-Earth3-Veg	EC-EARTH consortium, The Netherlands/Ireland	r1ilp1f1	$0.7 \times 0.7$
13. EC-Earth3	EC-EARTH consortium, The Netherlands/Ireland	r1ilp1f1	$0.7 \times 0.7$
14. FGOALS-f3-L	Chinese Academy of Sciences, China	r1ilp1f1	$2.3 \times 2.0$
15. FGOALS-g3	Chinese Academy of Sciences, China	r1ilp1f1	$2.3 \times 2.0$
16. GFDL-CM4	NOAA Geophysical Fluid Dynamics Laboratory	r1ilp1f1	$1.0 \times 1.3$
17. GFDL-ESM4	NOAA Geophysical Fluid Dynamics Laboratory	r1ilp1f1	$1.0 \times 1.3$
18. HadGEM3-GC31-LL	Met Office Hadley Centre, UK	r1ilp1f3	$2.2 \times 2.2$
19. HadGEM3-GC31-MM	Met Office Hadley Centre, UK	r1ilp1f3	$0.9 \times 0.9$
20. INM-CM4-8	Institute for Numerical Mathematics (INM), Russia	r1ilp1f1	$1.5 \times 2.0$
21. INM-CM5-0	Institute for Numerical Mathematics (INM), Russia	r1ilp1f1	$1.5 \times 2.0$
22. IPSL-CM6A-LR	Institute Pierre Simon Laplace, France	r1ilp1f1	$1.3 \times 2.5$
23. KACE-1-0-G	National Institute of Meteorological Sciences/Korea Meteorological Administration	r1ilp1f1	$2.2 \times 2.2$
24. KIOST-ESM	Korean Institute of Ocean Science and technology	r1ilp1f1	$2.2 \times 2.2$
25. MIR $^{\circ}$ C-ES2L	National Institute for Environmental Studies, Japan	r1ilp1f2	$4.5 \times 4.5$
26. MIR $^{\circ}$ C6	National Institute for Environmental Studies, Japan	r1ilp1f1	$1.4 \times 1.4$
27. MPI-ESM1-2-HAM	Max Planck Institute for Meteorology (MPI), Germany	r1ilp1f1	$2.2 \times 2.2$
28. MPI-ESM1-2-HR	Max Planck Institute for Meteorology (MPI), Germany	r1ilp1f1	$\sim 0.9$
29. MPI-ESM1-2-LR	Max Planck Institute for Meteorology (MPI), Germany	r1ilp1f1	$\sim 2.0$
30. MRI-ESM2-0	Meteorological Research Institute, Japan	r1ilp1f1	$1.1 \times 1.1$
31. NESM3	Nanjing University of Information Science and Technology, Nanjing	r1ilp1f1	$1.9 \times 1.9$
32. NorCPM1	Norwegian Climate Centre, Norway	r1ilp1f1	$1.9 \times 2.5$
33. NorESM2-LM	Norwegian Climate Centre, Norway	r1ilp1f1	$1.9 \times 2.5$
34. NorESM2-MM	Norwegian Climate Centre, Norway	r1ilp1f1	$0.9 \times 0.9$
35. SAM0-UNICON	Seoul National University	r1ilp1f1	$0.9 \times 1.3$
36. TaiESM1	Taiwan Earth System Model	r1ilp1f1	$0.9 \times 0.9$
37. UKESM1-0-LL	UK Met Office and NERC research centres	r1ilp1f2	$1.3 \times 1.9$

While extreme climate and weather events are generally multifaceted phenomena, in this study, we evaluate climate extremes based on daily precipitation and temperature as defined by the Expert Team on Sector-specific Climate Indices (ET-SCI; [47,48]). We use the ClimPACT version 2 software to calculate the ET-SCI indices (<https://climpact-sci.org/>; accessed on 1 November 2020), focusing on daily precipitation, maximum temperature, and minimum temperature.

Although ClimPACT produces 34 core indices (more than 80 indices in total), we use 12 indices based on the following considerations. 1. To capture key aspects of climate extremes; for example, we choose absolute indices (e.g., maximum 1-day precipitation

(Rx1day), hottest day (TXx), coldest day (TNn)), threshold-based indices (e.g., number of heavy rain days (R10mm), tropical nights, i.e., the annual count of days when daily minimum temperature  $> 20^{\circ}\text{C}$  (TR), summer days, i.e., the annual count of days when daily maximum temperature  $> 25^{\circ}\text{C}$  (SU), number of days when maximum temperature is greater than  $35^{\circ}\text{C}$  (TXge35)), percentile indices (e.g., total annual precipitation from very heavy rain days (R99p)), and duration indices (e.g., consecutive wet days (CWD), consecutive dry days (CDD), warm spell duration index (WSDI), cold spell duration index (CSDI)). 2. To capture extremes which have an impact on society and infrastructure; for example, extreme indices like TXge35, TR, and SU have the largest impacts on health [49], whereas indices like Rx1day, CDD, and CWD have the largest impact on agriculture, water resources and the economy [50,51].

## 2.2. Observations

We use observational data from the Australian Gridded Climate Dataset (AGCD Evans et al. 2020 [52]) to compare the CMIP6 GCM simulated extremes with observations for the present-day period. AGCD data have a spatial resolution of  $0.05^{\circ} \times 0.05^{\circ}$  and are obtained from an interpolation of station observations across the Australian continent. Most of these stations are in the more heavily populated coastal areas with a sparser representation inland over central Australia. Taking the same approach as for the GCMs, we re-grid the AGCD data to a common resolution of  $1.5^{\circ} \times 1.5^{\circ}$  and then calculate the ET-SCI indices using ClimPACT.

Since AGCD observations are only available for land, in this study, we evaluate CMIP6 GCMs only for grids over land. Due to the sparse distribution of precipitation gauges in central Australia, this region has no quality gridded observations available [53]. Therefore, we mask both ocean and parts of central Australia to examine precipitation extremes.

## 2.3. Methodology

Several studies have proposed multiple metrics for evaluating climate models. The metrics range from summarising features of the spatial distribution of the climatological mean state to summarising the temporal variability [54–57]. Some of the common and widely used metrics for evaluating the spatial climatology distribution of models against observations are biased (difference between the modelled and observed climatological mean field; here, smaller values indicate better performance of the model simulation), root mean square error (RMSE: square root of the difference between the modelled and observed climatological mean field; here smaller values indicate better performance of the model simulation) and pattern correlation (PCorr: correlation between the modelled and observed climatological mean field; Equation (1); Benesty et al. 2009 [58]). Here  $x_m$  and  $x_o$  denote the values of the variable in a sample of model simulations and the observational data sets, respectively, whereas  $\bar{x}_m$  and  $\bar{x}_o$  denote the means of the variable in a sample of model simulations and the observational data sets, respectively. Here, larger PCorr values indicate better performance of the model simulation.

$$PCorr = \frac{\sum (x_m - \bar{x}_m)(x_o - \bar{x}_o)}{\sqrt{\sum (x_m - \bar{x}_m)^2 \sum (x_o - \bar{x}_o)^2}} \quad (1)$$

For analysing the performance of the models with respect to temporal variation, studies have proposed the evaluation of interannual variability score (IVS, Equation (2); Chen et al. 2011 [59]; Fan et al. 2020 [60]). Here  $STD_m$  and  $STD_o$  denote the interannual standard deviation of the model simulations and the observational data sets, respectively. This score analyses the performance of models with respect to temporal variation and



suggests how well the models can reproduce the temporal standard deviation. Smaller IVS values indicate better performance of the model simulation.

$$IVS = \left( \frac{STD_m}{STD_o} - \frac{STD_o}{STD_m} \right)^2 \quad (2)$$

For each extreme index mentioned in Table 2, we calculate all the above-mentioned metrics, i.e., bias, RMSE, PCorr and IVS, to determine the performance of GCMs in capturing the observed extremes.

**Table 2.** List of ET-SCI Indices evaluated in this study.

	Index	Definition	Units	Timescale	Sectors
1.	R99p	Annual total precipitation when precipitation is greater than the 99th percentile	mm	Annual	Coasts
2.	Rx1day	Maximum 1-day precipitation	mm	Annual/Monthly	Agriculture, Forestry
3.	R10mm	Number of very heavy rain days (rain > 10 mm)	days	Annual/Monthly	coasts
4.	CWD	Consecutive wet days	days	Annual/Monthly	Agriculture, Food security, Water resources
5.	CDD	Consecutive dry days	days	Annual/Monthly	Agriculture, Food security, Water resources
6.	TXge35	Number of days when maximum temperature is greater than 35 °C	days	Annual/Monthly	Health, Agriculture and Disaster and risk management
7.	TR	Tropical nights (Number of days when minimum temperature > 20 °C)	days	Annual/Monthly	Health, forestry
8.	SU	Summer days (Number of days when maximum temperature > 25 °C)	days	Annual/Monthly	Health, forestry
9.	TXx	Maximum maximum-temperature	°C	Annual/Monthly	Agriculture and food, Energy, forestry
10.	TNn	Minimum minimum-temperature	°C	Annual/Monthly	Agriculture and food, Energy, forestry
11.	CSDI	Cold Spell Duration Indicator (Annual count of nights with at least 6 consecutive nights when daily minimum temperature < 10th percentile)	days	Annual	Health, Energy, disaster risk reduction, Agriculture
12.	WSDI	Warm Spell Duration Indicator (Annual count of days with at least 6 consecutive days when daily maximum temperature > 90th percentile)	days	Annual	Health, Energy, disaster risk reduction, Agriculture

Not all metrics are independent, and some metrics will inevitably be correlated with others. Still, the effort has been made to include roughly the same number of metrics from each variable, minimizing this issue. Once performance metrics are calculated (i.e., biases, spatial RMSEs, spatial correlations, IVS score for each variable/statistic and for each timescale), these are normalized following Rupp et al. 2013 [54], as per Equation (3);

$$E_{i,j}^* = \frac{E_{i,j} - \min(E_{i,j})}{\max(E_{i,j}) - \min(E_{i,j})} \quad (3)$$

where  $E_{i,j}$  is an error for a given model  $i$  and metric  $j$ . Note that for correlations, each min () or max () function in Equation (1) is reversed. The relative error is then summed across all  $m$  metrics following Rupp et al. (2013), as per Equation (4):

$$E_{i,tot}^* = \sum_{i=1}^m E_{i,j}^* \quad (4)$$

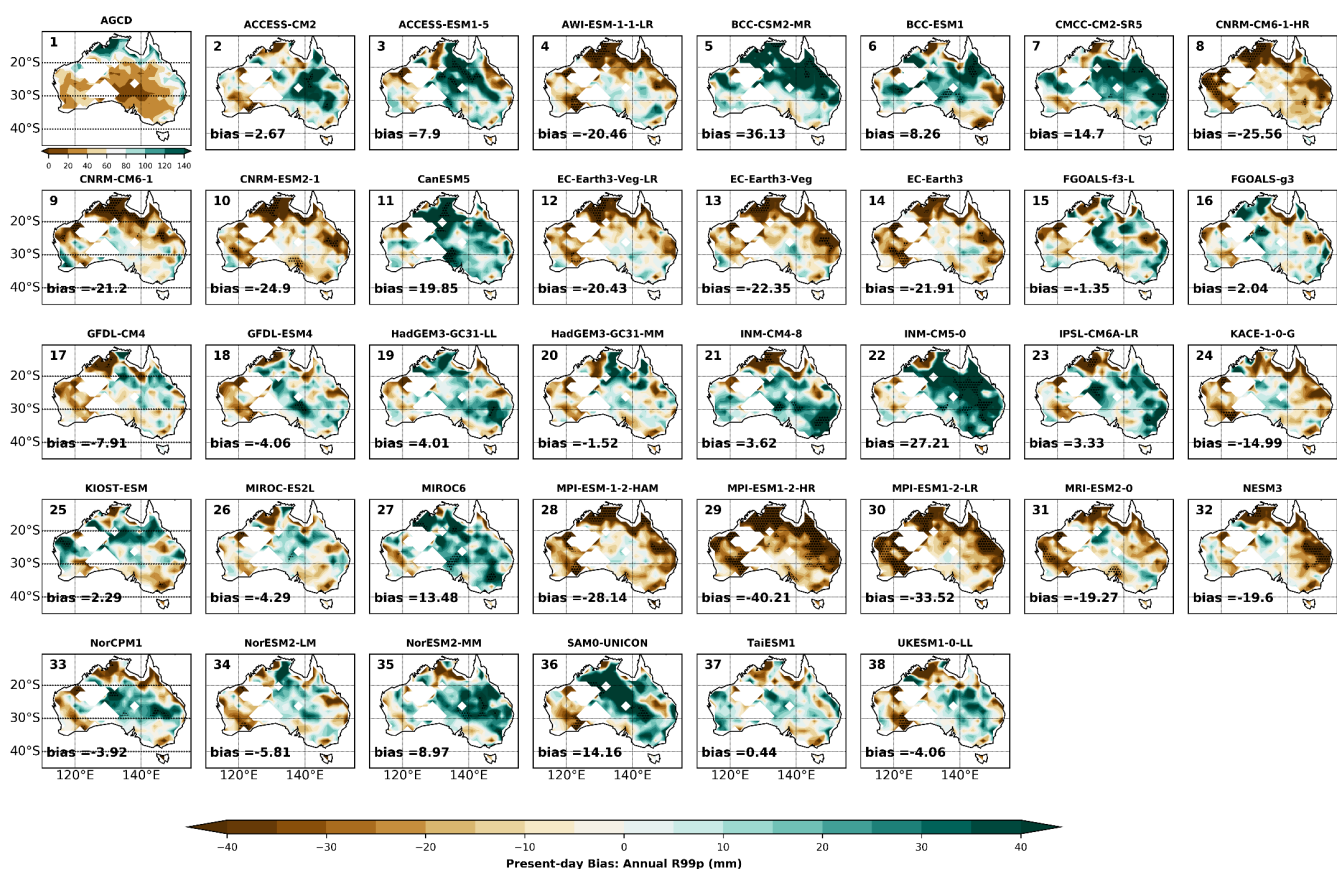
### 3. Results

We evaluate indices for either monthly or annual periods (Table 2). For the annual indices, evaluation metrics (bias, RMSE, Pcorr and IVS) are only calculated at the annual timescale, whereas for the monthly indices, the evaluation metrics are calculated at both the annual and seasonal (December-January-February (DJF), March-April-May (MAM), June-July-August (JJA) and September-October-November (SON)) timescales. Due to space limitations, we only show the biases at the annual scale in the main text. The biases at the seasonal scale are provided in the Supplementary Materials.

#### 3.1. Climatology of Indices

##### 3.1.1. Precipitation Extremes

Figures 1–5 show the observed climatological mean and biases in individual CMIP6 GCMs for annual R99p, Rx1day, R10mm, CWD, and CDD, respectively. When assessing biases, we calculate the statistical significance of bias for each grid cell using  $t$ -tests ( $\alpha = 0.05$ ) and Mann–Whitney  $U$  test for temperature and precipitation-based extremes, respectively. Stippling shows statistically significant bias.

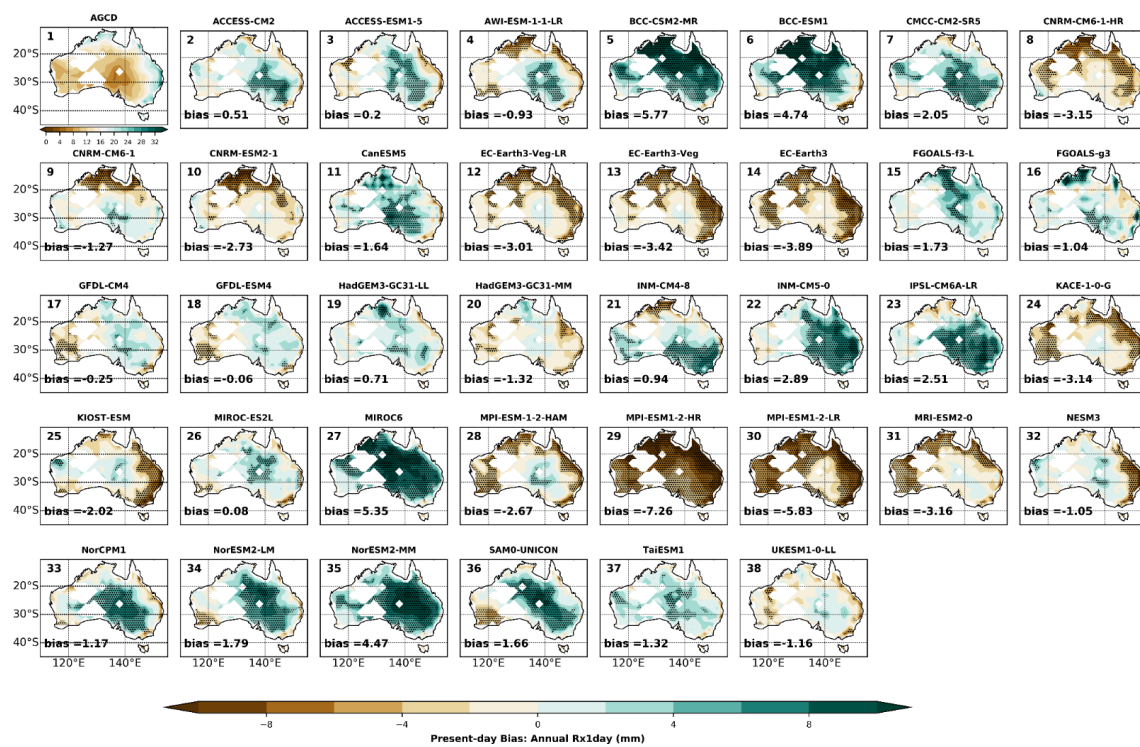


**Figure 1.** Climatological mean annual bias in total annual precipitation from very heavy rain days (R99p: mm) relative to the Australian Gridded Climate Data dataset (AGCD; panel 1) for the individual CMIP6 GCMs (panels 2–38). Data spans 1951–2014. Stippling indicates statistically significant differences using a student's  $t$ -test at the 95% confidence level. The white masks in part of the inland domain are the regions with no station data.

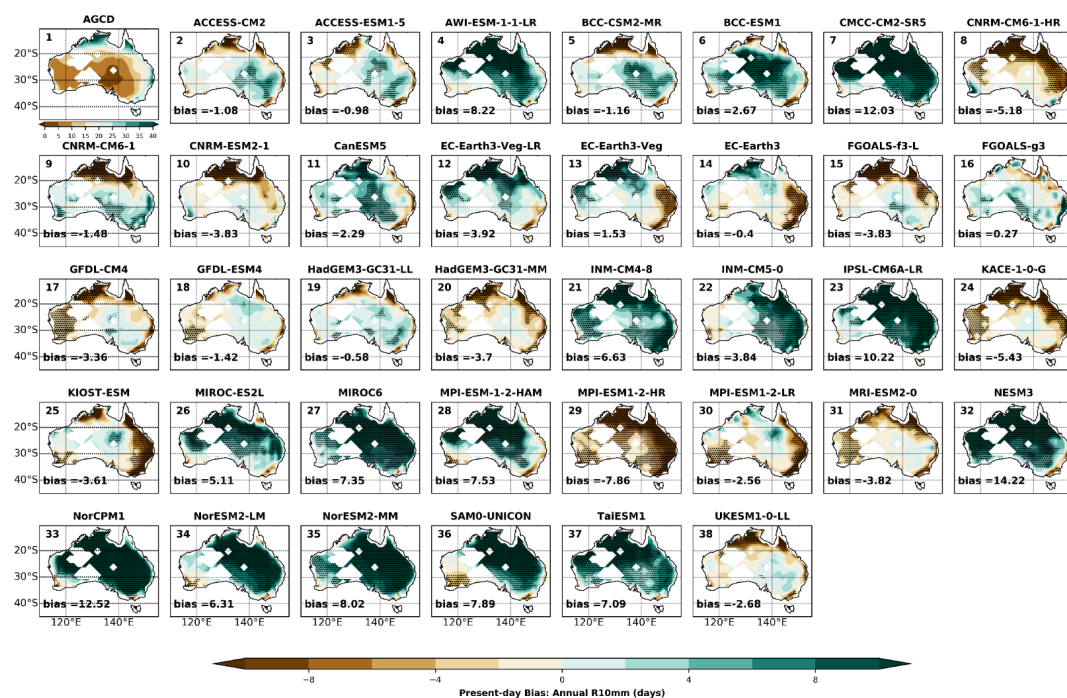
The spatial distribution of biases in R99p and Rx1day (Figures 1 and 2) look similar to each other. The majority of GCMs show the largest bias in the eastern half of Australia for both extremes. For R99p, TaiESM1 shows the least bias of 0.4 mm, and MPI-ESM1-1-HR and BCC-CSM2-MR show the maximum dry and wet bias of 40.2 and 36.1 mm, respectively. HadGEM3-GC31-MM, INM-CM4-8, TaiESM1, KIOST-ESM, ACCESS-CM2, FGOALS-f3-L and FGOALS-f3 show lower biases, whereas BCC-CSM2-MR and INM-CM5-0 show higher wet bias and CNRM-CM6-1-HR, MPI-ESM-1-2-HAM, MPI-ESM1-2-HR and MPI-ESM1-2LR (MPI Germany) show the highest dry bias respectively (Table 3).

**Table 3.** Table of continentally averaged absolute bias recorded in 37 CMIP6 GCMs for the 12 extreme indices. A smaller value of bias corresponds to the better performance of GCMs. Here, the colour green and red denote the best (top 25% subset) and worst (bottom 25% subset) performing GCMs.

	GCM	R99p	Rx1day	R10mm	CWD	CDD	Txge35	TR	SU	TXX	TNN	WSDI	CSDI
1	ACCESS-CM2	2.67	0.51	1.08	0.43	3.49	1.05	3.4	2.01	1.44	2.78	7.09	0.07
2	ACCESS-ESM1-5	7.9	0.2	0.98	0.69	4.12	0.13	3.63	1.25	0.45	2.56	5.07	0.81
3	AWI-ESM-1-1-LR	20.46	0.93	8.22	1.08	1.34	3.13	3.45	2.69	2.42	2.79	3.95	1.34
4	BCC-CSM2-MR	36.13	5.77	1.16	0.99	8.25	1.19	2.87	0.86	0.98	2.34	1.13	0.24
5	BCC-ESM1	8.26	4.74	2.67	0.58	6.12	0.19	4.62	0.68	0.1	4.2	0.55	0.06
6	CMCC-CM2-SR5	14.7	2.05	12.03	2.48	7.73	7.37	11	8.28	7.73	8.32	5.51	1.25
7	CNRM-CM6-1-HR	25.56	3.15	5.18	0.72	8.39	0.68	0.18	1.02	0.98	0.89	1.43	0.77
8	CNRM-CM6-1	21.2	1.27	1.48	0.18	3.52	1.18	1.51	1.21	1.35	1.74	3.73	0.09
9	CNRM-ESM2-1	24.9	2.73	3.83	0.5	8.86	0.02	2.19	0.45	0.69	2.08	2.65	0.1
10	CanESM5	19.85	1.64	2.29	1.12	6.03	1.38	2.97	0.39	1.34	1.71	5.88	1.42
11	EC-Earth3-Veg-LR	20.43	3.01	3.92	1.39	2.8	1.45	1.66	0.86	1.39	2.23	6.76	0.27
12	EC-Earth3-Veg	22.35	3.42	1.53	0.89	4.76	0.47	2.46	0.27	0.48	2.56	9.38	0.91
13	EC-Earth3	21.91	3.89	0.4	0.53	5.37	0.06	2.6	0.72	0.08	2.82	13.37	1.19
14	FGOALS-f3-L	1.35	1.73	3.83	0.62	6.99	1.83	1.33	2.55	1.95	2.64	8.03	1.33
15	FGOALS-g3	2.04	1.04	0.27	0.91	8.18	0.8	2.5	0.94	1.01	3.1	2.23	0.67
16	GFDL-CM4	7.91	0.25	3.36	0.25	3.05	2.23	0.92	3.49	1.97	0.09	2.08	0.54
17	GFDL-ESM4	4.06	0.06	1.42	0.04	6.54	1.88	2.78	2.82	1.79	1.81	3.66	1.37
18	HadGEM3-GC31-LL	4.01	0.71	0.58	0.45	4.17	0.83	0.52	1.48	0.73	0.8	3.84	1.62
19	HadGEM3-GC31-MM	1.52	1.32	3.7	0.03	1.84	0.63	0.21	0.48	0.22	0.46	1.96	1.89
20	INM-CM4-8	3.62	0.94	6.63	3.69	14.56	2.24	1.77	1.64	1.84	1.67	3.87	0.19
21	INM-CM5-0	27.21	2.89	3.84	2.76	13.8	2.03	0.6	2.51	2.01	0.49	1.98	0.83
22	IPSL-CM6A-LR	3.33	2.51	10.22	2.51	8.08	5.01	2.91	3.92	4.05	2.61	6.94	1.28
23	KACE-1-0-G	14.99	3.14	5.43	0.14	0.33	0.66	0.32	0.45	0.1	0.87	3.09	1.67
24	KIOST-ESM	2.29	2.02	3.61	1.15	1.62	8	3.43	3.97	8.54	2.56	0.4	0.59
25	MIR <sup>o</sup> C-ES2L	4.29	0.08	5.11	2.2	9.67	1.77	4.58	0.72	1.01	4.72	3.93	1.12
26	MIR <sup>o</sup> C6	13.48	5.35	7.35	1.25	10.31	4.88	4.56	1.75	4.75	4.51	1.86	0.34
27	MPI-ESM-1-2-HAM	28.14	2.67	7.53	0.55	4.5	2.57	3.46	1.8	2.11	2.93	1.45	1.13
28	MPI-ESM1-2-HR	40.21	7.26	7.86	1.02	30.31	1.09	4.98	0.03	0.3	3.95	2.42	0.51
29	MPI-ESM1-2-LR	33.52	5.83	2.56	0.51	19.27	0.45	4.41	0.76	0.39	3.1	4.25	0.01
30	MRI-ESM2-0	19.27	3.16	3.82	0.09	0.44	1.99	3.59	0.08	1.81	2.09	0.51	1.37
31	NESM3	19.6	1.05	14.22	1.04	5.21	6.13	9.27	7.16	6.61	7.05	5.01	0.27
32	NorCPM1	3.92	1.17	12.52	3.19	10.02	6.07	0.56	5	5.11	1.93	2.57	1.96
33	NorESM2-LM	5.81	1.79	6.31	1.34	6.61	2.81	4.65	2.79	2.53	4.43	1.45	0.71
34	NorESM2-MM	8.97	4.47	8.02	1.16	7.79	3.09	3.41	2.98	2.48	3.54	0.11	0.39
35	SAM0-UNICON	14.16	1.66	7.89	1.62	2.66	2.58	1.92	1.64	1.81	2.55	6	1.39
36	TaiESM1	0.44	1.32	7.09	1.62	6.41	7.1	11.14	8.52	7.34	8.35	3.9	0.95
37	UKESM1-0-LL	4.06	1.16	2.68	0.04	1.31	0.88	0.71	1.7	0.89	0.26	7.11	1.01

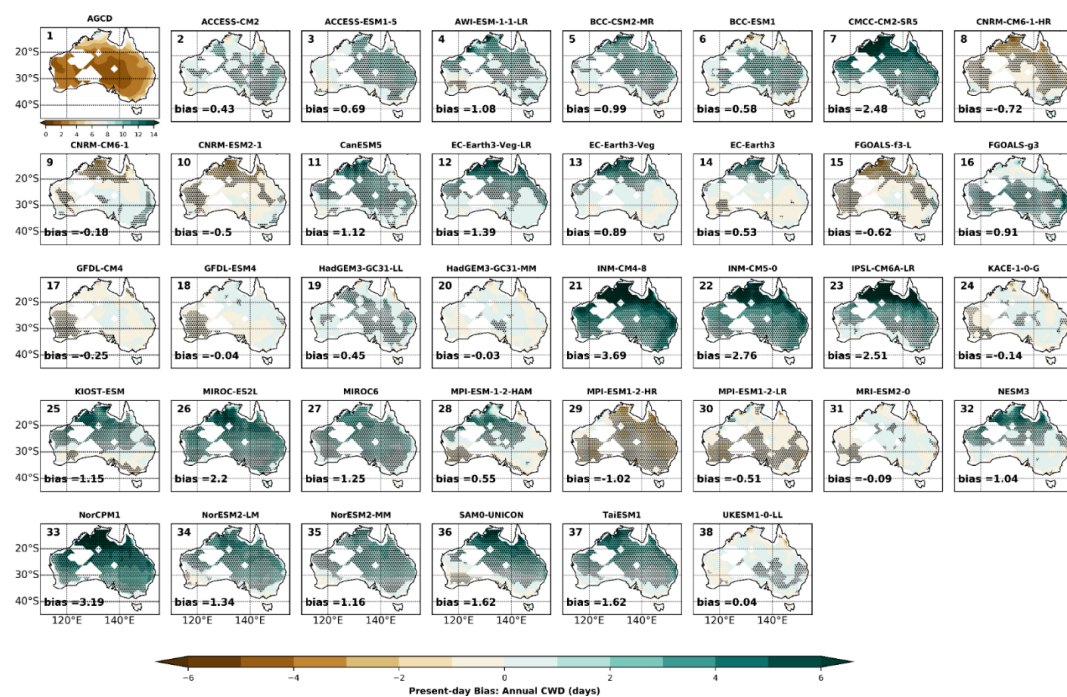


**Figure 2.** Climatological mean annual bias in maximum 1-day precipitation (Rx1Day: mm) relative to the Australian Gridded Climate Data dataset (AGCD; panel 1) for the individual CMIP6 GCMs (panels 2–38). Data spans 1951–2014. Stippling indicates statistically significant differences using a student's *t*-test at the 95% confidence level. The white masks in part of the inland domain are the regions with no station data.

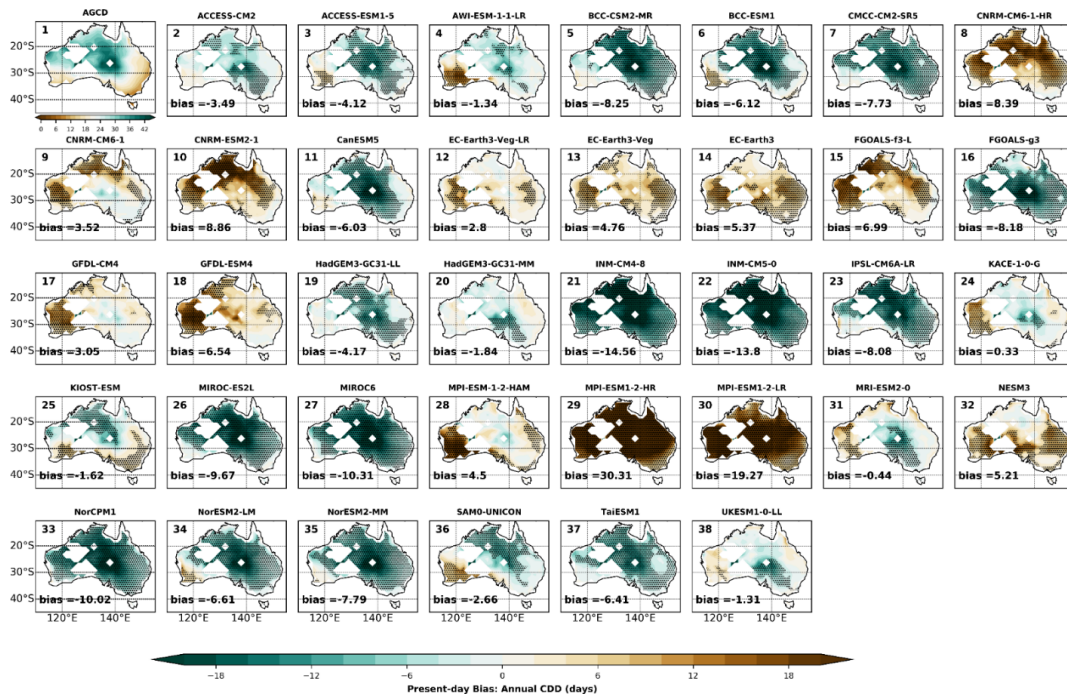


**Figure 3.** Climatological mean annual bias in the number of very heavy rain days (rain > 10 mm) (R10mm: days) relative to the Australian Gridded Climate Data dataset (AGCD; panel 1) for the individual CMIP6 GCMs (panels 2–38). Data spans 1951–2014. Stippling indicates statistically significant differences using a student's *t*-test at the 95% confidence level. The white masks in part of the inland domain are the regions with no station data.





**Figure 4.** Climatological mean annual bias in consecutive wet days (CWD: days) relative to the Australian Gridded Climate Data dataset (AGCD; panel 1) for the individual CMIP6 GCMs (panels 2–38). Data spans 1951–2014. Stippling indicates statistically significant differences using a student's *t*-test at the 95% confidence level. The white masks in part of the inland domain are the regions with no station data.



**Figure 5.** Climatological mean annual bias in consecutive dry days (CDD: days) relative to the Australian Gridded Climate Data dataset (AGCD; panel 1) for the individual CMIP6 GCMs (panels 2–38). Data spans 1951–2014. Stippling indicates statistically significant differences using a student's *t*-test at the 95% confidence level. The white masks in part of the inland domain are the regions with no station data.



For Rx1day, GCMs typically show absolute mean biases in the range of 0.1 mm (MIR<sup>°</sup>C-ES2L) and 7.3 mm (MPI-ESM1-2-HR). Here, ACCESS-CM2, ACCESS-ESM1-5 (Australian Community), AWI-ESM-1-1-LR, FGOALS-g3, HadGEM3-GC31-LL, INM-CM4-8, GFDL-ESM4 and GFDL-CM4 (NOAA) show the lowest bias whereas BCC-CSM2-MR, BCC-ESM1 (Beijing Climate Centre), MPI-ESM1-2-HR, MPI-ESM1-2LR, MIR<sup>°</sup>C6, NorESM2-MM, EC-Earth3-Veg and EC-Earth3 (EC-Earth Consortium) show the largest bias (~ 4–7 mm) (Table 3) both wet and dry.

For both R99p and Rx1day, there are two subsets of GCMs showing greater dry bias (EC-Earth3-Veg-LR, EC-Earth3-Veg and EC-Earth3 (EC-Earth Consortium), MPI-ESM-1-2-HAM, MPI-ESM1-2-HR and MPI-ESM1-2-LR (MPI-Germany)) and wet bias (BCC-CSM2-MR and BCC-ESM1 (Beijing Climate Centre), INM-CM4-8 and INM-CM5-0 (INM Russia)).

For R10mm (Figure 3), ACCESS-CM2, ACCESS-ESM1-5 (Australian Community), BCC-CSM2-MR, CNRM-CM6-1, FGOALS-g3, GFDL-ESM4, HadGEM3-GC31-LL show the smallest wet bias whereas AWI-ESM-1-1-LR, CMCC-CM2-SR5, IPSL-CM6A-LR, MPI-ESM-1-2-HAM, MPI-ESM1-2-HR, NESM3, NorCPM1, NorESM2-MM and SAM0-UNICON show the largest wet bias. Like R99p and Rx1day, GCMs for R10mm also show greater bias in eastern Australia than elsewhere.

For CDD and CWD (Figures 4 and 5), GCMs do not show large spatial variation in the sign of the bias. In addition to this, the GCMs with dry and wet biases in R99p and Rx1day overestimate (CDD) and underestimate (CWD) consecutive days, respectively. For CWD, GCMs typically show a small bias (between 0.5–2.5 days); however, there are outlier GCMs (INM-CM4-8 and NorCPM1) which show biases greater than 3 days. Here, ACCESS-CM2, GFDL-ESM4, GFDL-CM4 (NOAA), HadGEM3-GC31-LL and HadGEM3-GC31-MM (Hadley Centre) are found to be top performers (top 25% subset). In addition, CNRM-CM6-1, KACE-1-0-G, MRI-ESM2-0 and UKESM1-0-LL also show small biases and are among the top performers. For CDD, GCMs typically show a very strong bias (between 4–10 days) compared to CWD. Outlier GCMs show bias greater than 10 days; for example, MPI-ESM1-2-HR show the strongest bias of 30.3 days, whereas INM-CM4-8 and INM-CM5-0 (INM Russia) show biases greater than 15 days.

The large biases in CDD for GCMs from CNRM, INM Russia, MPI Germany, and National Institute for Environmental Studies Japan suggest that some GCMs typically struggle to capture the CDD. This can be attributed to the drizzle problem in GCMs. Past studies have suggested that convective precipitation generated by the cumulus parameterisation in coarse resolution GCMs is too frequent and long-lasting with reduced intensity, leading to the “drizzling” bias [61]. The drizzling bias impedes realistic representation of precipitation extremes like CDD and results in a large bias.

Spatial maps of biases at seasonal timescales for Rx1day (Figures S1–S4), R10mm (Figures S5–S8), CWD (Figures S9–S12) and CDD (Figures S13–S16) are provided in the Supplementary Material. A comparison of seasonal biases showed a similar story as the annual biases for all the precipitation extremes, i.e., GCMs with annual dry and wet biases demonstrate similar seasonal dry and wet biases, and those biases have similar spatial distributions.

We observe that models from the same parent institution tend to have similar biases for precipitation extremes (i.e., in terms of spatial structure, sign and magnitude). For example, both GCMs from the Beijing Climate Centre: BCC-CSM2-MR and BCC-ESM1, show wet biases in most parts of the continent. Similarly, GCMs from Norwegian Climate Centre: NorCPM1, NorESM2-LM and NorESM2-MM and GCMs from INM Russia: INM-CM4-8 and INM-CM5-0 also have wet biases over most parts of Australia. The GCMs from MPI Germany: MPI-ESM-1-2-HAM, MPI-ESM1-2-HR, and MPI-ESM1-2-LR, respectively, show dry biases across the continent. We note that GCMs from the same parent institution often share key modelling components such as land-surface components, atmospheric and °Ocean models and are thus not highly independent of each other.

Although GCMs from the same parent institution can show large inter-model similarity, GCMs from the Chinese Academy of Sciences appear more independent. Here,

FGOALS-f3-L shows dry bias, whereas FGOALS-g3 shows wet bias for most precipitation indices except for Rx1day. After investigating details about the model configurations, we found that both these GCMs have substantial differences in their components. For example, FGOALS-f3-L is a fully coupled climate system model consisting of four component models, whereas FGOALS-g3 comprises five components [62]. In FGOALS-f3-L, all the component models are coupled via the NCAR Coupler 7 [63], whereas, in FGOALS-g3, all the component models are coupled via Community Coupler, version 2 (C-Coupler2), developed by Tsinghua University [64]. These differences between the modelling components of these two GCMs are likely the reason for their dissimilarity.

Typically, for most precipitation extremes, the GCMs which show large bias also show large RMSE. At the annual timescale, for Rx1day and R99p, CNRM-CM6-1-HR, MPI-ESM-1-2-HAM, NESM3, and EC-Earth3-Veg-LR show the smallest RMSE, whereas BCC-CSM2-MR, BCC-ESM1 (Beijing Climate Centre), CanESM5, FGOALS-f3-L, INM-CM5-0, MIR°C6 and NorESM2-MM show the largest RMSE (Table 4). For R10mm and CWD, HadGEM3-GC31-LL, HadGEM3-GC31-MM (Hadley Centre), GFDL-ESM4, GFDL-CM4 (NOAA), and UKESM1-0-LL show the smallest RMSE whereas SAM0-UNICON, NorCPM1, IPSL-CM6A-LR and CMCC-CM2-SR5 show the largest RMSE.

**Table 4.** Table of continentally averaged root mean square error (RMSE) recorded in 37 CMIP6 GCMs for the 12 extreme indices. A smaller value of RMSE corresponds to the better performance of GCMs. Here, the colour green and red denote the best (top 25% subset) and worst (bottom 25% subset) performing GCMs.

	GCM	R99p	Rx1day	R10mm	CWD	CDD	Txge35	TR	SU	TXX	TNN	WSDI	CSDI
0	ACCESS-CM2	74.87	5.71	0.65	1.05	7.5	2.25	3.47	2.3	1.81	2.79	18.7	4.66
1	ACCESS-ESM1-5	73.24	5.46	0.58	1.08	7.84	1.98	3.67	1.78	1.28	2.58	15.37	4.12
2	AWI-ESM-1-1-LR	59.12	5.32	0.97	1.54	9.38	3.49	3.59	2.96	2.58	2.81	17.41	5.37
3	BCC-CSM2-MR	92.9	8.11	0.62	1.28	9.56	1.93	3.1	1.44	1.45	2.48	11.12	4.15
4	BCC-ESM1	77.92	7.66	0.72	1.22	8.9	1.88	4.72	1.56	1.44	4.23	12.65	4.5
5	CMCC-CM2-SR5	76	5.62	1.15	2.62	9.19	7.37	11	8.28	7.73	8.32	16.22	3.76
6	CNRM-CM6-1-HR	57.55	5.14	0.66	1.03	11.53	1.82	1.83	1.61	1.24	1.33	11.96	4.08
7	CNRM-CM6-1	60.66	5.14	0.65	0.93	9.43	2.14	2.29	1.9	1.6	1.9	14.28	4.89
8	CNRM-ESM2-1	60.04	5.42	0.66	0.98	13.31	1.76	2.7	1.66	1.2	2.2	12.85	4.52
9	CanESM5	82.74	6.4	0.7	1.5	9.15	2.66	3.12	1.79	2.48	2.25	16.77	3.76
10	EC-Earth3-Veg-LR	58.24	5.11	0.75	1.71	8.47	2.67	2.21	2.09	1.9	2.38	17.32	4.1
11	EC-Earth3-Veg	58.84	5.29	0.66	1.37	9.79	2.52	2.71	1.71	1.43	2.61	18.36	3.94
12	EC-Earth3	57.8	5.31	0.59	1.13	9.66	2.39	2.85	1.61	1.24	2.85	20.48	3.61
13	FGOALS-f3-L	73.3	6.41	0.64	1.05	11.76	2.31	1.86	2.76	2.05	2.68	17.66	3.56
14	FGOALS-g3	68.62	5.37	0.61	1.29	9.52	1.91	2.73	1.59	1.48	3.12	12.86	3.74
15	GFDL-CM4	65.14	5.22	0.56	0.81	8.64	2.61	1.6	3.52	2.06	1.11	13.51	4.21
16	GFDL-ESM4	68.16	5.48	0.56	0.9	11.76	2.48	2.93	3.01	1.96	1.92	14.85	3.46
17	HadGEM3-GC31-LL	71.27	5.25	0.59	1.01	7.45	2.11	1.73	1.82	1.33	1.67	14.47	3.63
18	HadGEM3-GC31-MM	68.15	5.03	0.55	0.75	6.68	1.72	1.48	1.19	0.98	1.29	12.15	3.33
19	INM-CM4-8	70.85	5.35	0.8	3.72	14.6	2.92	2.23	2.09	2.06	2.24	14.42	4.08
20	INM-CM5-0	88.48	6.26	0.71	2.81	13.86	2.86	1.91	2.73	2.17	1.92	14.45	4.8
21	IPSL-CM6A-LR	69.71	6.03	1.09	2.72	9.96	5.06	3.02	3.95	4.06	2.64	16.42	3.77
22	KACE-1-0-G	61.6	5.36	0.64	0.89	8.05	1.82	1.77	1.44	1.26	1.68	13.22	3.34
23	KIOST-ESM	69.4	5.47	0.62	1.55	7.83	8.18	3.58	4.38	8.59	2.71	15.19	4.85
24	MIR°C-ES2L	61.93	4.74	0.74	2.3	10.43	2.49	4.65	1.51	1.77	4.72	14.34	5.14
25	MIR°C6	75.53	7.43	0.85	1.47	11.06	5.04	4.58	2.15	4.85	4.51	15.75	4.24
26	MPI-ESM-1-2-HAM	54.65	5.08	1.02	1.38	10.63	2.99	3.5	2.18	2.29	2.93	12.41	5.27
27	MPI-ESM1-2-HR	53.7	7.71	0.78	1.3	32.47	2.04	4.99	1.26	1.11	3.95	14.84	4.07
28	MPI-ESM1-2-LR	55.17	6.69	0.63	1.09	21.65	1.98	4.42	1.56	1.31	3.11	15.38	4.29
29	MRI-ESM2-0	62.8	5.69	0.63	0.86	8.42	2.51	3.62	1.49	2.23	2.21	13	3.79
30	NESM3	59.32	4.55	1.39	1.48	9.77	6.14	9.27	7.18	6.61	7.05	16.01	4.46
31	NorCPM1	66.8	5.41	1.27	3.28	10.86	6.07	2.02	5.03	5.12	2.13	14.84	5.57
32	NorESM2-LM	64.93	5.99	0.9	1.64	9.31	3.09	4.66	2.87	2.62	4.43	13.61	5.02
33	NorESM2-MM	72.54	6.86	0.88	1.38	9.36	3.21	3.5	3.06	2.52	3.54	12.47	4.43
34	SAM0-UNICON	79.68	6.16	1	1.96	8.65	3.14	2.25	2.08	2	2.63	15.61	5.53
35	TaiESM1	67.91	5.45	0.87	1.86	9.25	7.1	11.14	8.52	7.34	8.35	16.41	4.04
36	UKESM1-0-LL	71.46	5.09	0.57	0.85	7.25	1.91	1.71	1.92	1.35	1.55	15.48	3.88

For CDD, MPI-ESM1-2-HR and MPI-ESM1-2-LR (MPI Germany) show the largest RMSE, whereas HadGEM3-GC31-LL, HadGEM3-GC31-MM (Hadley Centre), UKESM1-0-LL, ACCESS-CM2 and ACCESS-ESM1-5 (Australian Community) show the smallest RMSE (Table 4).

For PCorr and IVS (Tables 5 and 6), GCMs show approximately similar results to the bias and RMSE evaluation, i.e., GCMs which show low bias and RMSE also capture spatial (have high PCorr) and temporal patterns (have low IVS score) well, like HadGEM3-GC31-LL, HadGEM3-GC31-MM, GFDL-ESM4, GFDL-CM4 and UKESM1-0-LL. There are also exceptions to this pattern. For example, EC-Earth-Veg-LR, EC-Earth3-Veg and EC-Earth3 (EC-EARTH Consortium) show top and average performance for PCorr and IVS, respectively, but they are not always in the top performing models for bias and RMSEs for most indices.

**Table 5.** Table of pattern correlation (PCorr) recorded in 37 CMIP6 GCMs for the 12 extreme indices. A larger value of PCorr corresponds to the better performance of GCMs. Here, the colour green and red denote the best (top 25% subset) and worst (bottom 25% subset) performing GCMs.

	GCM	R99p	Rx1day	R10mm	CWD	CDD	Txge35	TR	SU	TXX	TNN	WSDI	CSDI
0	ACCESS-CM2	0.62	0.79	0.9	0.87	0.81	0.96	0.98	0.98	0.96	0.97	0.67	0.41
1	ACCESS-ESM1-5	0.64	0.73	0.89	0.9	0.75	0.95	0.97	0.98	0.94	0.95	0.76	0.61
2	AWI-ESM-1-1-LR	0.43	0.69	0.85	0.83	0.65	0.88	0.98	0.98	0.96	0.95	0.64	0.53
3	BCC-CSM2-MR	0.8	0.77	0.83	0.87	0.62	0.97	0.96	0.99	0.95	0.94	0.68	0.32
4	BCC-ESM1	0.54	0.57	0.8	0.68	0.52	0.94	0.9	0.98	0.89	0.85	0.64	0.39
5	CMCC-CM2-SR5	0.62	0.8	0.83	0.77	0.78	0.37	0.96	0.95	0.95	0.98	0.67	0.32
6	CNRM-CM6-1-HR	0.64	0.86	0.72	0.79	0.83	0.97	0.94	0.98	0.97	0.96	0.77	0.39
7	CNRM-CM6-1	0.46	0.76	0.77	0.83	0.78	0.97	0.96	0.97	0.97	0.96	0.71	0.48
8	CNRM-ESM2-1	0.59	0.8	0.8	0.83	0.81	0.97	0.95	0.98	0.97	0.95	0.66	0.41
9	CanESM5	0.65	0.69	0.86	0.8	0.62	0.91	0.94	0.98	0.92	0.89	0.65	0.36
10	EC-Earth3-Veg-LR	0.63	0.89	0.92	0.91	0.84	0.84	0.97	0.97	0.92	0.95	0.73	0.22
11	EC-Earth3-Veg	0.66	0.9	0.9	0.91	0.86	0.85	0.97	0.98	0.92	0.96	0.74	0.31
12	EC-Earth3	0.73	0.91	0.92	0.93	0.85	0.87	0.97	0.98	0.93	0.96	0.68	0.13
13	FGOALS-f3-L	0.67	0.87	0.87	0.76	0.8	0.95	0.98	0.96	0.95	0.96	0.73	0.44
14	FGOALS-g3	0.72	0.74	0.86	0.81	0.69	0.97	0.97	0.98	0.94	0.93	0.73	0.54
15	GFDL-CM4	0.73	0.85	0.94	0.95	0.81	0.95	0.99	0.98	0.98	0.98	0.58	0.23
16	GFDL-ESM4	0.68	0.86	0.94	0.95	0.81	0.92	0.98	0.96	0.97	0.96	0.72	0.48
17	HadGEM3-GC31-LL	0.73	0.84	0.92	0.89	0.77	0.94	0.96	0.99	0.95	0.9	0.71	0.52
18	HadGEM3-GC31-MM	0.76	0.88	0.96	0.96	0.81	0.98	0.97	0.99	0.98	0.94	0.7	0.38
19	INM-CM4-8	0.36	0.68	0.92	0.81	0.72	0.78	0.95	0.98	0.94	0.87	0.74	0.43
20	INM-CM5-0	0.59	0.8	0.9	0.82	0.66	0.81	0.94	0.97	0.95	0.87	0.75	0.45
21	IPSL-CM6A-LR	0.63	0.74	0.77	0.8	0.59	0.65	0.97	0.97	0.98	0.95	0.7	0.27
22	KACE-1-0-G	0.72	0.82	0.91	0.87	0.73	0.95	0.95	0.98	0.91	0.91	0.7	0.32
23	KIOST-ESM	0.7	0.74	0.79	0.79	0.72	0.69	0.96	0.91	0.72	0.93	0.57	0.5
24	MIR <sup>2</sup> C-ES2L	0.8	0.8	0.87	0.81	0.65	0.92	0.9	0.97	0.87	0.89	0.62	0.48
25	MIR <sup>2</sup> C6	0.73	0.73	0.9	0.86	0.6	0.91	0.96	0.98	0.92	0.95	0.59	0.49
26	MPI-ESM1-2-HAM	0.64	0.73	0.71	0.69	0.59	0.92	0.97	0.99	0.95	0.96	0.76	0.42
27	MPI-ESM1-2-HR	0.49	0.72	0.88	0.86	0.78	0.98	0.98	0.99	0.98	0.97	0.65	0.55
28	MPI-ESM1-2-LR	0.69	0.78	0.86	0.88	0.75	0.95	0.98	0.99	0.94	0.96	0.72	0.46
29	MRI-ESM2-0	0.62	0.8	0.92	0.89	0.69	0.97	0.99	0.99	0.97	0.96	0.66	0.53
30	NESM3	0.65	0.76	0.76	0.8	0.77	0.44	0.94	0.92	0.89	0.96	0.74	0.19
31	NorCPM1	0.6	0.61	0.73	0.73	0.71	0.61	0.95	0.95	0.9	0.93	0.49	0.4
32	NorESM2-LM	0.75	0.68	0.79	0.8	0.59	0.9	0.97	0.98	0.93	0.95	0.6	0.61
33	NorESM2-MM	0.71	0.74	0.86	0.89	0.7	0.92	0.98	0.99	0.96	0.97	0.68	0.58
34	SAM0-UNICON	0.63	0.69	0.8	0.85	0.65	0.81	0.98	0.98	0.94	0.96	0.66	0.51
35	TaiESM1	0.77	0.9	0.9	0.87	0.82	0.51	0.96	0.93	0.96	0.97	0.67	0.27
36	UKESM1-0-LL	0.66	0.86	0.92	0.91	0.76	0.97	0.96	0.99	0.95	0.91	0.64	0.37

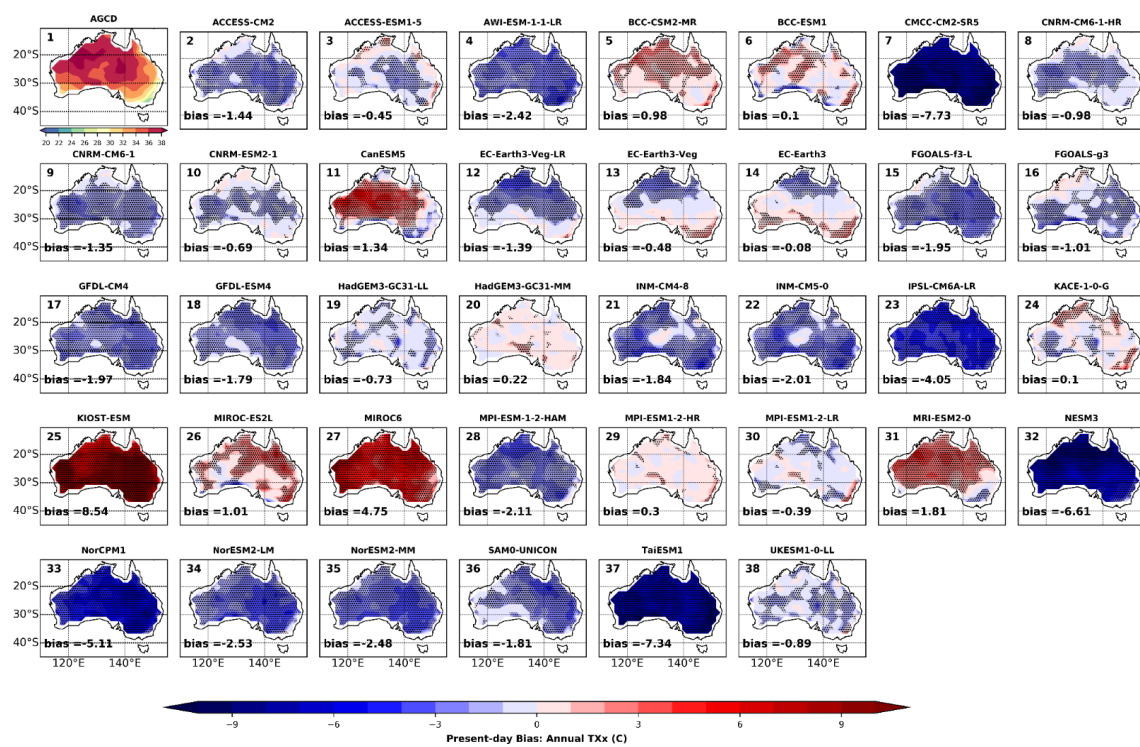
**Table 6.** Table of interannual variability score (IVS) recorded in 37 CMIP6 GCMs for the 12 extreme indices. A smaller value of IVS corresponds to the better performance of GCMs. Here, the colour green and red denote the best (top 25% subset) and worst (bottom 25% subset) performing GCMs.

	GCM	R99p	Rx1day	R10mm	CWD	CDD	Txge35	TR	SU	TXX	TNN	WSDI	CSDI
0	ACCESS-CM2	7.41	0.26	0.05	0.07	1.55	0.09	0.2	0.09	0.06	0.07	4.56	0.3
1	ACCESS-ESM1-5	5.57	0.15	0.04	0.04	1.59	0.36	0.14	0.23	0.03	0.14	2.2	0.49
2	AWI-ESM-1-1-LR	6.15	0.06	0.17	0.28	0.85	0.15	0.32	0.65	0.43	0.1	5.46	0.6
3	BCC-CSM2-MR	6.53	0.18	0.07	0.01	3.01	0.55	0.04	0.08	0.14	0.05	3.06	0.46
4	BCC-ESM1	1.44	0.4	0.03	0.06	2.39	0.42	0.03	0.17	0.09	0.07	1.17	0.19
5	CMCC-CM2-SR5	12.05	1.12	0.28	0.51	0.38	0.73	0.4	0.29	0.14	0.28	7.59	0.33
6	CNRM-CM6-1-HR	6.13	0.65	0.11	0.12	1.06	0.03	0.22	0.08	0.14	0	0.24	0.37
7	CNRM-CM6-1	5.41	0.45	0.05	0.06	0.04	0.26	0.22	0.13	0.08	0.11	0.07	0.55
8	CNRM-ESM2-1	4.2	0.37	0.04	0.04	0.77	0.21	0.1	0.24	0.05	0.03	0.18	0.23
9	CanESM5	6.31	0.19	0.06	0.12	1.75	0.13	0.03	0.26	0.14	0.05	3.35	0.96
10	EC-Earth3-Veg-LR	3.38	0.25	0.13	0.48	0.03	0.1	0.17	0.68	0.41	0.03	7.03	0.06
11	EC-Earth3-Veg	5.85	0.56	0.01	0.22	0.01	0.14	0.15	0.26	0.07	0.03	4.94	0.6
12	EC-Earth3	4.99	0.7	0.01	0.05	0.92	0.11	0.15	0.08	0.09	0.05	6.74	0.64
13	FGOALS-f3-L	1.64	0.57	0.07	0.1	0.06	0.35	0.17	0.21	0.1	0	3.71	0.41
14	FGOALS-g3	4.02	0.06	0.02	0.11	2.21	0.32	0.02	0.02	0.14	0.11	1.55	0.16
15	GFDL-CM4	1.18	0.08	0.05	0.03	0.87	0.3	0.12	0.1	0.05	0.18	1.09	0.24
16	GFDL-ESM4	0.61	0.3	0.01	0.13	1.72	0.3	0.21	0.37	0.06	0.04	2.8	0.82
17	HadGEM3-GC31-LL	0.3	0.27	0.05	0.01	1.9	0.37	0.17	0.03	0.11	0.14	1.53	1.2
18	HadGEM3-GC31-MM	0.8	0.46	0.11	0.1	2.02	0.46	0.11	0.06	0.09	0.07	0.66	1.22
19	INM-CM4-8	2.1	0.47	0.07	0.21	3.51	0.22	0.05	0.08	0.02	0.14	3.26	0
20	INM-CM5-0	3.58	0.33	0.02	0.06	3.18	0.25	0.02	0.02	0.05	0.12	0.58	0.22
21	IPSL-CM6A-LR	4.62	0.42	0.17	0.48	2.36	0.33	0	0.09	0.05	0	3.46	0.73
22	KACE-1-0-G	0.07	0.39	0.08	0.03	0.88	0.34	0.04	0.06	0.16	0.01	0.36	1
23	KIOST-ESM	4.8	0.67	0.13	0.05	2.45	0.24	0.55	0.33	0.55	0.14	0.37	0.91
24	MIR <sup>o</sup> C-ES2L	1.51	0.38	0.03	0.24	2.38	0.16	0.03	0.06	0.06	0.04	1.19	0.76
25	MIR <sup>o</sup> C6	2.88	0.82	0.12	0.1	2.44	0.31	0.43	0.26	0.69	0.11	3.88	0.14
26	MPI-ESM1-2-HAM	5.57	0.58	0	0.02	0.15	0.4	0.08	0.29	0.05	0.01	0.23	0.65
27	MPI-ESM1-2-HR	10.16	0.6	0.03	0.04	7.84	0.06	0.01	0.12	0.06	0.04	1.03	0.26
28	MPI-ESM1-2-LR	9.41	0.43	0.02	0.05	3.69	0.02	0.25	0.26	0.14	0.11	2.98	0.41
29	MRI-ESM2-0	1.81	0.26	0.03	0.03	0.69	0.2	0.12	0.23	0.01	0.11	1.95	0.18
30	NESM3	3.92	0.86	0.07	0.04	1.75	0.59	0.02	0.03	0.07	0.02	2.6	0.31
31	NorCPM1	0.09	0.34	0.14	0.37	2.29	0.39	0.06	0.42	0.15	0.02	1.97	0.9
32	NorESM2-LM	0.4	0.59	0.15	0.17	1.81	0.35	0.07	0.2	0.19	0.09	0.18	0.42
33	NorESM2-MM	2.06	1.35	0.24	0.28	0.79	0.04	0.12	0.1	0.01	0.07	0.07	0.13
34	SAM0-UNICON	6.38	0.5	0.23	0.38	0.07	0.06	0.19	0.2	0.09	0.07	1.39	1.36
35	TaiESM1	5.02	0.66	0.16	0.35	0.33	0.79	0.25	0.13	0.08	0.17	3.3	0.43
36	UKESM1-0-LL	2.85	0.6	0.11	0.11	1.82	0.3	0.16	0.14	0.03	0.11	0.88	0.47

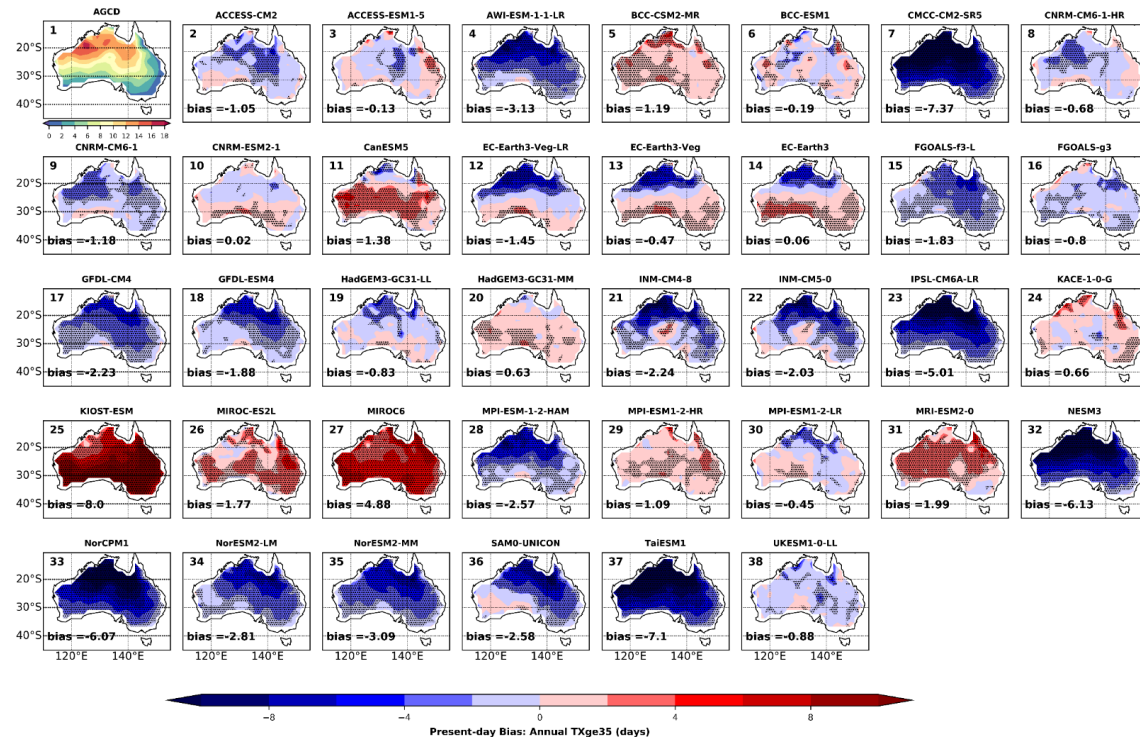
### 3.1.2. Temperature Extremes

Figures 6–12 show the observed annual climatological mean and biases in CMIP6 GCMs for TXx, TXge35, SU, TNN, TR, WSDI and CSDI, respectively. On average, most GCMs typically show cold biases for hot extremes and warm biases for cold extremes, except for WSDI and CSDI. However, the magnitude of bias for cold extremes is generally larger than that of hot extremes. There also exists an exception to this pattern. For example, CanESM5, KIOST-ESM, MIR<sup>o</sup>C-ES2L, MIR<sup>o</sup>C6 and MRO-ESM2-0 show strong hot bias for hot extremes, whereas GFDL-CM4, USESM1-0-LL and NorCPM1 show slight cold bias for cold extremes. In addition, GCMs which struggle to capture the extremes of maximum temperature also perform poorly in capturing extremes of minimum temperature.



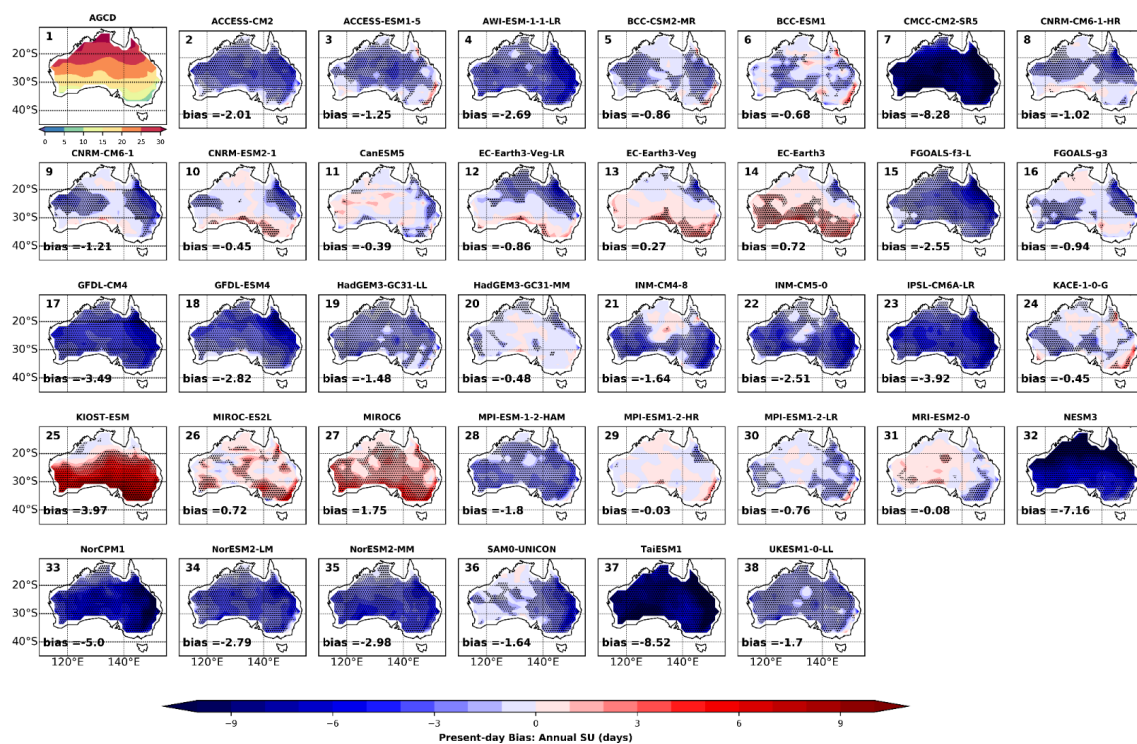


**Figure 6.** Climatological mean annual bias in maximum maximum temperature (TXx: °C) relative to the Australian Gridded Climate Data dataset (AGCD; panel 1) for the individual CMIP6 GCMs (panels 2–38). Data spans 1951–2014. Stippling indicates statistically significant differences using a student's *t*-test at the 95% confidence level.

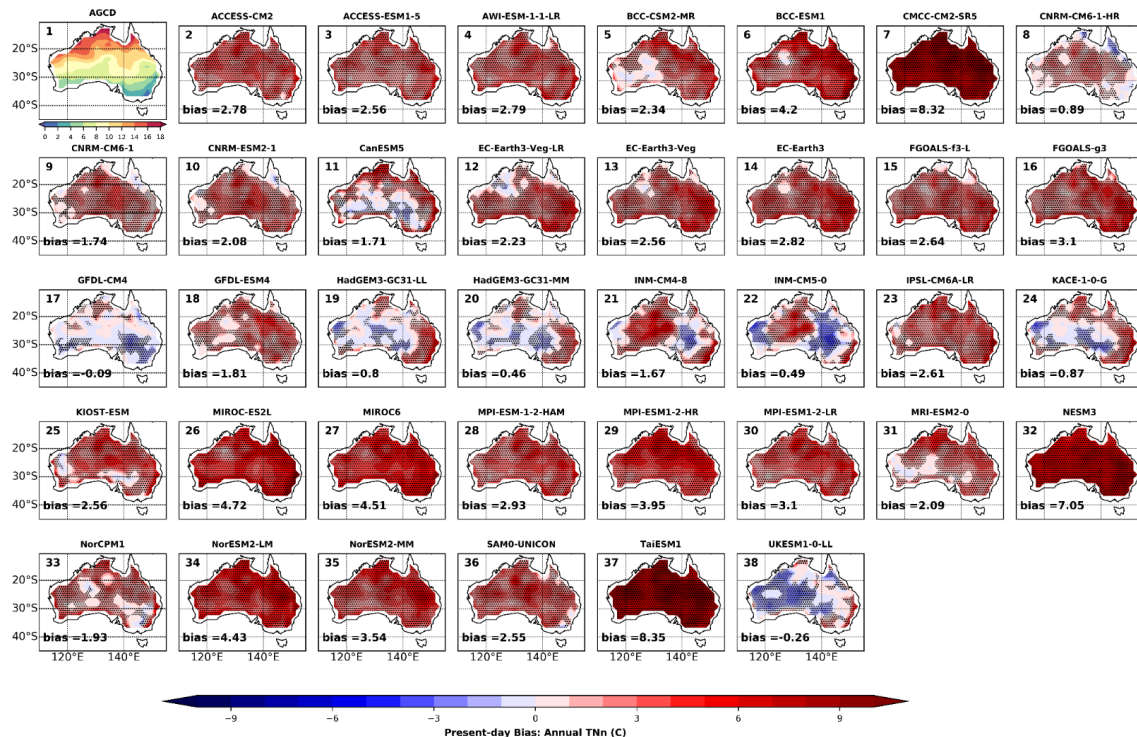


**Figure 7.** Climatological mean annual bias in the number of days when maximum temperature is greater than 35 C (Txge35: days) relative to the Australian Gridded Climate Data dataset (AGCD; panel 1) for the individual CMIP6 GCMs (panels 2–38). Data spans 1951–2014. Stippling indicates statistically significant differences using a student's *t*-test at the 95% confidence level.

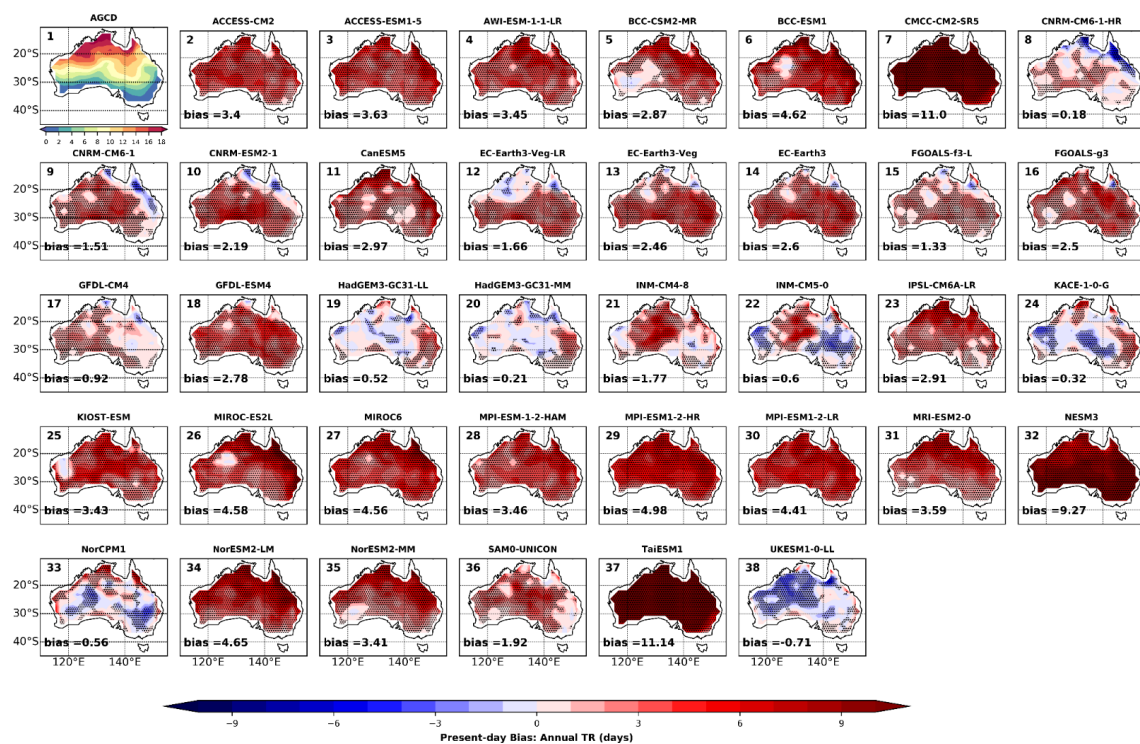




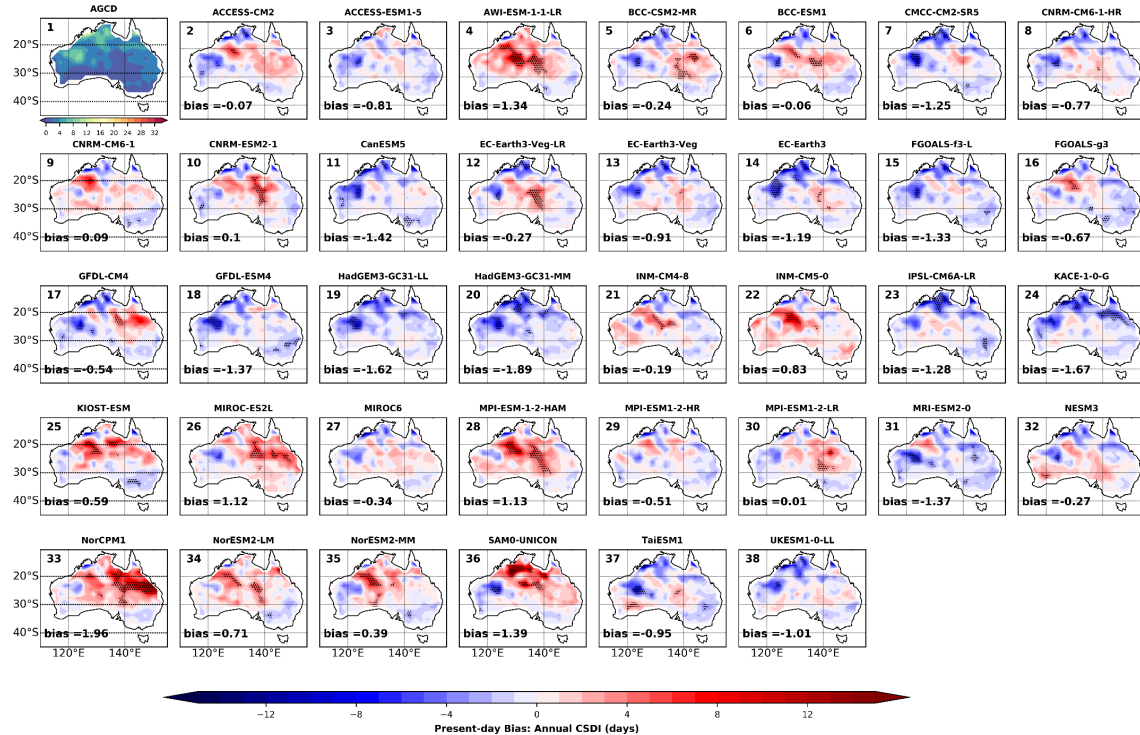
**Figure 8.** Climatological mean annual bias in summer days (number of days when maximum temperature  $> 25^{\circ}\text{C}$ ) (SU: days) relative to the Australian Gridded Climate Data dataset (AGCD; panel 1) for the individual CMIP6 GCMs (panels 2–38). Data spans 1951–2014. Stippling indicates statistically significant differences using a student's  $t$ -test at the 95% confidence level.



**Figure 9.** Climatological mean annual bias in the minimum minimum temperature (TNn:  $^{\circ}\text{C}$ ) relative to the Australian Gridded Climate Data dataset (AGCD; panel 1) for the individual CMIP6 GCMs (panels 2–38). Data spans 1951–2014. Stippling indicates statistically significant differences using a student's  $t$ -test at the 95% confidence level.

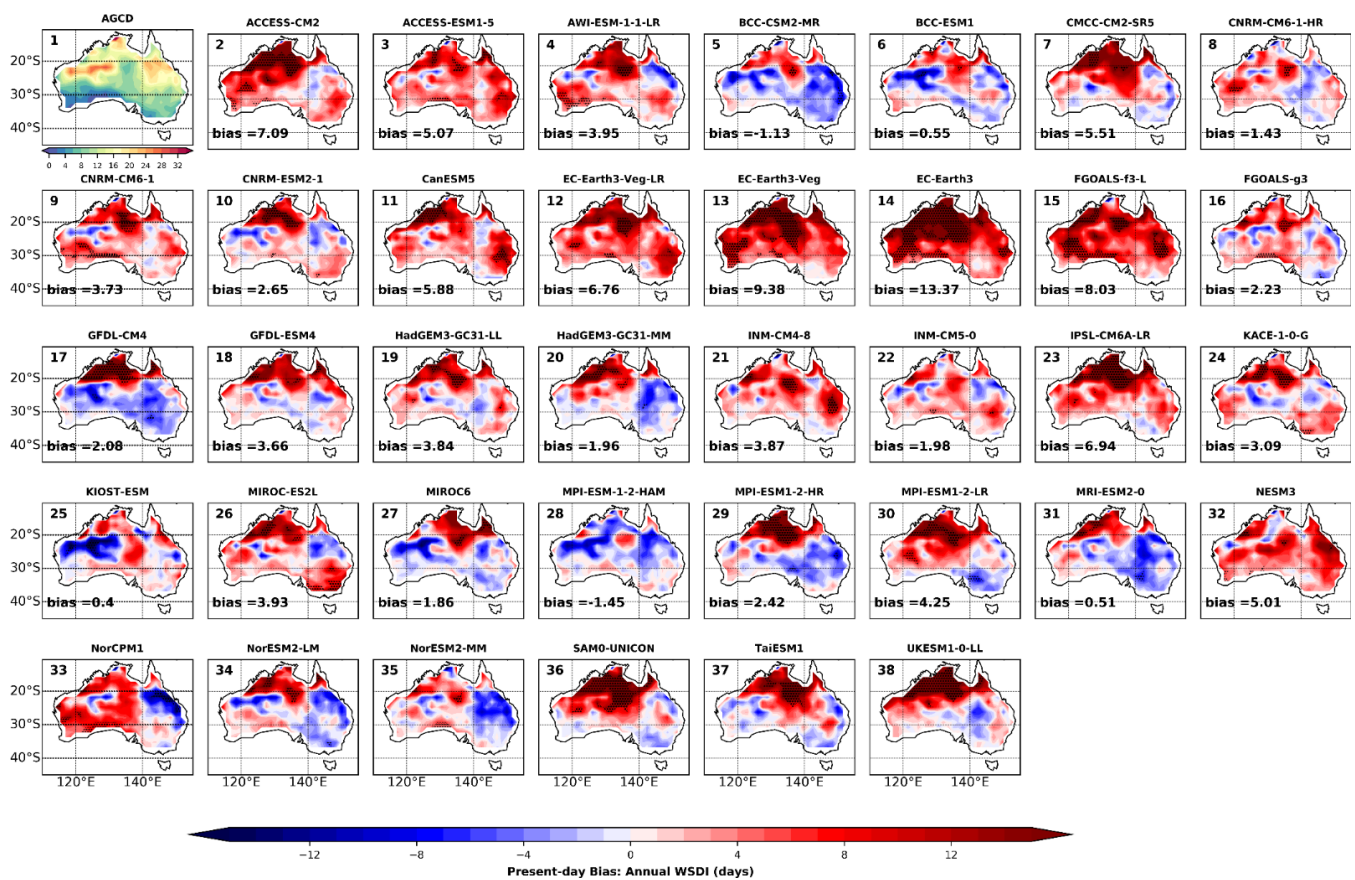


**Figure 10.** Climatological mean annual bias in tropical nights (number of days when minimum temperature > 20 °C (TR: days) relative to Australian Gridded Climate Data dataset (AGCD; panel 1), for the individual CMIP6 GCMs (panels 2–38). Data spans 1951–2014. Stippling indicates statistically significant differences using a student's *t*-test at the 95% confidence level.



**Figure 11.** Climatological mean annual bias in the annual count of nights with at least 4 consecutive nights when daily minimum temperature < 10th percentile (CSDI: days) relative to Australian Gridded Climate Data dataset (AGCD; panel 1), for the individual CMIP6 GCMs (panels 2–38). Data spans 1951–2014. Stippling indicates statistically significant differences using a student's *t*-test at the 95% confidence level.





**Figure 12.** Climatological mean annual bias in the annual count of days with at least 4 consecutive days when daily maximum temperature > 90th percentile (WSDI: days) relative to Australian Gridded Climate Data dataset (AGCD; panel 1), for the individual CMIP6 GCMs (panels 2–38). Data spans 1951–2014. Stippling indicates statistically significant differences using a student's *t*-test at the 95% confidence level.

In Figure 6, GCMs KIOST-ESM, MIR<sup>o</sup>C6, CanESM5 and MRI-ESM2-0 show the strongest warm biases ( $\sim 2$ – $8$  °C), whereas NorCPM1, NorESM2-LM and NorESM2-MM (Norwegian Climate Centre), TaiESM1, IPSL-CM6A-LR and CMCC-CM2-SR5 show the strongest cold bias ( $\sim -2$  to  $-10$  °C) throughout the continent. Previously, Alexander & Arblaster (2017) [11] evaluated CMIP5 GCMs and noted that GCMs from the MIR<sup>o</sup>C family show a strong warm bias over Australia. Our results are broadly consistent with these findings. Most GCMs typically show cold bias ranging from  $-1$  to  $-3$  °C. Txge35 (Figure 7) and SU (Figure 8) follow a similar pattern to TXx, and the GCMs with warm bias in TXx also overestimate TXge35 and SU, whereas GCMs with cold bias in TXx, underestimate TXge35 and SU.

Overall, for hot extremes (TXx, Txge35 and SU), BCC-ESM1, CNRM-ESM2-1, EC-Earth3-Veg, EC-Earth3, HadGEM3-GC31-MM, MPI-ESM1-2-HR and KACE-1-0-G are found to be top-performing GCMs whereas CMCC-CM2-SR5, IPSL-CM6A-LR, KIOST-ESM, MIR<sup>o</sup>C6, NESM3, NorCPM1, NorESM2-MM and TaiESM1 are found to be the worst-performing GCMs (Table 3).

For TNn (Figure 9), most GCMs typically show warm biases over the entire continent. The GCMs GFDL-CM4 and UKESM1-0-LL show slight cold biases in western and eastern Australia. However, the magnitude of these biases is small ( $\sim -0.2$ – $0.5$  °C). CanESM5, GFDL-CM4, KACE-1-0-G, UKESM1-0-LL, HadGEM3-GC31-LL, HadGEM3-GC31-MM (Hadley Centre), INM-CM4-8 and INM-CM5-0 (INM Russia) show the smallest bias. TaiESM1, CMCC-CM2-SR5 and NESM3 show the largest bias of  $\sim -7$  to  $-8.5$  °C (Table 3). The GCMs with large warm biases in TNn (TaiESM1, CMCC-CM2-SR5 and

NESM3) also show large overestimations in TR (Figure 10 ~10–12 days). For all other GCMs, overestimation in TR varies between 2–5 days.

The GCMs typically overestimate CSDI in northern and central Australia, whereas they underestimate cold spell durations over other parts of Australia, including the heavily populated east coast (Figure 11). However, these biases are not found to be significant. GFDL-ESM4, HadGEM3-1LL, HadGEM3-GC31-MM and KACE-1-0G show underestimation throughout the domain. In terms of bias magnitude, ACCESS-CM2, BCC-CSM2-MR and BCC-ESM1 (Beijing Climate Centre), CNRM-CM6-1, CNRM-ESM2-1, EC-Earth3-Veg-LR, INM-CM4-8, MPI-ESM1-2-LR and NESM3 show the smallest bias (<0.3 days), whereas NorCPM1, SAMO-UNICON, CanESM5, GFDL-ESM4, HadGEM3-GC31-LL and HadGEM3-GC31-MM (Hadley Centre), MRI-ESM2-0 and AWI-ESM-1-1LR show the largest biases (>1 day).

For WSDI (Figure 12), biases are much larger than CSDI. The large biases in WSDI compared to CSDI can be attributed to the observed WSDI (Figure 12, panel 1) typically being 4-times higher than CSDI (Figure 11, panel 1). Here, except for ACCESS-CM2 and ACCESS-ESM1-5 (Australian Community) and EC-Earth-Veg-LR, EC-Earth3-Veg and EC-Earth3 (EC-Earth consortium), all other GCMs overestimate warm spells in northern and central Australia and underestimate them in other parts of the domain. Positive biases are mostly significant, whereas negative biases are mostly not significant. The EC-Earth consortium GCMs show the largest bias (~ >8 days), followed by ACCESS-CM2, CanESM5, FGOALS-f3-L, IPSL-CM6A-LR, TaiESM1 and UKESM1-0-LL.

Similar to precipitation extremes, the biases at the seasonal scale for temperature extremes (TXx (Figures S17–S20), Txge35 (Figure S21–S24), SU (Figures S25–S28), TNn (Figures S29–S32) and TR (Figures S33–S36) mimicked the annual biases (Supplementary Materials), i.e., GCMs with cold and warm biases at annual scales also exhibit cold and warm seasonal biases, respectively, and these biases have approximately similar spatial distributions. For impact modellers and stakeholders using GCM data, bias corrections of extremes are recommended.

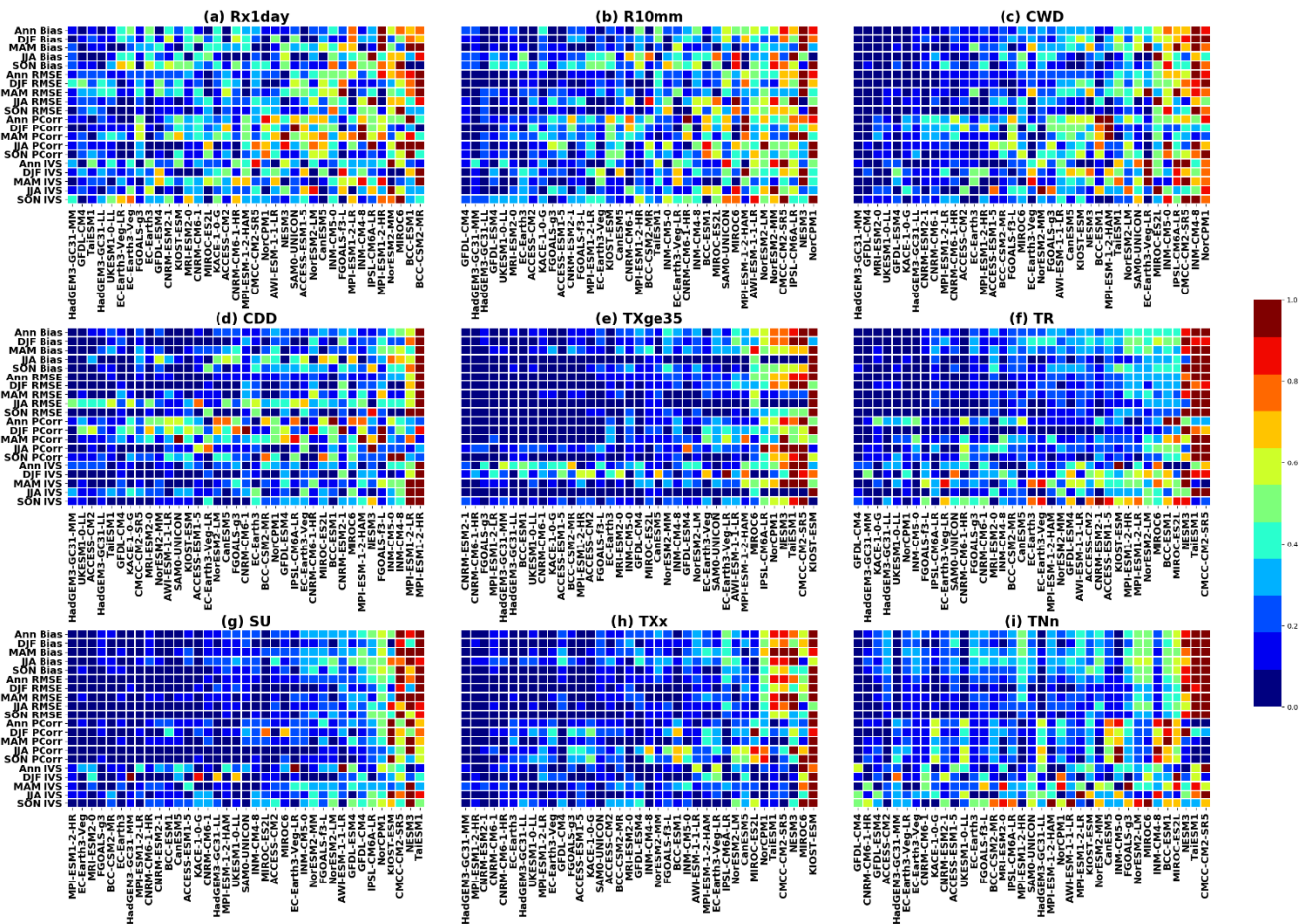
In terms of RMSE, for all the temperature extremes except WSDI and CSDI, a common pattern is seen, i.e., GCMs that show large biases also show large RMSEs and vice versa (Table 4). Hence, the evaluation of GCM performance remains robust across bias and RMSE for all temperature extremes except WSDI and CSDI. For WSDI, there seems to be some overlap in the best and worst performing GCMs as measured by bias and RMSE, with a few exceptions such as AWI-ESM-1-1-LR and TaiESM1 (which show large RMSEs and small bias). For CSDI, there is more variance in results, and the biases are found to be mostly not significant. Due to this, GCMs which are the worst performers for bias (GFDL-ESM4, HadGEM3-GC31-LL, HadGEM3-GC31-MM, CanESM5 and KACE-1-0-G) are found to be among the best performers for RMSE.

For PCorr (Table 5), the GCM performance ranking does not completely align with that of other metrics (bias, RMSE and IVS). For example, MPI-ESM-1-2-HAM, MPI-ESM-1-2-LR and MPI-ESM-1-2-HR (MPI Germany) and MRI-ESM2-0 are among the best performers for PCorr for more than three extremes. However, these GCMs are not in the best performing subset for other metrics (bias, RMSE and IVS). For IVS, GCMs show overlap in the best and worst performing subset for most of the extremes; for example, INM-CM4-8, INM-CM5-0, IPSL-CM6A-LR, KACE-1-0-G and CNRM-ESM2-1 are found to be the best performers whereas AWI-ESM-1-1-LR, CMCC-CM2-SR5, KIOST-ESM, MIR<sup>o</sup>C6, MPI-ESM1-2-LR are found to be the worst performers for more than three extremes. The GCMs ranking for IVS is thus roughly comparable with bias and RMSE.

### 3.2. Normalisation and Averaging

We calculate the continental means of biases and RMSEs, then normalise (based on Equations (3) and (4)) and average all the metrics (bias, RMSE, PCorr, IVS) for all indices. For the extreme indices available at the monthly timescale (Table 2), 20 normalised metrics (i.e., bias, RMSE, PCorr and IVS for Annual, DJF, MAM, JJA and SON, respectively) are

calculated (Figure 13). In contrast, for the extreme indices only available at the annual timescale (Table 2), four normalised metrics (i.e., bias, RMSE, PCorr and IVS for Annual) are generated (Figure 14).

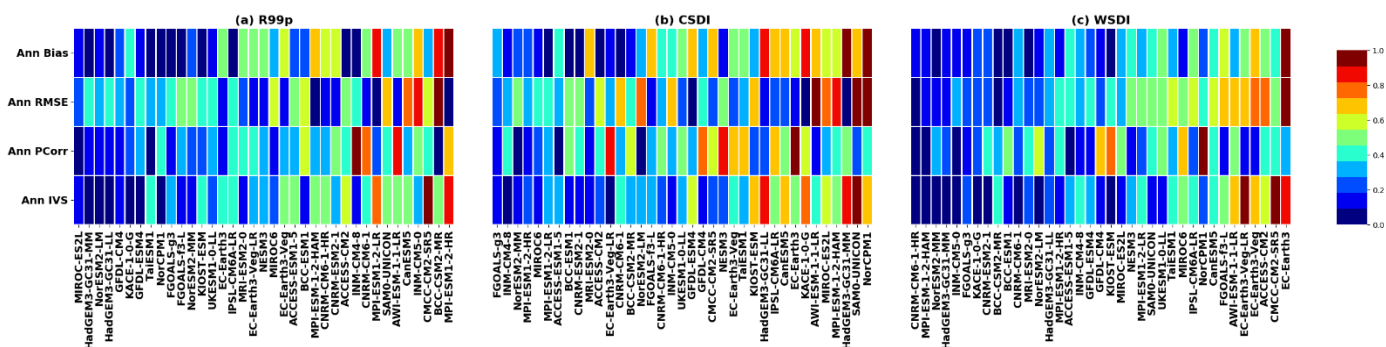


**Figure 13.** Normalised and continentally averaged annual and seasonal means of individual metrics (bias, root mean square error (RMSE), pattern correlation (PCorr) and interannual variability score (IVS)) for the extreme indices (a) maximum 1-day precipitation (Rx1Day), (b) number of very heavy rain days (rain > 10 mm) (R10mm), (c) consecutive wet days (CWD), (d) consecutive dry days (CDD), (e) number of days when maximum temperature is greater than 35 °C (Txge35), (f) tropical nights (number of days when minimum temperature > 20 °C) (TR), (g) summer days (number of days when maximum temperature > 25 °C) (SU), (h) maximum maximum-temperature (TXx) and (i) minimum minimum-temperature (TNn). Here, the smaller values of normalised metrics correspond to the better performance of GCMs. Moreover, for each extreme index, GCMs are arranged from best to worst performance as we move from left and right. Data spans 1951–2014.

Overall, for both precipitation and temperature extremes (Figures 13 and 14), we find that the top performing GCMs consistently score well (<0.2) in most metrics, capturing the magnitude, spatial and temporal patterns of the observed field. In contrast, the poorly performing GCMs are less skilled at capturing spatial and temporal patterns. For temperature extremes (except for CSDI and WSDI), approximately 5–7 GCMs perform poorly (normalised mean score of all metrics > 0.6) (Figure 13e–i). For precipitation extremes, 15–17 GCMs perform poorly (normalised score > 0.6) at capturing the spatial and temporal pattern of the observed field (Figure 13a–d). These results suggest that GCMs show higher overall skill in capturing observed temperature extremes. In contrast, many GCMs struggle to replicate observed precipitation extremes, with only a small subset of high-skill GCMs.



These results reflect the findings of previous studies [11,21,22], which have suggested that GCMs struggle to capture precipitation and its extremes.



**Figure 14.** Normalised and continentally averaged annual means of individual metrics (bias, root mean square error (RMSE), pattern correlation (PCorr) and interannual variability score (IVS)) for the extreme indices (a) total annual precipitation from very heavy rain days (R99p), (b) annual count of nights with at least 4 consecutive nights when daily minimum temperature < 10th percentile (CSDI) and (c) annual count of days with at least 4 consecutive days when daily maximum temperature > 90th percentile (WSDI). Here, the smaller values of normalised metrics correspond to the better performance of GCMs. Moreover, for each extreme index, GCMs are arranged from best to worst performance as we move from left and right. Data spans 1951–2014.

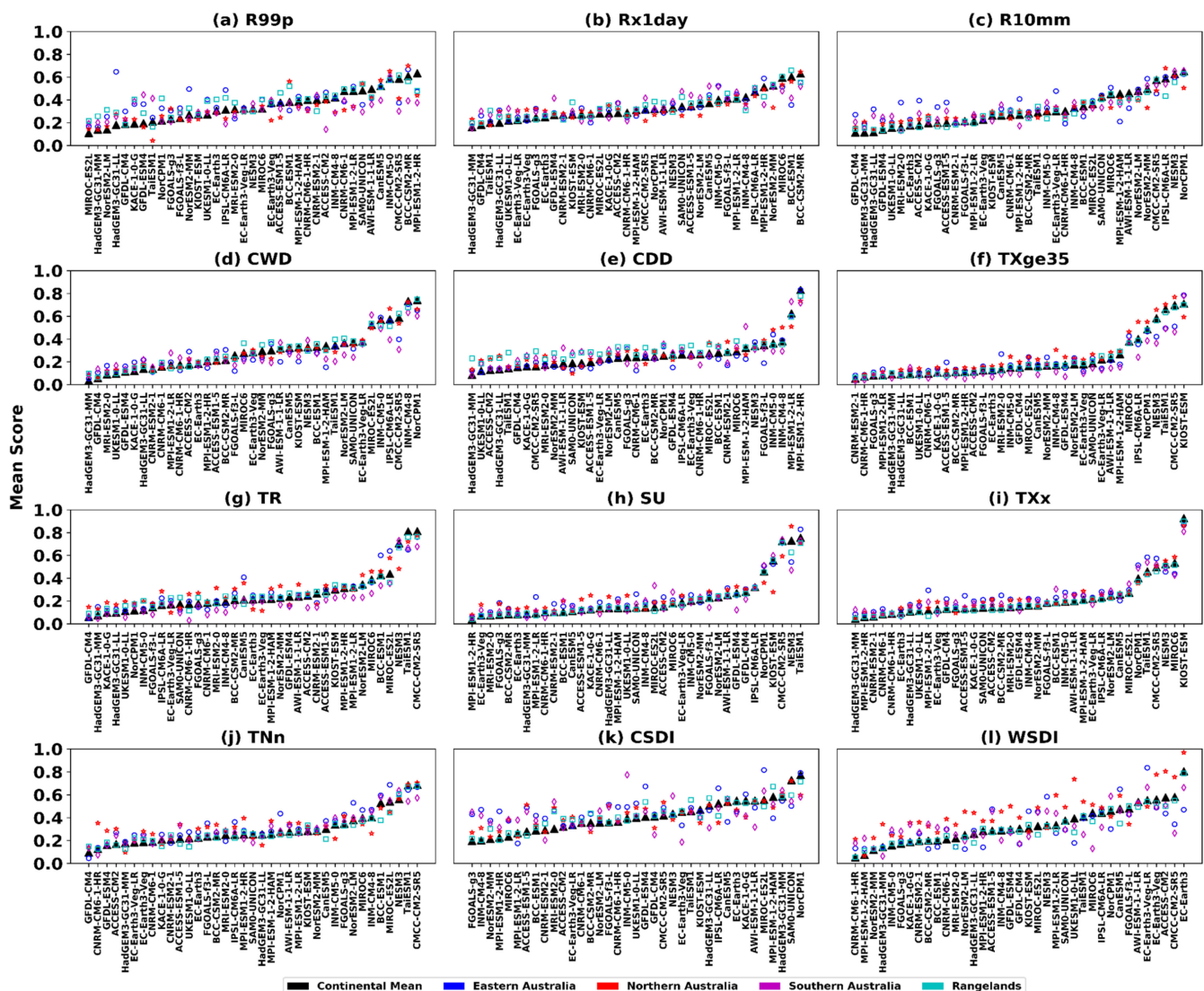
Based on these results, we find that HadGEM3-GC31-MM, HadGEM3-GC31-LL (Hadley Centre), GFDL-CM4, GFDL-ESM4 (NOAA), UKESM1-0-LL, KACE-1-0-G, and MRI-ESM2-0 tend to show consistently good performance for more than 50% of the extreme precipitation indices analysed in this study. Whereas GCMs INM-CM4-8, INM-CM5-0 (INM Russia), MPI-ESM1-2-LR, MPI-ESM1-2-HR (MPI Germany), MIR<sup>o</sup>C6, NorCPM1, BCC-ESM1 and IPSL-CM6A-LR tend to show consistently poor performance for more than 50% of the extreme precipitation indices.

For Australian temperature extremes, consistently good performing GCMs for more than 50% of the extreme indices are: BCC-ESM1, HadGEM3-GC31-MM, CNRM-CM6-1-HR, CNRM-ESM2-1, FGOALS-g3, MPI-ESM1-2-LR, GFDL-ESM4 and GFDL-CM4 (NOAA), and KACE-1-0G whereas the consistently poor performing GCMs are: KIOST-ES, MIR<sup>o</sup>C6, CMCC-CM2-SR5, TaiESM1, MIR<sup>o</sup>C-ES2L, NorCPM1, and NESM3.

Considering both precipitation and temperature extremes, HadGEM3-GC31-MM, HadGEM3-GC31-LL (Hadley Centre), UKESM1-0-LL, FGOALS-g3, GFDL-ESM4 and GFDL-CM4 (NOAA) show consistently good performance. Conversely, CMCC-CM2-SR5, MIR<sup>o</sup>C6, NESM3 and NorCPM1 consistently perform poorly in replicating Australian temperature and precipitation extremes.

We also compare the overall GCM performance ranking at the continental scale with performance calculated for the four Natural Resource Management (NRM) regions (Figure S37) to assess if GCM performance varies for different regions (Figure 15). The four NRM regions are named southern Australia (south), eastern Australia (east), northern Australia (north), and Rangelands covering the central area (centre). These regions are broadly aligned with climatological boundaries (CSIRO and Bureau of Meteorology 2015 [65]; Fiddes et al. 2021 [66]) and align with the IPCC reference regions [67]. Examining the performance of GCMs over these regions is particularly important when considering the design of regional downscaling experiments or assessing regional scale climate impacts. For example, regional climate projections like New South Wales (NSW) and Australian Regional Climate Modelling (NARClIM) are made by choosing a subset of best-performing GCMs over a smaller sub-domain which falls within eastern and southern Australia NRM regions [68,69]. This analysis will be of strong interest for producing updated national

climate change projections for Australia and providing new insights into the climate system and climate change relevant to the region.



**Figure 15.** Spatially averaged normalised scores of the GCMs for the extreme indices (a) total annual precipitation from very heavy rain days (R99p), (b) maximum 1-day precipitation (Rx1Day), (c) number of very heavy rain days (rain > 10 mm) (R10mm), (d) consecutive wet days (CWD), (e) consecutive dry days (CDD), (f) number of days when maximum temperature is greater than 35 °C (Txge35), (g) tropical nights (number of days when minimum temperature > 20 °C) (TR), (h) summer days (number of days when maximum temperature > 25 °C) (SU), (i) maximum maximum-temperature (TXx), (j) minimum minimum-temperature (TNn), (k) annual count of nights with at least 4 consecutive nights when daily minimum temperature < 10th percentile (CSDI) and (l) annual count of days with at least 4 consecutive days when daily maximum temperature > 90th percentile (WSDI). Here, the smaller score values correspond to the better performance of GCMs. Also, for each extreme index, GCMs are arranged from best to worst performance as we move from left and right. Here black, blue, red, magenta and cyan colours denote means over the continent, eastern Australia, northern Australia, southern Australia and rangelands. The markers denote the continental (triangle), northern Australia (star), eastern Australia (circle), southern Australia (diamond) and rangelands (square) values. Data spans 1951–2014.

The result reveals that for all the extreme precipitation indices (except R99p), GCM performance does not show much regional variance and performance rankings remain

the same for the continent and the four sub-regions (Figure S37). In contrast, for R99p (Figure 15a), some GCMs show notable inter-regional differences. For example, HadGEM3-GC31-LL is one of the worst performers in eastern Australia whereas it is among the best for continental scale and other regions. A similar pattern is also seen for ACCESS-CM2. This GCM is one of the best performers for southern Australia; however, it remains in the mid-range performance on a continental scale and for other regions.

For all the temperature extremes except WSDI and CSDI (Figure 15k,l), GCM performance at the continental and regional scale show little variance. However, for WSDI and CSDI, strong inter-regional differences are observed for eastern, northern, and southern Australia.

Previous studies have also shown that for different climate variables, GCM performance varies in different regions [70,71]. This is related to how GCMs represent the land cover, terrain morphology, and climate zones which are crucial for the simulation of particular climate variables.

#### 4. Discussion and Conclusions

Di Virgilio et al. 2022 [33] evaluated CMIP6 GCMs for downscaling over Australia by assessing individual GCMs against criteria like performance simulating daily climate variable distributions, climate means, extremes, and modes; model independence; and climate change signal diversity. Grose et al. 2020 [32] evaluated how the CMIP6 ensemble performs in comparison to CMIP5 for present and future climates. A recent study by Zhang et al. 2022 [72] evaluated the performance of CMIP6 GCMs over the Australia CORDEX domain by examining the climatological mean and interannual variability. However, none of these studies comprehensively examined the climate extremes in Australia. We, in this study, present the detailed evaluation of 12 extreme precipitation and temperature indices as simulated by 37 CMIP6 GCMs. Although this study did not evaluate all possible aspects of GCM performance, it aimed to select some key evaluation metrics fundamental to characterising GCM performance for simulating climate extremes.

Our results show that overall, CMIP6 GCMs underestimate precipitation intensity and overestimate wet days, overestimate minimum temperature extremes, and underestimate maximum temperature extremes. We also find that a greater number of GCMs show high skill in capturing the observed temperature extremes, whereas GCMs typically show low skill for precipitation extremes. Past studies have suggested that GCMs typically struggle to capture the observed changes in precipitation extremes due to multiple factors [73]. For example, the inability of GCMs to simulate internal variability in heavy regional precipitation, statistical effects of different grid resolutions of GCMs and observations, and the inability to resolve convective processes due to convective parametrisations used by the GCMs [74].

We find a subset of GCMs (HadGEM3-GC31-MM and HadGEM3-GC31-LL (Hadley Centre), UKESM1-0-LL, FGOALS-g3, and GFDL-ESM4 and GFDL-CM4 (NOAA)) that tend to show consistently good performance for more than 50% of climate extremes analysed in this study and across a broad suite of evaluation metrics. Our results are comparable to the recent study by Srivastava et al. (2020) [43] and Ayugi et al. (2021) [26], who evaluated the performance of CMIP6 GCMs for simulating precipitation extremes over the United States and East Africa, respectively. Both studies showed that HadGEM3-GC31-LL and UKESM1-0-LL consistently show good performance for the United States and East Africa.

We also find that a subset of GCMs (CMCC-CM2-SR5, MIR<sup>2</sup>C6, NESM3 and NorCPM1) tend to perform poorly overall. It is well known that some aspects of climate are poorly represented in all GCMs due to common limitations in model parametrisations and spatial resolution. For example, Chen et al. (2021) [45] showed that the convective parametrisation in most CMIP5 GCMs typically leads to a slight drizzling bias. In contrast, Alexander and Arblaster (2017) [11] showed that CMIP5 GCMs tend to slightly overestimate minimum temperature extremes and underestimate maximum temperature extremes. However, we find that the poor performing GCMs tend to show unrealistic errors compared to

common errors found in most GCMs. For example, CMCC-CM2-SR5, MIR<sup>o</sup>C6, NESM3 and NorCPM1 were found to perform extremely poorly for most of the extreme indices analysed in this study. The magnitude of biases shown in these GCMs was exceptionally high compared to the median bias across the 37 GCMs. Past study has shown that GCMs showing unrealistic errors are often those for which projections lie on the margins of or outside the range of most of the ensemble [75].

The paper's results also highlighted that resolution does not appear to be a deciding factor in how well a model performs at reproducing temperature/precipitation indices for Australia. For example, the relatively high-resolution CNRM-CM6-1 does not necessarily 'outperform' lower resolution models such as BCC-ESM1. The key factor that is more closely tied to the model performance is the 'model family'. GCMs from the same parent institution and GCMs from different institutions that share key model components either do a good or bad job in capturing the broad-scale features and processes of the climate, resulting in similar model performance. For example, in GCMs from the same parent institution (GCMs from MPI Germany, INM Russia, Beijing Climate Centre etc.), if one model version is too hot or too wet, then this is apparent in all the model versions. GCMs from the Australian community, UK Met Office Hadley Centre, NOAA-GFDL and UKESM1-0-LL, which share atmospheric model codes, also indicate similar results. Past studies have also reported a similar impact of the 'model family' on the GCM performance of CMIP3 [10] and CMIP5 models [11]. The large dependence of different versions of models on the model 'family' highlights an important issue regarding model dependence and whether we are using false assumptions that these multiple simulations all constitute independent samples (e.g., Abramowitz 2010 [76]; Abramowitz and Bishop 2015 [77]).

Although this study provides a comprehensive evaluation of CMIP6 GCMs, it also has some caveats. This study assessed the performance of CMIP6 GCMs in simulating climate extremes using a metric-based approach. It should be recognised that models are complex in nature, and summarising the performance of models using a few metrics is quite challenging and may be inadequate. In addition to the limitations in performance metrics, the choice of reference dataset also affects the performance of models. Past studies have shown that the reference dataset's choice can bias the model performance results [22,78,79].

In this study, we used the Rupp et al. (2013) [54] normalisation method, which normalised the GCMs based on the maximum and minimum value of the metric among the GCMs. Normalisation is performed such that different metrics can be fairly compared. This method assumes the spread of each metric for all the GCMs is relatively equal. This assumption is not necessarily true; thus, this method can bias the results, as it is possible to subdivide the GCMs and manipulate rankings by changing the relative weights between the different metrics. For example, if a single GCM performs extraordinarily poorly in only one metric, the inclusion of that GCM greatly reduces the importance of that metric as the normalisation factor becomes too small because of the range being extremely large.

Despite these limitations, our study has important implications for multiple systems affected by climate extremes. For instance, our evaluation shows that precipitation extremes typically record the largest bias in Australia's northern and eastern parts. The northern part of Australia is the region belonging to the northern monsoon cluster. This region receives widespread rainfall from the large oceanic mesoscale convective systems [80]. Past studies have shown that GCMs typically struggle to simulate the characteristic of mesoscale convective systems with reasonable accuracy and thus record large errors in monsoon-dominated regions [70,81–84]. The eastern part of Australia is a region of complex topography composed of mountains, alpine, and lakes [85]. Past studies have shown that the topography of GCMs is generally smoother and lower in elevation than that in reality [86]. For this reason, GCMs typically record large biases over these regions [87]. Northern Australia is a major region for agricultural production, whereas eastern coastlines are where more than 50% of the Australian population resides. The large uncertainty in GCM performances over these economically and societally important regions reveals



the limitations of GCMs for their application in studying impacts from precipitation and temperature extremes over these areas.

The results of this study are useful for identifying potential CMIP6 GCMs for performing exercises like dynamical downscaling experiments as part of endeavours such as CORDEX Australasia. Regional climate modelling helps provide information at fine, sub-GCM grid scales, which is more suitable for studies of regional phenomena and application to vulnerability, climate impacts, and adaptation assessments. However, we acknowledge that for selecting potential GCMs for dynamical downscaling, additional and more comprehensive evaluations like the GCM performance for simulating mean climate, and climate drivers like El Niño, Southern Annular mode, etc., should be undertaken to rank the overall GCM performance. An in-depth analysis of physical processes is also required to identify the cause of the model bias. Another crucial factor is selecting GCMs based on the sign of GCM future climate change signals.

Modelling factors like choice of model physics and land-atmosphere coupling in the GCMs can directly affect precipitation and temperature extremes. Analysing and attributing the GCMs with respect to these factors was beyond the scope of this study. However, it will be crucial to investigate further to identify what modelling factors contribute to the high and poor skill in GCMs for simulating different extreme indices.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/atmos13091478/s1>, Figure S1. Climatological mean December-January-February bias in maximum 1-day precipitation (Rx1Day: mm) relative to Australian Gridded Climate Data dataset (AGCD; panel 1), for the individual CMIP6 GCMs (panels 2–38). Data spans 1951–2014. Stippling indicates statistically significant differences using a student's *t*-test at the 95% confidence level. The white mask in part of the inland domain are the regions with no station data. Figure S2. Same as S1 but for March-April-May. Figure S3. Same as S1 but for June-July-August. Figure S4. Same as S1 but for September-October-November. Figure S5. Climatological mean December-January-February bias in number of very heavy rain days (rain > 10 mm) (R10mm: days) relative to Australian Gridded Climate Data dataset (AGCD; panel 1), for the individual CMIP6 GCMs (panels 2–38). Data spans 1951–2014. Stippling indicates statistically significant differences using a student's *t*-test at the 95% confidence level. The white mask in part of the inland domain are the regions with no station data. Figure S6. Same as S5 but for March-April-May. Figure S7. Same as S5 but for June-July-August. Figure S8. Same as S5 but for September-October-November. Figure S9. Climatological mean December-January-February bias in consecutive wet days (CWD: days) relative to Australian Gridded Climate Data dataset (AGCD; panel 1), for the individual CMIP6 GCMs (panels 2–38). Data spans 1951–2014. Stippling indicates statistically significant differences using a student's *t*-test at the 95% confidence level. The white mask in part of the inland domain are the regions with no station data. Figure S10. Same as S9 but for March-April-May. Figure S11. Same as S9 but for June-July-August. Figure S12. Same as S9 but for September-October-November. Figure S13. Climatological mean December-January-February bias in consecutive dry days (CDD: days) relative to Australian Gridded Climate Data dataset (AGCD; panel 1), for the individual CMIP6 GCMs (panels 2–38). Data spans 1951–2014. Stippling indicates statistically significant differences using a student's *t*-test at the 95% confidence level. The white mask in part of the inland domain are the regions with no station data. Figure S14. Same as S13 but for March-April-May. Figure S15. Same as S13 but for June-July-August. Figure S16. Same as S13 but for September-October-November. Figure S17. Climatological mean December-January-February bias in maximum maximum-temperature (TXx: °C) relative to Australian Gridded Climate Data dataset (AGCD; panel 1), for the individual CMIP6 GCMs (panels 2–38). Data spans 1951–2014. Stippling indicates statistically significant differences using a student's *t*-test at the 95% confidence level. The white mask in part of the inland domain are the regions with no station data. Figure S18. Same as S17 but for March-April-May. Figure S19. Same as S17 but for June-July-August. Figure S20. Same as S17 but for September-October-November. Figure S21. Climatological mean December-January-February bias in number of days when maximum temperature is greater than 35 °C (Txge35: days) relative to Australian Gridded Climate Data dataset (AGCD; panel 1), for the individual CMIP6 GCMs (panels 2–38). Data spans 1951–2014. Stippling indicates statistically significant differences using a student's *t*-test at the 95% confidence level. The white mask in part of the inland domain are the regions with no station data. Figure S22. Same as S21



but for March–April–May. Figure S23. Same as S21 but for June–July–August. Figure S24. Same as S21 but for September–October–November. Figure S25. Climatological mean December–January–February bias in summer days (number of days when maximum temperature  $> 25^{\circ}\text{C}$ ) (SU: days) relative to Australian Gridded Climate Data dataset (AGCD; panel 1), for the individual CMIP6 GCMs (panels 2–38). Data spans 1951–2014. Stippling indicates statistically significant differences using a student’s  $t$ -test at the 95% confidence level. The white mask in part of the inland domain are the regions with no station data. Figure S26. Same as S25 but for March–April–May. Figure S27. Same as S25 but for June–July–August. Figure S28. Same as S25 but for September–October–November. Figure S29. Climatological mean December–January–February bias in minimum minimum-temperature (TNn:  $^{\circ}\text{C}$ ) relative to Australian Gridded Climate Data dataset (AGCD; panel 1), for the individual CMIP6 GCMs (panels 2–38). Data spans 1951–2014. Stippling indicates statistically significant differences using a student’s  $t$ -test at the 95% confidence level. The white mask in part of the inland domain are the regions with no station data. Figure S30. Same as S29 but for March–April–May. Figure S31. Same as S29 but for June–July–August. Figure S32. Same as S29 but for September–October–November. Figure S33. Climatological mean December–January–February bias in tropical nights (number of days when minimum temperature  $> 20^{\circ}\text{C}$ ) (TR: days) relative to Australian Gridded Climate Data dataset (AGCD; panel 1), for the individual CMIP6 GCMs (panels 2–38). Data spans 1951–2014. Stippling indicates statistically significant differences using a student’s  $t$ -test at the 95% confidence level. The white mask in part of the inland domain are the regions with no station data. Figure S34. Same as S33 but for March–April–May. Figure S35. Same as S33 but for June–July–August. Figure S36. Same as S33 but for September–October–November. Figure S37. (Figure 3 from Grose et al. 2020). Here red lines indicate the borders of the four “supercluster” averaging regions: Northern Australia (North), Rangelands (Centre), Southern Australia (South), Eastern Australia (East). Given that there is a substantial distinction between the coastal (temperate) East, versus the semi-arid inland East.

**Author Contributions:** N.N. conducted all the analyses and led the writing of the paper. G.D.V., F.J., E.T., K.B. and M.L.R. assisted in the interpretation and writing of the paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by NSW Climate Change Fund.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** CMIP6 data supporting this study’s findings are freely available from ESGF (<https://esgf-node.llnl.gov/search/cmip6/>, accessed on 1 November 2020). Details about AGCD are available at the Australian Bureau of Meteorology website (<http://www.bom.gov.au/metadata/catalogue/19115/ANZCW0503900567>, accessed on 1 November 2020). The dataset is available on the NCI (National Computational Infrastructure) server in project zv2. Detail on how to access the data can be found at <http://climate-cms.wikis.unsw.edu.au/AGCD> (accessed on 1 November 2020).

**Acknowledgments:** This work is made possible by funding from the NSW Climate Change Fund for NSW and the Australian Regional Climate Modelling (NARCLIM) Project. We thank the climate modelling groups for producing and making available their model outputs, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP3, CMIP5 and the Earth System Grid Federation (ESGF).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Masson-Delmotte, V.; Zhai, P.; Pirani, A.; Connors, S.L.; Péan, C.; Berger, S.; Zhou, B. Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, 2. 2021. Available online: <https://www.ipcc.ch/report/ar6/wg1/> (accessed on 1 November 2021).
2. Kiem, A.S.; Johnson, F.; Westra, S.; van Dijk, A.; Evans, J.P.; O'Donnell, A.; Mehrotra, R. Natural hazards in Australia: Droughts. *Clim. Chang.* **2016**, *139*, 37–54. [CrossRef]
3. Walsh, K.; White, C.J.; McInnes, K.; Holmes, J.; Schuster, S.; Richter, H.; Warren, R.A. Natural hazards in Australia: Storms, wind and hail. *Clim. Chang.* **2016**, *139*, 55–67. [CrossRef]
4. Johnson, F.; Hutchinson, M.F.; The, C.; Beesley, C.; Green, J. Topographic relationships for design rainfalls over Australia. *J. Hydrol.* **2016**, *533*, 439–451. [CrossRef]

5. Perkins-Kirkpatrick, S.E.; White, C.J.; Alexander, L.V.; Argüeso, D.; Boschat, G.; Cowan, T.; Purich, A. Natural hazards in Australia: Heatwaves. *Clim. Chang.* **2016**, *139*, 101–114. [[CrossRef](#)]
6. Zander, K.K.; Botzen, W.J.; Oppermann, E.; Kjellstrom, T.; Garnett, S.T. Heat stress causes substantial labour productivity loss in Australia. *Nat. Clim. Chang.* **2015**, *5*, 647–651. [[CrossRef](#)]
7. Westra, S.; White, C.J.; Kiem, A.S. Introduction to the special issue: Historical and projected climatic changes to Australian natural hazards. *Clim. Chang.* **2016**, *139*, 1–26. [[CrossRef](#)]
8. Plummer, N.; Salinger, M.J.; Nicholls, N.; Suppiah, R.; Hennessy, K.J.; Leighton, R.M.; Lough, J.M. Changes in climate extremes over the Australian region and New Zealand during the twentieth century. *Weather. Clim. Extrem.* **1999**, *42*, 183–202.
9. Haylock, M.; Nicholls, N. Trends in extreme rainfall indices for an updated high-quality data set for Australia, 1910–1998. *Int. J. Climatol. A J. R. Meteorol. Soc.* **2000**, *20*, 1533–1541. [[CrossRef](#)]
10. Alexander, L.V.; Arblaster, J.M. Assessing trends in observed and modelled climate extremes over Australia in relation to future projections. *Int. J. Climatol. A J. R. Meteorol. Soc.* **2009**, *29*, 417–435. [[CrossRef](#)]
11. Alexander, L.V.; Arblaster, J.M. Historical and projected trends in temperature and precipitation extremes in Australia in observations and CMIP5. *Weather Clim. Extrem.* **2017**, *15*, 34–56. [[CrossRef](#)]
12. Min, S.K.; Cai, W.; Whetton, P. Influence of climate variability on seasonal extremes over Australia. *J. Geophys. Res. Atmos.* **2013**, *118*, 643–654. [[CrossRef](#)]
13. Nishant, N.; Sherwood, S.C. How strongly are mean and extreme precipitation coupled? *Geophys. Res. Lett.* **2021**, *48*, e2020GL092075. [[CrossRef](#)]
14. Herold, N.; Ekström, M.; Kala, J.; Goldie, J.; Evans, J.P. Australian climate extremes in the 21st century according to a regional climate model ensemble: Implications for health and agriculture. *Weather Clim. Extrem.* **2018**, *20*, 54–68. [[CrossRef](#)]
15. King, A.D.; Karoly, D.J.; Henley, B.J. Australian climate extremes at 1.5 C and 2 C of global warming. *Nat. Clim. Chang.* **2017**, *7*, 412–416. [[CrossRef](#)]
16. King, A.D.; Black, M.T.; Min, S.K.; Fischer, E.M.; Mitchell, D.M.; Harrington, L.J.; Perkins-Kirkpatrick, S.E. Emergence of heat extremes attributable to anthropogenic influences. *Geophys. Res. Lett.* **2016**, *43*, 3438–3443. [[CrossRef](#)]
17. Bradstock, R.A.; Hammill, K.A.; Collins, L.; Price, O. Effects of weather, fuel and terrain on fire severity in topographically diverse landscapes of south-eastern Australia. *Landsc. Ecol.* **2010**, *25*, 607–619. [[CrossRef](#)]
18. Zheng, B.; Chenu, K.; Fernanda Dreccer, M.; Chapman, S.C. Breeding for the future: What are the potential impacts of future frost and heat events on sowing and flowering time requirements for Australian bread wheat (*Triticum aestivum*) varieties? *Glob. Chang. Biol.* **2012**, *18*, 2899–2914. [[CrossRef](#)]
19. Yilmaz, A.G.; Hossain, I.; Perera, B.J.C. Effect of climate change and variability on extreme rainfall intensity–frequency–duration relationships: A case study of Melbourne. *Hydrol. Earth Syst. Sci.* **2014**, *18*, 4065–4076. [[CrossRef](#)]
20. Nairn, J.R.; Fawcett, R.J. The excess heat factor: A metric for heatwave intensity and its use in classifying heatwave severity. *Int. J. Environ. Res. Public Health* **2015**, *12*, 227–253. [[CrossRef](#)]
21. Seneviratne, S.; Nicholls, N.; Easterling, D.; Goodess, C.; Kanae, S.; Kossin, J.; Zwiers, F.W. Changes in climate extremes and their impacts on the natural physical environment. In *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*; A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC); Field, C.B., Barros, V.T.F., Stocker, D., Qin, D.J., Dokken, K.L., Ebi, M.D., Mastrandrea, K.J., Mach, G.-K., Plattner, S.K., Allen, M., et al., Eds.; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2012; pp. 109–230.
22. Sillmann, J.; Kharin, V.V.; Zhang, X.; Zwiers, F.W.; Bronaugh, D. Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *J. Geophys. Res. Atmos.* **2013**, *118*, 1716–1733. [[CrossRef](#)]
23. Ongoma, V.; Chen, H.; Gao, C.; Nyongesa, A.M.; Polong, F. Future changes in climate extremes over Equatorial East Africa based on CMIP5 multimodel ensemble. *Nat. Hazards* **2018**, *90*, 901–920. [[CrossRef](#)]
24. Xu, K.; Wu, C.; Hu, B.X. Projected changes of temperature extremes over nine major basins in China based on the CMIP5 multimodel ensembles. *Stoch. Environ. Res. Risk Assess.* **2019**, *33*, 321–339. [[CrossRef](#)]
25. Wu, C.; Yeh, P.J.F.; Chen, Y.Y.; Hu, B.X.; Huang, G. Future precipitation-driven meteorological drought changes in the CMIP5 multimodel ensembles under 1.5 °C and 2 °C global warming. *J. Hydrometeorol.* **2020**, *21*, 2177–2196. [[CrossRef](#)]
26. Ayugi, B.; Jiang, Z.; Zhu, H.; Ngoma, H.; Babaousmail, H.; Karim, R.; Dike, V. Comparison of CMIP6 and CMIP5 models in simulating mean and extreme precipitation over East Africa. *Int. J. Climatol.* **2021**, *41*, 6474–6496. [[CrossRef](#)]
27. Jiang, Z.; Li, W.; Xu, J.; Li, L. Extreme precipitation indices over China in CMIP5 models. Part I: Model evaluation. *J. Clim.* **2015**, *28*, 8603–8619. [[CrossRef](#)]
28. Kumar, D.; Mishra, V.; Ganguly, A.R. Evaluating wind extremes in CMIP5 climate models. *Clim. Dyn.* **2015**, *45*, 441–453. [[CrossRef](#)]
29. Eyring, V.; Bony, S.; Meehl, G.A.; Senior, C.A.; Stevens, B.; Stouffer, R.J.; Taylor, K.E. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **2016**, *9*, 1937–1958. [[CrossRef](#)]
30. Taylor, K.E.; Stouffer, R.J.; Meehl, G.A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **2012**, *93*, 485–498. [[CrossRef](#)]
31. Stouffer, R.J.; Eyring, V.; Meehl, G.A.; Bony, S.; Senior, C.; Stevens, B.; Taylor, K.E. CMIP5 scientific gaps and recommendations for CMIP6. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 95–105. [[CrossRef](#)]

32. Grose, M.R.; Narsey, S.; Delage, F.P.; Dowdy, A.J.; Bador, M.; Boschat, G.; Power, S. Insights from CMIP6 for Australia's future climate. *Earth's Future* **2020**, *8*, e2019EF001469. [\[CrossRef\]](#)
33. Di Virgilio, G.; Ji, F.; Tam, E.; Nishant, N.; Evans, J.P.; Thomas, C.; Delage, F. (Selecting CMIP6 GCMs for CORDEX Dynamical Downscaling: Model Performance, Independence, and Climate Change Signals. *Earth's Future* **2022**, *10*, e2021EF002625. [\[CrossRef\]](#)
34. Deng, X.; Perkins-Kirkpatrick, S.E.; Lewis, S.C.; Ritchie, E.A. Evaluation of Extreme Temperatures over Australia in the Historical Simulations of CMIP5 and CMIP6 Models. *Earth's Future* **2021**, *9*, e2020EF001902. [\[CrossRef\]](#)
35. Masud, B.; Cui, Q.; Ammar, M.E.; Bonsal, B.R.; Islam, Z.; Faramarzi, M. Means and Extremes: Evaluation of a CMIP6 Multi-Model Ensemble in Reproducing Historical Climate Characteristics across Alberta, Canada. *Water* **2021**, *13*, 737. [\[CrossRef\]](#)
36. Gusain, A.; Ghosh, S.; Karmakar, S. Added value of CMIP6 over CMIP5 models in simulating Indian summer monsoon rainfall. *Atmos. Res.* **2020**, *232*, 104680. [\[CrossRef\]](#)
37. Akinsanola, A.A.; Ongoma, V.; Koopman, G.J. Evaluation of CMIP6 models in simulating the statistics of extreme precipitation over Eastern Africa. *Atmos. Res.* **2021**, *254*, 105509. [\[CrossRef\]](#)
38. Ukkola, A.M.; De Kauwe, M.G.; Roderick, M.L.; Abramowitz, G.; Pitman, A.J. Robust future changes in meteorological drought in CMIP6 projections despite uncertainty in precipitation. *Geophys. Res. Lett.* **2020**, *47*, e2020GL087820. [\[CrossRef\]](#)
39. Seneviratne, S.I.; Hauser, M. Regional Climate Sensitivity of Climate Extremes in CMIP6 Versus CMIP5 Multimodel Ensembles. *Earth's Future* **2020**, *8*, e2019EF001474. [\[CrossRef\]](#)
40. White, C.J.; McInnes, K.L.; Cechet, R.P.; Corney, S.P.; Grose, M.R.; Holz, G.K.; Bindoff, N.L. On regional dynamical downscaling for the assessment and projection of temperature and precipitation extremes across Tasmania, Australia. *Clim. Dyn.* **2013**, *41*, 3145–3165. [\[CrossRef\]](#)
41. Tamoffo, A.T.; Dosio, A.; Vondou, D.A.; Sonkoué, D. Process-based analysis of the added value of dynamical downscaling over Central Africa. *Geophys. Res. Lett.* **2020**, *47*, e2020GL089702. [\[CrossRef\]](#)
42. Tegegne, G.; Melesse, A.M.; Worqlul, A.W. Development of multi-model ensemble approach for enhanced assessment of impacts of climate change on climate extremes. *Sci. Total Environ.* **2020**, *704*, 135357. [\[CrossRef\]](#)
43. Srivastava, A.; Grotjahn, R.; Ullrich, P.A. Evaluation of historical CMIP6 model simulations of extreme precipitation over contiguous US regions. *Weather Clim. Extrem.* **2020**, *29*, 100268. [\[CrossRef\]](#)
44. Ge, F.; Zhu, S.; Luo, H.; Zhi, X.; Wang, H. Future changes in precipitation extremes over Southeast Asia: Insights from CMIP6 multi-model ensemble. *Environ. Res. Lett.* **2021**, *16*, 024013. [\[CrossRef\]](#)
45. Chen, C.A.; Hsu, H.H.; Liang, H.C. Evaluation and comparison of CMIP6 and CMIP5 model performance in simulating the seasonal extreme precipitation in the Western North Pacific and East Asia. *Weather Clim. Extrem.* **2021**, *31*, 100303. [\[CrossRef\]](#)
46. Wehner, M.; Lee, J.; Risser, M.; Ullrich, P.; Gleckler, P.; Collins, W.D. Evaluation of extreme sub-daily precipitation in high-resolution global climate model simulations. *Philos. Trans. R. Soc. A* **2021**, *379*, 20190545. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Alexander, L.V.; Herold, N. Climactv2 Indices and Software. A Document Prepared on Behalf of the Commission for Climatology (CCL) Expert Team on Sector-Specific Climate Indices (ET-SCI). 2015. Available online: [https://epic.awi.de/id/eprint/49274/1/ClimPACTv2\\_manual.pdf](https://epic.awi.de/id/eprint/49274/1/ClimPACTv2_manual.pdf) (accessed on 1 November 2020).
48. Herold, N.; Alexander, L. Climact 2. 2016. Available online: <https://github.com/ARCCSS-extremes/climact2> (accessed on 1 November 2020).
49. Zivin, J.G.; Shrader, J. Temperature extremes, health, and human capital. *Future Child.* **2016**, *26*, 31–50. [\[CrossRef\]](#)
50. Tabari, H. Climate change impact on flood and extreme precipitation increases with water availability. *Sci. Rep.* **2020**, *10*, 13768. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Pei, F.; Zhou, Y.; Xia, Y. Assessing the Impacts of Extreme Precipitation Change on Vegetation Activity. *Agriculture* **2021**, *11*, 487. [\[CrossRef\]](#)
52. Evans, A.; Jones, D.; Smalley, R.; Lellyett, S. An Enhanced Gridded Rainfall Analysis Scheme for Australia. Bureau Research Report-41. 2020. Available online: <http://www.bom.gov.au/research/publications/researchreports/BRR-041.pdf> (accessed on 1 November 2020).
53. Jones, D.A.; Wang, W.; Fawcett, R. High-quality spatial climate data-sets for Australia. *Aust. Meteorol. Oceanogr. J.* **2009**, *58*, 233–248. [\[CrossRef\]](#)
54. Rupp, D.E.; Abatzoglou, J.T.; Hegewisch, K.C.; Mote, P.W. Evaluation of CMIP5 20th century climate simulations for the Pacific Northwest USA. *J. Geophys. Res. Atmos.* **2013**, *118*, 10–884. [\[CrossRef\]](#)
55. Irving, D.B.; Perkins, S.E.; Brown, J.R.; Gupta, A.S.; Moise, A.F.; Murphy, B.F.; Brown, J.N. Evaluating global climate models for the Pacific island region. *Clim. Res.* **2011**, *49*, 169–187. [\[CrossRef\]](#)
56. Overland, J.E.; Wang, M.; Bond, N.A.; Walsh, J.E.; Kattsov, V.M.; Chapman, W.L. Considerations in the selection of global climate models for regional climate projections: The Arctic as a case study. *J. Clim.* **2011**, *24*, 1583–1597. [\[CrossRef\]](#)
57. Randall, D.A.; Wood, R.A.; Bony, S.; Colman, R.; Fichefet, T.; Fyfe, J.; Taylor, K.E. Climate models and their evaluation. In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the IPCC (FAR)*; Cambridge University Press: Cambridge, UK, 2007; pp. 589–662.
58. Benesty, J.; Chen, J.; Huang, Y.; Cohen, I. Pearson correlation coefficient. In *Noise Reduction in Speech Processing*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 1–4.
59. Chen, W.; Jiang, Z.; Li, L. Probabilistic projections of climate change over China under the SRES A1B scenario using 28 AOGCMs. *J. Clim.* **2011**, *24*, 4741–4756. [\[CrossRef\]](#)



60. Fan, X.; Miao, C.; Duan, Q.; Shen, C.; Wu, Y. The performance of CMIP6 versus CMIP5 in simulating temperature extremes over the global land surface. *J. Geophys. Res. Atmos.* **2020**, *125*, e2020JD033031. [\[CrossRef\]](#)
61. Dai, A. Precipitation characteristics in eighteen coupled climate models. *J. Clim.* **2006**, *19*, 4605–4630. [\[CrossRef\]](#)
62. Guo, Y.; Yu, Y.; Lin, P.; Liu, H.; He, B.; Bao, Q.; Wang, X. Overview of the CMIP6 historical experiment datasets with the climate system model CAS FGOALS-f3-L. *Adv. Atmos. Sci.* **2020**, *37*, 1057–1066. [\[CrossRef\]](#)
63. Craig, A.P.; Versteinsten, M.; Jacob, R. A new flexible coupler for earth system modeling developed for CCSM4 and CESM1. *Int. J. High Perform. Comput. Appl.* **2012**, *26*, 31–42. [\[CrossRef\]](#)
64. Liu, L.; Zhang, C.; Li, R.; Wang, B.; Yang, G. C-Coupler2: A flexible and user-friendly community coupler for model coupling and nesting. *Geosci. Model Dev.* **2018**, *11*, 3557–3586. [\[CrossRef\]](#)
65. CSIRO and Bureau of Meteorology: Climate Change in Australia Information for Australia's Natural Resource Management Regions: Technical Report, CSIRO and Bureau of Meteorology, Australia. 2015. Available online: <https://apo.org.au/sites/default/files/resource-files/2015-01/apo-nid52475.pdf> (accessed on 1 November 2020).
66. Fiddes, S.; Pepler, A.; Saunders, K.; and Hope, P. Redefining southern Australia's climatic regions and seasons. *J. South Hemisph. Earth Syst. Sci.* **2021**, *71*, 92–109. [\[CrossRef\]](#)
67. Iturbide, M.; Gutiérrez, J.M.; Alves, L.M.; Bedia, J.; Cerezo-Mota, R.; Gimadevilla, E.; Vera, C.S. An update of IPCC climate reference regions for subcontinental analysis of climate model data: Definition and aggregated datasets. *Earth Syst. Sci. Data* **2020**, *12*, 2959–2970. [\[CrossRef\]](#)
68. Evans, J.P.; Ji, F.; Lee, C.; Smith, P.; Argüeso, D.; Fita, L. Design of a regional climate modelling projection ensemble experiment—NARCLIM. *Geosci. Model Dev.* **2014**, *7*, 621–629. [\[CrossRef\]](#)
69. Nishant, N.; Evans, J.P.; Di Virgilio, G.; Downes, S.M.; Ji, F.; Cheung, K.K.W.; Tam, E.; Miller, J.; Beyer, K.; Riley, M.L. Introducing NARCLIM1.5: Evaluating the performance of regional climate projections for southeast Australia for 1950–2100. *Earth's Future* **2021**, *9*, e2020EF001833. [\[CrossRef\]](#)
70. Cai, X.; Wang, D.; Zhu, T.; Ringler, C. Assessing the regional variability of GCM simulations. *Geophys. Res. Lett.* **2009**, *36*. [\[CrossRef\]](#)
71. Shi, F.; Wang, Z.; Qi, L.; Chen, R. An Assessment of GCM Performance at a Regional Scale Using a Score-Based Method. *Adv. Meteorol.* **2018**, *2018*, 7641019. [\[CrossRef\]](#)
72. Zhang, M.Z.; Xu, Z.; Han, Y.; Guo, W. Evaluation of CMIP6 models toward dynamical downscaling over 14 CORDEX domains. *Clim. Dyn.* **2022**, 1–15. [\[CrossRef\]](#)
73. Trenberth, K.E. Conceptual framework for changes of extremes of the hydrological cycle with climate change. *Weather Clim. Extrem.* **1999**, *42*, 327–339.
74. Fischer, E.M.; Knutti, R. Observed heavy precipitation increase confirms theory and early models. *Nat. Clim. Chang.* **2016**, *6*, 986–991. [\[CrossRef\]](#)
75. McSweeney, C.F.; Jones, R.G.; Lee, R.W.; Rowell, D.P. Selecting CMIP5 GCMs for downscaling over multiple regions. *Clim. Dyn.* **2015**, *44*, 3237–3260. [\[CrossRef\]](#)
76. Abramowitz, G. Model independence in multi-model ensemble prediction. *Aust. Meteorol. Oceanogr. J.* **2010**, *59*, 3–6. [\[CrossRef\]](#)
77. Abramowitz, G.; Bishop, C.H. Climate model dependence and the ensemble dependence transformation of CMIP projections. *J. Clim.* **2015**, *28*, 2332–2348. [\[CrossRef\]](#)
78. Gleckler, P.J.; Taylor, K.E.; Doutriaux, C. Performance metrics for climate models. *J. Geophys. Res. Atmos.* **2008**, *113*, D6. [\[CrossRef\]](#)
79. Diaconescu, E.P.; Mailhot, A.; Brown, R.; Chaumont, D. Evaluation of CORDEX-Arctic daily precipitation and temperature-based climate indices over Canadian Arctic land areas. *Clim. Dyn.* **2018**, *50*, 2061–2085. [\[CrossRef\]](#)
80. Pope, M.; Jakob, C.; Reeder, M.J. Convective systems of the north Australian monsoon. *J. Clim.* **2008**, *21*, 5091–5112. [\[CrossRef\]](#)
81. Ashfaq, M.; Rastogi, D.; Mei, R.; Touma, D.; Ruby Leung, L. Sources of errors in the simulation of south Asian summer monsoon in the CMIP5 GCMs. *Clim. Dyn.* **2017**, *49*, 193–223. [\[CrossRef\]](#)
82. Agel, L.; Barlow, M. How well do CMIP6 historical runs match observed Northeast US Precipitation and extreme precipitation-related circulation? *J. Clim.* **2020**, *33*, 9835–9848. [\[CrossRef\]](#)
83. Taylor, G.P.; Loikith, P.C.; Aragon, C.M.; Lee, H.; Waliser, D.E. CMIP6 model fidelity at simulating large-scale atmospheric circulation patterns and associated temperature and precipitation over the Pacific Northwest. *Clim. Dyn.* **2022**, 1–20. [\[CrossRef\]](#)
84. Cannon, A.J. Reductions in daily continental-scale atmospheric circulation biases between generations of global climate models: CMIP5 to CMIP6. *Environ. Res. Lett.* **2020**, *15*, 064006. [\[CrossRef\]](#)
85. Johnson, F.; White, C.J.; van Dijk, A.; Ekstrom, M.; Evans, J.P.; Jakob, D.; Westra, S. Natural hazards in Australia: Floods. *Clim. Chang.* **2016**, *139*, 21–35. [\[CrossRef\]](#)
86. Zhao, T.; Guo, W.; Fu, C. Calibrating and evaluating reanalysis surface temperature error by topographic correction. *J. Clim.* **2008**, *21*, 1440–1446. [\[CrossRef\]](#)
87. Lun, Y.; Liu, L.; Cheng, L.; Li, X.; Li, H.; Xu, Z. Assessment of GCMs simulation performance for precipitation and temperature from CMIP5 to CMIP6 over the Tibetan Plateau. *Int. J. Climatol.* **2021**, *41*, 3994–4018. [\[CrossRef\]](#)