Climate Services 30 (2023) 100368

Contents lists available at ScienceDirect

Climate Services

journal homepage: www.elsevier.com/locate/cliser

# A CMIP6-based multi-model downscaling ensemble to underpin climate change services in Australia

Michael R. Grose [a,*], Sugata Narsey [b], Ralph Trancoso [c], Chloe Mackallah [a], Francois Delage [b], Andrew Dowdy [b], Giovanni Di Virgilio [d], Ian Watterson [a], Peter Dobrohotoff [a], Harun A. Rashid [a], Surendra Rauniyar [b], Ben Henley [e,f,g], Marcus Thatcher [a], Jozef Syktus [c], Gab Abramowitz [h], Jason P. Evans [h], Chun-Hsu Su [b], Alicia Takbash [a]

[a] *CSIRO Climate Science Centre, Australia*
[b] *Bureau of Meteorology, Docklands, VIC, Australia*
[c] *University of Queensland, Brisbane, Qld, Australia*
[d] *Department of Planning and Environment, NSW, Australia*
[e] *School of Earth, Atmospheric and Life Sciences, University of Wollongong, Australia, University of Wollongong, NSW, Australia*
[f] *Securing Antarctica's Environmental Future, University of Wollongong, Australia*
[g] *Department of Infrastructure Engineering, University of Melbourne, Australia*
[h] *University of New South Wales, NSW, Australia*

## HIGHLIGHTS

- Australian climate change services can use a regional climate model ensemble as the main data source.
- A sparse matrix of a selected CMIP6 models and multiple regional climate models is outlined.
- Climate sensitivity is considered, treating models outside the *very likely* range with care.
- A storyline framework is used in model selection, ensuring different climate signals are captured.

## ARTICLE INFO

*Keywords:*
Climate services
Regional climate modelling
Storylines

## ABSTRACT

A multi-scenario, multi-model ensemble of simulations from regional climate models is outlined to provide the core data source for a set of climate projections and a climate change service. A subset of realisations from CMIP6 Global Climate Models (GCMs) are selected for downscaling by Regional Climate Models (RCMs) under a 'sparse matrix' framework using the CORDEX guidelines for Shared Socio-economic Pathways that feature low emissions (SSP1-2.6) and high emissions (SSP3-7.0). The subset excludes poor performing models, with performance assessed by the climatology over a large Indo-Pacific domain and an Australian-specific domain, the simulation of atmospheric circulation and teleconnections to major drivers, then incorporating other evaluation from the literature. The models are selected to be relatively independent by simply choosing one model from each 'family' where possible. The projected change in temperature and rainfall in climatic regions of Australia in the selected models are broadly representative of that from the whole CMIP6 ensemble, after deliberately treating models with very high climate sensitivity separately. A limited but carefully constructed ensemble will not represent statistically balanced estimates but can be used effectively under a 'storylines' style approach and can maximise representativeness within limits. The resulting ensemble can be used as a key data source for the future climate component of climate services in Australia. The ensemble will be used in conjunction with CMIP6 and large ensembles of GCM simulations as important context, and targeted 'convective permitting resolution' modelling, deep learning models and emulators for added insights to inform climate change planning in Australia.

* Corresponding author.
*E-mail address:* Michael.Grose@csiro.au (M.R. Grose).

https://doi.org/10.1016/j.cliser.2023.100368
Received 8 August 2022; Received in revised form 14 February 2023; Accepted 11 April 2023

**Practical implications**

Climate services are now called on to include data and information about future projections, and this requires a core data source of model simulations to draw on. This ensemble of simulations should be consistent with and also complement the scientific assessment of multiple lines of evidence on climate change. To be useful in regional applications, it must provide locally relevant insights. Therefore, the ensemble needs to be scientifically credible, as well as relevant at the spatial and temporal scales of interest. The CMIP ensembles of global climate models (GCMs) are internationally respected and credible for producing projections, but have coarse spatial resolution so have limitations in producing locally relevant information. Projections of extreme events and hazards are of strong interest, especially in Australia where they have severe impacts. Downscaling using regional climate models (RCMs) can 'add value' to GCM projections, but often faces limitations in terms of ensemble size due to production costs and can draw from only a restricted set of host models due to data demands.

Here we present one strategy for selecting GCMs from CMIP6 for downscaling by RCMs to provide the core dataset for national climate projections for Australia that can underpin climate services for the coming years. The trade-off between ensemble size and resolution is addressed by employing concepts from the 'storyline' approach, where the choice of models samples the major categories of potential climate futures and takes the emphasis off producing frequentist statistics of projected change. This means ensuring the major categories of change are represented, such as warm and hot, wet and dry, for the major climate regions of the country, and for features than affect climate hazards, such as the shift in the boundary of circulation regimes. Consistency between RCMs, and with other regions in the world is attained by following CORDEX protocols. The implication of this strategy is that a large multi-model, multi-scenario ensemble of RCM simulations, rather than treating CMIP as the core data source, can provide the core data source for national climate projections in Australia for the first time. While the ensemble of RCM simulations at CORDEX resolution can form the core data source, it will need to be complemented by other sources in a coherent framework. These sources include providing the context from CMIP6, including models and concentration pathways not downscaled and insights from large ensembles, as well as finer scale modelling and other modelling from deep learning and emulators.

## Introduction

Climate projections are information products of Earth's climate into future decades and centuries, based on a set of scenarios describing plausible concentrations of future conditions, primarily atmospheric greenhouse gas (GHG) emissions. Projections rely heavily on outputs from global climate models (GCMs), which are typically of coarse spatial resolution (~100–200 km) that don't resolve local scale geographic features and atmospheric processes, so don't directly provide locally relevant information for all purposes. The dynamical downscaling of GCMs using higher resolution regional climate models (RCMs) offers the potential for regional-scale 'added value' - information at finer temporal and spatial scales - which, therefore, may provide more locally applicable climate information (Di Luca et al., 2013). The issue of added value from RCMs has been the topic for some debate (Lloyd et al., 2021), but a convincing case for new insights from downscaling is when there is both added value in the simulation of the current climate together with new detail about the climate change signal in the projections (Ciarlo et al., 2020), termed Realised Added Value (RAV) by Di Virgilio et al. (2020). The source of RAV may be the enhanced spatial resolution of features and their interactions with the parameterised physical processes in the simulation, which can be important factors in determining the climate change signal. These features include the effects of topography, vegetation and soil characteristics, land–water contrasts, and finer-scale processes in the atmosphere (e.g., convection), and the ocean if coupled regional modelling is used (e.g., ocean eddies); see Kotamarthi et al. (2021) and references therein. Higher spatial and temporal resolution is particularly useful when producing information on climate extremes and climate hazards relevant to regional and local scales.

It is possible to use the added value and regional information from dynamical downscaling as part of a comprehensive set of climate projections and as a core part of the future climate component of a climate service. To do this, the structural and sampling uncertainty of climate model ensembles must be accounted for, and in the current context, this requires a systematic approach to using global and regional models (e.g., McGinnis and Mearns, 2021). A current standard for examining future climate projections (e.g., IPCC, 2021) is the Coupled Model Intercomparison Project Phase 6 (CMIP6) multi-model database (Eyring et al., 2016) of climate simulations under the shared Socio-Economic Pathway (SSP) framework (Riahi et al., 2017). The SSPs are five narratives representing alternative plausible future scenarios, with various subsequent GHG concentration pathways. CMIP6 simulations for the SSPs can be used as an input into the downscaling performed by RCMs. Different RCMs may simulate different regional changes given the same GCM input, so a range of RCMs can be surveyed along with the range of GCMs. A large Multi-scenario Multi-model Ensemble (MME) of RCM experiments using multiple GCMs and SSPs is usable as a core part of climate projections, and not just in a complementary role to the GCMs.

Such an MME is necessarily limited in size due to the large amount of computing resources needed to run dynamical downscaling. Therefore, trade-offs must be made and a subset of host GCMs, RCMs and SSPs must be selected, with a limit to the number of ensemble members that can be utilised. Other key decisions include the size of the spatial domains, spatial resolution, and the length of the simulations. Many decisions can now be standardised through international guidelines from the Coordinated Regional Climate Downscaling Experiment (CORDEX-CMIP6), such as the spatial domain, spatial resolutions, SSP priority order and some modelling details (CORDEX, 2021). There are advantages to following these international guidelines in terms of comparability to other domains around the world, the legitimacy bestowed by an international program, and the ease of collaboration when using a common framework. This move to a coordinated program is similar to the move from *ad hoc* programs to CMIP for global modelling. However, even when following CORDEX guidelines, there are several key decisions remaining; primarily the choice of which GCMs and RCMs to use.

A 'sparse matrix' approach to the GCM and RCM selection means that each GCM is downscaled by more than one RCM, but not every RCM needs to be run for each GCM which is the case with a 'filled matrix' approach (Mearns et al., 2012). A sparse matrix MME allows a wider range of GCMs to be sampled for a given number of RCM simulations compared to a filled matrix, but still ensures some estimation of the uncertainty from the RCM step. Given that the uncertainty space is larger from GCMs than RCMs (Abramowitz et al., 2019), the sparse matrix approach has an advantage. There are a range of constraints on RCM applications, mainly from the high computational cost of running RCMs at fine spatiotemporal scales over long time periods suitable for climate analyses. A sparse matrix can include every RCM that is available to contribute and does not rely on the complete agreement of GCM selection by each RCM, so allows for a range of priorities to be considered in its development.

The CMIP6 contributions are not centrally planned, so it is an 'ensemble of opportunity'. But there is the possibility of subsampling CMIP6 using a systematic strategy. Here we describe one strategy for the sub-selection of a group of CMIP6 GCMs for use in a sparse matrix CORDEX MME of RCMs for Australasia, along with some other key details about the strategy behind the ensemble generation. This analysis is not exhaustive, and the recommendations are specific to an initial MME for CORDEX; other downscaling programs may have different needs or priorities. The analysis is also not rigid or final, so it can be built upon in future.

## GCM selection strategy

There are three interrelated aspects to consider when selecting a subset of GCMs for dynamical downscaling: 1) performance of the GCMs, 2) independence of GCMs, and 3) the plausible uncertainty space of projected climate change produced by GCMs. Past studies have placed different emphases on each of the three elements, and how to combine them (e.g., Evans et al., 2014; Teichmann et al., 2021). An additional limitation is that not all modelling groups have published the sub-daily data for all CMIP6 models that many RCMs require. Limited Area Models (LAMs) participating in CORDEX-Australasia, including the Weather Research and Forecasting (WRF, Skamarock et al., 2008) system and the Bureau of Meteorology Atmospheric Regional Projections for Australia (BARPA, Su et al., 2022) require sub-daily inputs. This study follows the strategy of rejecting poor-performing models, then selecting a subset of those models that remain that are relatively independent and representatively sample the projected range of change of interest and have the data required to run RCMs, following previous studies (e.g., Evans et al., 2014).

On the evaluation of GCM performance, we aim to reject models with a simulation of the historical climate that is poor enough to cast doubt on its future projection or is otherwise unsuitable. In this way, the strategy employed here differs from studies that aim to rank models by evaluation metrics and identify best performing models, as for example Desmet et al. (2022) present for southeast Asia. The aim of this study is to reject models that are likely to be unsuitable and retain for consideration all models that are likely to be suitable. However, suitability will inevitably depend on the purpose, and varies by geographic region. Also, the approach assumes that the evaluation of the current climate is useful to inform projections, which may be vulnerable to the effects of model tuning and compensating errors when applied in the RCM framework (e.g., Di Luca et al., 2021). Also, there are many dimensions on which to assess suitability, the range of metrics available is typically not comprehensive, some metrics correlate with others, observation datasets vary in quality and there are often not clear objective criteria or thresholds to determine what is acceptable.

There are open questions about the appropriate categories of metrics to use and the spatial domain over which to assess surface climate statistics for GCM selection. Detailed statistics on surface variables from within the target domain can utilise high-quality observed datasets and are potentially useful indicators of the GCM simulation of the local climate (e.g., Syktus et al., 2022), assuming this translates quite directly

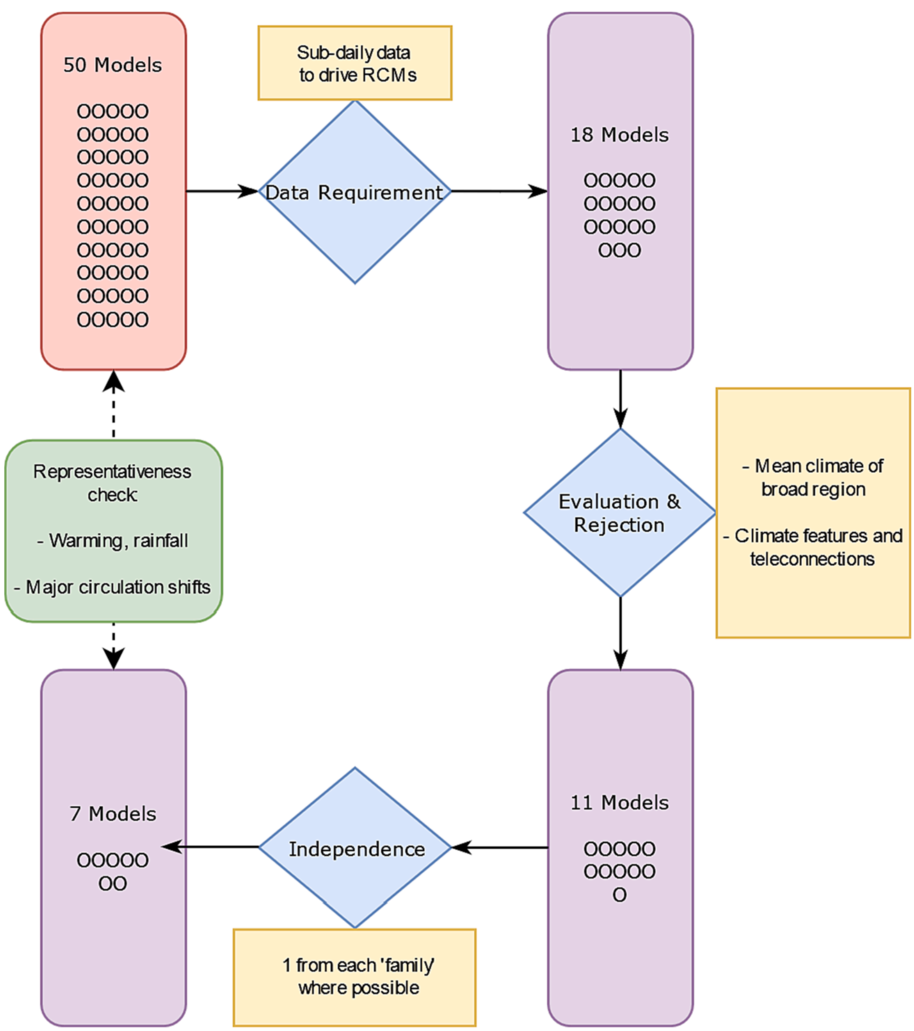


**Fig. 1.** A flow chart schematic of the CMIP6 model selection process, including the number of models at each stage.

from the GCM into the RCM simulation. However, the surface climate over the target domain may be seen as largely a function of the RCM simulation, so that it is more important to evaluate the climate at the RCM boundary (e.g., Moalafhi et al., 2016) or the relevant large-scale features and processes. These features may include global energy balance, sea surface temperature patterns, atmospheric circulation features, modes of variability and so on. The evaluation at the local and broad scales may be consistent or could conceivably differ in some cases. There is also an open question of how to incorporate historical trends (regional or large-scale) in evaluation.

Here, we take a simple and practical approach to GCM selection, see the schematic in Fig. 1. We examine a selection of evaluation metrics, including surface climate statistics, climate processes and modes of variability. Evaluation of surface climate statistics is focussed mainly on a large spatial domain over the Indo-Pacific region. This domain includes broad-scale features that commonly have model biases associated with them, such as the edge of the Western Pacific Warm Pool (WPWP) where models commonly have the so-called 'cold-tongue bias', the Intertropical Convergence Zone (ITCZ) where models may have the 'double ITCZ' bias, and the temperature gradient from the equator to the Southern Ocean, where models can have a warm bias. The domain also contains large rainfall features such as the southern edge of the Asia-Australia monsoon and the South Pacific Convergence Zone (SPCZ). A selection of large-scale circulation features, climate drivers and modes of variability are assessed. Daily data are also examined for some extremes, including for the projected changes in the annual maximum of daily rainfall, with a general comparison presented of those results in relation to the projected changes in mean values. The metrics were weighted by variable or category before weighting between variables, to avoid overweighting metrics that are highly correlated (following Rupp et al., 2013).

The results from this evaluation of broad spatial scale surface variables and climate drivers are then compared with analyses from the literature, including assessments of the GCM performance within the Australian domain (including extremes) and past trends. The results are compared, and the model list is reduced. We reject the poorest performing models rather than select a small group of highest performers. This will likely exclude the most unsuitable GCMs for the many uses of the MME, while retaining models that may be acceptable despite not performing best, thus retaining as much of the range of plausible future states as possible for consideration.

On the independence of GCMs, an ensemble with a small number of members may not sufficiently span model uncertainty if the models are closely related. Certain modelling choices or representations may dominate the GCM subsample and effectively provide redundant information in the final ensemble, potentially leading to overconfidence in a particular set of outcomes. There are various approaches to measuring independence and accounting for this issue (Abramowitz et al., 2019), but the overall goal for any GCM selection is to minimise the effects of model dependence by not treating the outputs of highly related models as independent samples. Independence in the MME can be addressed in various ways in practice, and one simple method is to select models that are above some threshold of independence from each other (e.g., Pennell & Reichler, 2011; Herger et al., 2018), or only selecting one model from each 'family' of models (Leduc et al., 2016; Abramowitz et al., 2019). As with GCM performance, model independence can be specific to the region, time period, variable and metric. We take a simple approach to independence suggested by Abramowitz et al. (2019), by ensuring that we select models that are not highly related (e.g., by sharing model components); we do not weight independence further. We place a much higher emphasis on the representativeness of the change signal.

On representativeness, the aim is to draw a representative sample of plausible climate change signals from the ensemble of models used as input (the host ensemble). This again is domain-specific and can be purpose-specific, and in some contexts might be considered an aspect of model dependence. One strategy is to reproduce similar statistics (e.g., mean and spread) in the sub-sample as in the full ensemble for key climate quantities (e.g., mean temperature and precipitation) for an appropriate spatial domain of interest. As for evaluation, there is no objective set of criteria to determine the appropriate spatial domain to assess representativeness, from the smallest domain of interest (e.g., a city) to larger scale climate zones (e.g., the monsoons or mid-latitude region). This is again a matter of judgement and depends on the purpose of the ensemble.

Rather than aim to reproduce the same distribution of projected change as the host ensemble (e.g., the mean and spread), it may be appropriate to deliberately draw from a constrained range of change signal. Rejecting models that evaluate poorly may do this, but also choosing selectively from non-rejected models may be appropriate if the sample of projected changes in the host ensemble are known to be skewed. In CMIP6, there is an uneven distribution of climate sensitivity values compared to an independent assessment (Sherwood et al., 2020), including many models outside the *likely* range (2.5 to 4 °C) and some outside the *very likely* range (2 to 5 °C), but still considered possible (IPCC, 2021). There is a strong case to not draw evenly across the CMIP6 spread in climate sensitivity (and therefore warming projection, as they are strongly correlated), and especially to limit sampling of the very high end of the range of climate sensitivities (Hausfather et al., 2022). In response, we select a limited number of models from the group above the *likely* range of Equilibrium Climate Sensitivity (ECS, also known as Effective Climate Sensitivity when measured using the standard method) as possible, ensure a model representative of the low end of the *likely* range in ECS is selected, and treat models with ECS above the *very likely* range as a 'low probability high impact' case, noting that it is not possible to rule out ECS values $> 5$ °C (IPCC, 2021).

Given the limited number of possible simulations, we don't attempt to sample the CMIP6 uncertainty space in a statistically detailed way. Rather, we sample the main categories of change in surface temperature and rainfall for eight broad climatic regions of Australia, including warm and hot, wet and dry. We use the change in these variables between the two 50-year periods of 1950–1999 and 2050–2099 under SSP3-7.0 as being representative of a strong climate change signal, with reduced effects of internal variability compared to using 20-year periods. Given the model ensemble will be used to examine hazards, we assess the simulation and change signal in some features known to affect the projection of coastal hazards: the movement of the subtropical ridge and monsoon shear line, which correlates with changes to the latitude of storms.

The study aims to choose 5–8 GCMs from the models that provide sub-daily data but allow for further GCMs with sub-daily data available to be downscaled if resources permit. Models that don't provide sub-daily data but can be used in RCMs that do not need it are also noted. The selection is compared to assessments from other studies, and a sparse matrix is devised, noting all modelling will follow the CORDEX guidelines (CORDEX, 2021).

## Detailed methods

All methods and choices are consistent with CORDEX guidelines, including the selection of SSPs with the two highest priorities being SSP1-2.6 and SSP3-7.0, followed by SSP2-4.5 and SSP5-8.5. The CORDEX guidelines also specify the Australasian spatial domain (Fig. 2) and spatial resolution (12 or 25 km), however, we note that further regional modelling for smaller sub-domains and at finer resolution will use these CORDEX simulations as input.

A set of 50 CMIP6 models are considered, with historical simulations available at the time of analysis, and one simulation from each is used in evaluation, typically r1i1p1f1 (Table 1). However, at the time of writing only 35 of the 50 models provide monthly and daily data for Historical and the two highest priority SSPs. A significant constraint is that only 18 of the 50 models provide sub-daily outputs for historical and these SSP simulations.

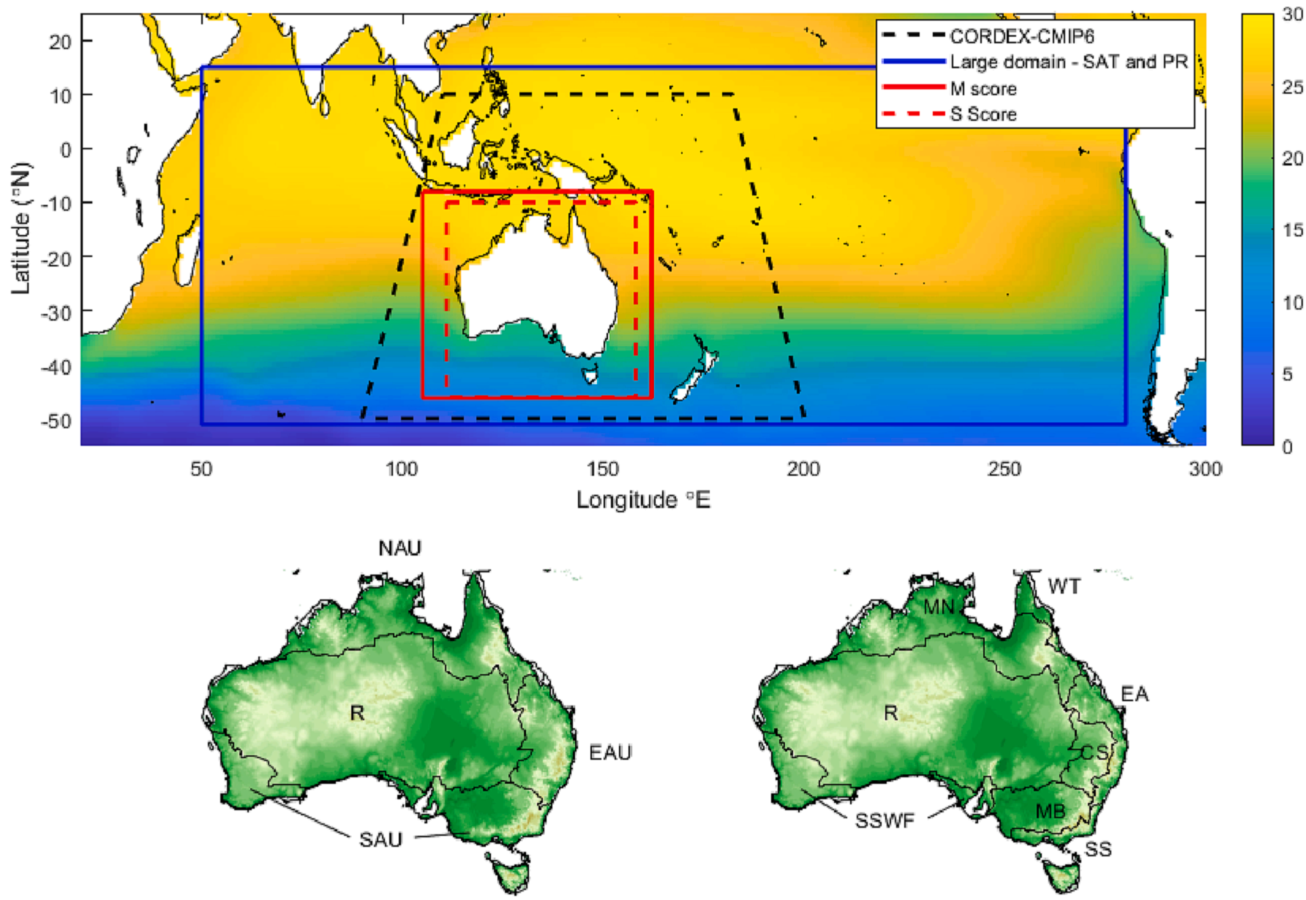


**Fig. 2.** Map of domains used, top shows the CORDEX Australasia domain and large Indo-Pacific domain used to assess surface air temperature (SAT) and precipitation (PR) spatial distributions, and the domain used to calculate the M Score for MSLP and S score, colour scale indicates SST from HadISST; bottom left shows the four 'supercluster' regions (NAU Northern Australia, R Rangelands, SAU Southern Australia, EAU Eastern Australia), bottom right shows the eight cluster regions (MN Monsoonal North, WT Wet Tropics, R Rangelands, SSWF Southern and Southwestern Flatlands, MB Murray Basin, CS Central Slopes, EA Eastern Australia, SS Southern Slopes).

Analysis of regional averages is generally done on the eight large 'cluster' regions used in CSIRO and BoM (2015), see Fig. 2. Four broad 'super-clusters' are also considered, broadly similar to the IPCC reference regions (Iturbide et al. 2020). These are Eastern Australia (EA), Northern Australia (NA), Southern Australia (SA) and the centre, referred to as Rangelands (R), also see Fig. 2.

*Evaluation*

Observed datasets used for evaluation of the historical simulations are the global rainfall from the Global Precipitation Climatology Project, GPCPv2.3 (Adler et al., 2017), sea surface temperature from Hadley Centre Sea Ice and Sea Surface Temperature dataset HadISST1.1 (Rayner et al., 2003), atmospheric pressure and circulation from ERA5 (Hersbach et al., 2020) and high-quality gridded climate data of surface air temperature (SAT) and precipitation (PR) from the Australian Gridded Climate Data (AGCD) of Evans et al. (2020) and Jones et al. (2009).

An assessment of only the first realisation is used, and it is assumed that this is broadly indicative of the model simulation across multiple members, as has been done previously (e.g., Grose et al., 2020; Di Virgilio et al., 2022). Evaluation is used to find generally poor performing models and is not used to make fine distinctions at the level of differences between ensemble members. Also, a large domain and long temporal periods are used in the evaluation of the mean climate, to be less sensitive to phases of internal variability. Model data were regridded to a 1.5° lat/lon grid (the mean resolution of CMIP6) using bilinear interpolation for analysis that requires this consistency, otherwise analysis was performed on the original grid. The evaluation is focussed mainly on the mean climate in terms of SAT and PR over the large Indo-Pacific domain and over the whole of Australia (Fig. 2), seasonal mean sea level pressure (MSLP) distribution, as well as atmospheric circulation, teleconnections and variability. Australian multi-year meteorological drought is assessed compared to AGCD from estimates of the 50- and 100-year Average Recurrence Interval 5 and 10-year drought totals using a probabilistic approach based on Srikanthan and McMahon (1986).

The evaluation in this study is then compared to and used jointly with material from the literature. These other studies include the parallel per-GCM assessments of mean and extreme climate and daily distributions of temperature and rainfall in the Australian domain by Di Virgilio et al. (2022) and the Kling-Gupta (Gupta et al., 2009) skill scores for annual cycle and seasonal temperature and rainfall characteristics calculated by Syktus et al. (2022). The evaluation also considers the past trends in global temperature in Tokarska et al. (2020), the assessment of Australian heatwaves in Hirsch et al. (2021), and the simulation of the Asia-Australia monsoons in Narsey et al. (2020). Also, a poor simulation of important features in the western Pacific are expected to have flow-on effects on Australia, so models are also flagged here from the assessment of the edge of the west Pacific warm pool (WPWP) by Grose et al. (2020), and the simulation of the South Pacific Convergence Zone in Narsey et al. (in press).

After weighting models for each category, a simple system of model

**Table 1**
Details of 50 CMIP6 models considered in this study and ranking for evaluation metrics for each model. The number in brackets indicates the number of realisations with daily data available, models highlighted in blue have daily sub-data available for all SSPs required. Numbered cells show the ranking of the model based on the average of tests within that category, binary (0/1) cells show the results of a threshold test, X marks where a rejection was made due to the number of tests failed, or egregious fails, even if more categories were not in the bottom third (see methods). Last column shows the final tally for each model: dark red = reject, light red = one flag, blue = insufficient evaluation, white = no flags.

| | Model | Atmos. Resolution | Precip. Corr. RMSE | Temp. Corr. RMSE | M score | S score | Pass or fail | ENSO | IOD | SAM | STR | Jet | Storm track | Drought | Pass or fail | Global warming | Qld | NSW | Heatwaves | WPWP/SPCZ | Flag |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | ACCESS-CM2 (5) | 1.2 x 1.8 | 30 | 7 | 5 | 14 | | 32 | 25 | 0 | 0 | 1 | 4 | 0 | | | | | | | |
| 2 | ACCESS-ESM1-5 (40) | 1.2 x 1.8 | 31 | 17 | 20 | 10 | | 27 | 6 | 0 | 0 | 2 | 5 | 0 | | | | | | | |
| 3 | AWI-CM-1-1-MR (5) | 0.9 x 0.9 | 23 | 9 | 28 | 25 | | 34 | 40 | | 0 | 13 | | 0 | | | | X | | | |
| 4 | BCC-CSM2-MR (1) | 1.1 x 1.1 | 33 | 21 | 22 | 20 | | 37 | 30 | 0 | 0 | 3 | 25 | 1 | | | | X | X | | |
| 5 | BCC-ESM1 | 2.8 x 2.8 | | | | | | | | | 0 | | | 1 | | | | | | | |
| 6 | CAM5-CSM1-0 (2) | ~1.0 | | | 37 | 35 | | 23 | 39 | 0 | 0 | | | 1 | | | | | | | |
| 7 | CanESM5 (50) | 2.8 x 2.8 | 26 | 18 | 20 | 24 | | 20 | 26 | | 0 | 18 | 18 | 1 | | X | | | X | | |
| 8 | CanESM5-CanOE (3) | 2.8 x 2.8 | | | 23 | | | | | | | | | | | | | | | | |
| 9 | CAS-ESM2-0 (2) | 1.4 x 1.4 | 35 | 40 | 38 | 38 | | 29 | 21 | | 0 | 32 | | 1 | | | | | | | |
| 10 | CESM2 (3) | ~1.0 | 5 | 11 | 6 | 4 | | 12 | 7 | | 0 | 19 | 3 | 0 | | | | | | | |
| 11 | CESM2-WACCM (1) | 1.0 x 1.3 | 4 | 10 | 12 | 5 | | 13 | 4 | | 0 | 12 | 1 | 0 | | | | | | | |
| 12 | CIESM | 1.0 x 1.0 | 25 | 12 | 45 | | | 5 | 3 | 1 | 0 | 7 | | 0 | | | X | X | | | |
| 13 | CMCC-CM2-SR5 (1) | ~0.9 | 18 | 10 | 26 | 33 | | 2 | 5 | | 0 | 20 | 6 | 1 | | | | X | | | |
| 14 | CMCC-ESM2 (1) | 0.9 x 1.3 | 14 | 4 | 16 | 30 | | 3 | 19 | | 0 | | 2 | | | | | | | | |
| 15 | CNRM-CM6-1 (6) | 1.4 x 1.4 | 13 | 15 | 13 | 19 | | 24 | 15 | 0 | | | | 0 | | | | | | | |
| 16 | CNRM-CM6-1-HR (1) | ~0.5 | | | 18 | | | | | 1 | | | | | | | | | | | |
| 17 | CNRM-ESM2-1 (5) | 1.4 x 1.4 | 20 | 15 | 28 | 17 | | 33 | 16 | | | | | 0 | | | | | | | |
| 18 | E3SM-1-1 | ~1.0 | 23 | 15 | 35 | 7 | | 14 | 10 | | 1 | 14 | | | | | | | | | |
| 18 | EC-Earth3 (57) | 0.7 x 0.7 | 6 | 22 | 9 | 11 | | 22 | 17 | 0 | 0 | 10 | 8 | 0 | | | | | | | |
| 20 | EC-Earth3-AerChem (1) | ~1.0 | 9 | 23 | 23 | 13 | | 30 | 36 | | 0 | | 22 | | | | | | | | |
| 21 | EC-Earth3-CC | ~1.0 | 9 | 25 | 13 | 12 | | 6 | 20 | | 0 | | 23 | | | | | | | | |
| 22 | EC-Earth3-Veg (6) | 0.7 x 0.7 | 8 | 22 | 10 | | | 40 | 38 | | 0 | 5 | 16 | 0 | | | | | | | |
| 23 | EC-Earth3-Veg-LR (3) | 0.7 x 0.7 | 13 | 24 | 27 | 18 | | 10 | 32 | | 1 | 8 | 26 | 0 | | | | | | | |
| 24 | FGOALS-f3-L (1) | 2.3 x 2.0 | 17 | 18 | 18 | 6 | | 11 | 24 | 0 | 0 | 9 | 14 | 0 | | | | X | | | |
| 25 | FGOALS-g3 | 2.3 x 2.0 | 35 | 15 | 39 | 37 | | 21 | 33 | 0 | 0 | 11 | 11 | 0 | | | | | | | |
| 26 | FIO-ESM-2-0 | 1.3 x 0.9 | 17 | 3 | 11 | | | 17 | 14 | | 0 | 22 | | 1 | | | | | | | |
| 27 | GFDL-CM4 | 1.0 x 1.3 | 5 | 8 | 8 | 1 | | 18 | 9 | 0 | 0 | 26 | 14 | 0 | | | | | | | |
| 28 | GFDL-ESM4 (1) | 1.0 x 1.3 | 13 | 2 | 4 | 2 | | 4 | 2 | 0 | 0 | 16 | | 0 | | | | | | | |
| 29 | GISS-E2-1-G (15) | 2.0 x 2.5 | 35 | 25 | 13 | 15 | | 38 | 27 | | 0 | 4 | 13 | 0 | | | | | | X | |
| 30 | HadGEM3-GC31-LL | 2.2 x 2.2 | | | 2 | | | | | | | | | | | X | | | | | |
| 31 | HadGEM3-GC31-MM | 0.9 x 0.9 | | | 1 | | | | | | | | | | | | | | | | |
| 32 | IITM-ESM (1) | ~1.89 | 23 | 26 | 34 | 32 | | 35 | 29 | | 0 | | | | | | | | | | |
| 33 | INM-CM4-8 (1) | 1.5 x 2.0 | 37 | 20 | 43 | 29 | | 41 | 34 | 0 | 0 | 30 | 24 | 1 | | | | | | X | |
| 34 | INM-CM5-0 (5) | 1.5 x 2.0 | 35 | 21 | 44 | 23 | | 43 | 42 | 0 | 0 | 21 | 9 | 0 | | | | | | X | |
| 35 | IPSL-CM5A2-INCA (1) | 1.3 x 2.5 | 38 | 24 | 41 | 26 | | 42 | 41 | | 1 | | 20 | | | | | | | | |
| 36 | IPSL-CM6A-LR (11) | 1.3 x 2.5 | 35 | 9 | 31 | 21 | | 28 | 23 | 1 | 0 | 28 | 12 | 0 | | | X | X | | | |
| 37 | KACE-1-0-G (3) | 2.2 x 2.2 | 34 | 11 | 7 | 9 | | 8 | 13 | 0 | 0 | | | 0 | | | | | | | |
| 38 | KIOST-ESM | 2.2 x 2.2 | | | | | | | | | 0 | | | | | | | X | | | |
| 39 | MCM-UA-1-0 (1) | 2.2 x 3.8 | | | 45 | | | | | | | | | | | | | | | | |
| 40 | MIROC6 (3) | 1.4 x 1.4 | 5 | 16 | 48 | 27 | | 15 | 11 | 0 | 1 | 29 | 30 | 1 | | | | X | | | |
| 41 | MIROC-ES2L (10) | 4.5 x 4.5 | 16 | 22 | 47 | 31 | | 26 | 37 | 0 | | | | 1 | | | | X | | | |
| 42 | MPI-ESM-1-2-HAM | 2.2 x 2.2 | 33 | 24 | 40 | 36 | | 25 | 18 | | 0 | 25 | 28 | 0 | X | | | | | X | |
| 43 | MPI-ESM1-2-HR (10) | ~0.9 | 20 | 12 | 36 | | | 36 | 35 | | 0 | 17 | 21 | 1 | | | | | | X | |
| 44 | MPI-ESM1-2-LR (10) | ~2.0 | 16 | 23 | 30 | 28 | | 31 | 43 | 0 | 0 | 31 | 19 | 0 | | | | | | X | |
| 45 | MRI-ESM2-0 (5) | 1.1 x 1.1 | 33 | 16 | 33 | 22 | | 9 | 1 | 0 | 0 | 27 | 27 | 0 | | | | X | | | |
| 46 | NESM3 | 1.9 x 1.9 | 27 | 29 | 42 | 34 | | 39 | 28 | 0 | 0 | 23 | 10 | 0 | | | | X | | | |
| 47 | NorESM2-LM (1) | 1.9 x 2.5 | 23 | 24 | 32 | | | 19 | 12 | | 0 | 6 | 29 | | | | X | X | | | |
| 48 | NorESM2-MM (1) | 0.9 x 0.9 | 1 | 6 | 16 | 3 | | 1 | 8 | 0 | 0 | 15 | 17 | 1 | | | X | | | | |
| 49 | TaiESM1 (1) | 0.9 x 0.9 | 10 | 4 | 25 | 16 | | 7 | 22 | | 0 | 24 | 7 | 1 | | | | X | | | |
| 50 | UKESM1-0-LL (16) | 1.3 x 1.9 | 23 | 2 | 3 | 8 | | 16 | 31 | 0 | | | | 0 | | | | | | | |

rejection is used, generally where models falling in the bottom third of models for the majority of tests in a category are deemed to fail that category, consistent with Rupp et al. (2013) and Di Virgilio et al. (2022). Models that are flagged in two of the three categories are chosen for rejection and models flagged in one are used with lower priority.

*Mean climate*

Four different scores of the spatial distribution in the mean climatology for both SAT and PR are calculated and compared, and one score also evaluates Mean Sea Level Pressure (MSLP). The first two scores are spatial correlation and Root Mean Square Error (RMSE) compared to observations for two six-month seasons (June-Nov and Dec-May) over a large Indo-Pacific domain (Fig. 2). The third score is the M statistic (Watterson, 1996), calculated with TAS and PR over the Australian land area relative to the higher resolution observed dataset of AGCD, and for MSLP over the 105–162 °E and 8–46 °S domain relative to ERA5. The fourth is the S score (Taylor, 2001) of SAT and PR was derived over a domain of 10 to 46 °S and 111 to 158 °E, using the formula:

$$S = \frac{4(1+R)^4}{\left(\sigma_f + 1/\sigma_f\right)^2 (1+R_0)^4}$$

where $\sigma_f$ is the ratio of pattern standard deviations of the simulated SATs and the reference SAT (from ERA5), R is the corresponding pattern correlation between the simulated and reference SATs and $R_0$ is the maximum attainable correlation (taken as 1 in our calculation).

The seasonal correlations and RMSE were averaged to make one score for SAT and one for PR, then weighted equally with M Statistic averaged for TAS, PR and MSLP and the S score averaged for TAS and PR. These evaluation metrics are not all independent, with evaluation of the same variable (e.g., all scores use TAS), and of calculations (e.g., the S score uses spatial correlation as one component). To reduce the redundancy of non-independent scores, we focus on one score from each method (making a total of four scores), averaging scores for different seasons first, and using only the final M score and S score (not the components). Then models in the bottom third for two or more of the four score metrics were flagged for possible rejection.

*Atmospheric circulation, teleconnections and variability*

Mean atmospheric circulation relevant to Australian climate in the models was assessed using a series of indices of specific features. Indices of the mean boundary between dominant zonal wind regimes were calculated, specifically the mean position and intensity of the Subtropical Ridge (STR) of pressure along a meridional transect in eastern Australia following the methods in Grose et al. (2015), and the mean position of the monsoon shear line in northern Australia defined by 925 hPa winds following methods in Colman et al. (2011). These features are not only relevant to precipitation, but also the latitude that storms affect coastlines in Australia.

The mean position and strength of the subtropical jet and the storm track over an Australian region in summer (DJF) and winter (JJA) were calculated using methods outlined in Grose et al. (2017).

The simulation of the El Niño Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD) and the teleconnection to Australian rainfall for the four regions were assessed using methods described in Grose et al. (2020). This evaluation compared the temporal correlation between indices of each driver and winter-spring rainfall variability to the relationship in observations. The correlation in each region was averaged to give a single score. In addition, the models with poor historical trends in the Southern Annular Mode (SAM) were flagged.

As for mean climate, seasonal scores of each type were averaged. The ranking of models for the simulation of drought were also combined, creating a total of seven scores. Models with four or more scores in the bottom third of all models were flagged for possible rejection. Exceptions to this 'bottom third' rule are where absolute thresholds are used for the STR (model is outside observed range for the location of the peak STR) and monsoon shear line (location is $> 2^\circ$ latitude from observed).

*Independence*

The simple approach to model independence of choosing no more than one model from each 'family' from Abramowitz et al. (2019) was employed where possible, as outlined in the strategy section. The phylogeny analysis and minimum generalised distance threshold were taken from Brunner et al. (2020). Given the main selection must be made from models with sub-daily data available and many of these are related, this means selecting only one model from the six 'families' with more than one model represented (Table 2). However, given that the model list is already restricted due to sub-daily data availability (from 50 to 18 models) and the rejection of some models (from 18 to 7 models with 2 more with flags), it is only possible to pick 6 independent models. An additional model CESM2 is also retained, even though it is considered to be in the same family as NorESM2, since the family connection is weaker than some other instances, including an entirely different ocean and biogeochemistry model, as well as different aerosol physics and chemistry (Seland et al., 2020). Note that the ACCESS-ESM1.5 model is within the UKMO family but is assessed to be more independent from the others at a level greater than the generalised distance threshold (due at least in part to the code version of the major ocean and atmosphere components used in this model).

*Representativeness*

Global warming trends were used as a line of evidence in the evaluation stage, but are also considered here in terms of representativeness, along with the Equilibrium Climate Sensitivity (ECS) value associated with each model. ECS estimates are taken from Zelinka (2022). The CMIP6 spread has an over-representation of the warm end of global warming, largely related to the climate sensitivity, as mentioned in the strategy section. This issue is exacerbated somewhat when considering only the 18 models with sub-daily data available (Fig. 3). Modelled high climate sensitivity generally also leads to high warming in Australia, so the model spread in projected temperature is not consistent with that expected from an independent assessment of temperature change. There are limited strategies that can be used in response to this issue when choosing a small subset of models from the group of 18, given the lack of models with required data, excluding rejected models and accounting for independence. So here a practical approach is taken by selecting as few models with ECS above the *likely* range as possible (but this space is still over represented), and deliberately including models' representative of other parts of the *likely* range. Models with ECS above the *very likely* range of climate sensitivity can be examined as a 'low probability, high impact' outcome, and taking care when using this as part of an ensemble.

After considering climate sensitivity, the subset of models is chosen to be representative of the CMIP6 spread in the climate change signal in two ways. The first is the projected change in mean annual temperature and precipitation over the eight cluster regions of Australia for a strong climate change signal (SSP3-7.0) over the whole century (1950–1999 to 2050–2099). These changes in the mean are expected to correlate with projected changes in associated extremes such as heatwaves and time in drought. The 50-year periods were chosen to show the change signal with reduced influence from natural variability. An additional analysis of the change between 20-year periods 1995–2014 to 2040–2059 and 2080–2099 was examined, to assess the consistency of the trends through time periods of interest for applications.

The second assessment is of the change in the circulation features known to affect the change in the latitude of coastal storms – the subtropical ridge and the monsoon shear line. Here, we aim to include at least one model that projects each direction of change in latitude of these features in CMIP6 models (poleward, little change or equatorward).

Climate extremes are of strong interest, but the purpose of this study is to inform the choices for dynamical downscaling, rather than the projection of extreme events by the GCMs themselves. RCMs inherit only broad-scale climate features from GCMs, then produce their own simulation of extremes at the regional scale. Therefore, only a general overview of the projected change in rainfall extremes relative to the mean warming is given here as general context, using the monthly maximum values of daily rainfall (rx1day) from the ETCCDI (Expert Team on Climate Change Detection and Indices). Data were obtained for many of the GCMs assessed in this study from the Copernicus Climate Data Store, as documented in Sillmann et al. (2013) and Kim et al.

**Table 2**
The group of 18 models with sub-daily data available, the colour code over the name and the letter in the Family column indicates their 'family' identified in Brunner et al. (2020), ECS denotes the Equilibrium Climate Sensitivity (ECS) of the model with those above the assessed likely range (light red) and above the very likely range (dark red). Subsequent columns indicate the flag from model evaluation by category of mean climate, drivers and other evidence, the group based on evaluation flags and selection for use in downscaling. For equivalent information for all 50 models, see Table S1.

| | Model | Family | ECS | Mean Climate | Drivers | Other evidence* | | Group | Selection |
|---|---|---|---|---|---|---|---|---|---|
| 1 | ACCESS-CM2 | A | 4.66 | | | | | 1 | |
| 2 | ACCESS-ESM1.5 | | 3.88 | | | | | 1 | |
| 3 | BCC-CSM2-MR | | 3.02 | X | X | HW | | 3 | |
| 4 | CanESM5 | | 5.64 | | X | Warming | HW | 3 | |
| 5 | CESM2 | B | 5.15 | | | Syk | | 1 | |
| 6 | CMCC-CM2-SR5 | C | 3.55 | | | DV | Syk | 2 | |
| 7 | CMCC-ESM2 | C | 3.58 | | | | | 1 | |
| 8 | CNRM-ESM2-1 | | 4.79 | | | | | 1 | |
| 9 | EC-Earth3 | D | 4.26 | | | | | 1 | |
| 10 | EC-Earth3-Veg | D | 4.33 | | | | | 1 | |
| 11 | IPSL-CM5A2-INCA | E | 3.82 | X | X | | | 3 | |
| 12 | IPSL-CM6A-LR | E | 4.70 | | X | DV | Syk | 3 | |
| 13 | MIROC6 | | 2.60 | X | X | DV | Syk | 3 | |
| 14 | MPI-ESM1-2-HR | F | 2.98 | | X | WPWP | Syk | 3 | |
| 15 | MPI-ESM1-2-LR | F | 3.03 | | X | SPCZ | | 3 | |
| 16 | NorESM2-LM | B | 2.56 | | | DV | Syk | 2 | |
| 17 | NorESM2-MM | B | 2.49 | | | Syk | | 1 | |
| 18 | UKESM1-0-LL | A | 5.36 | | | | | 1 | High ECS |

* Other Evidence: Warming denotes global temperature trends from Tokarska et al. (2020), DV denotes evaluation within Australia by Di Virgilio et al. (2022), Syk denotes evaluation by Syktus et al. (2022), WPWP denoted West Pacific Warm Pool from Grose et al. (2020), SPCZ denotes South Pacific Convergence Zone from Narsey et al. (in press), HW denotes heatwaves from Hirsch et al. (2021).

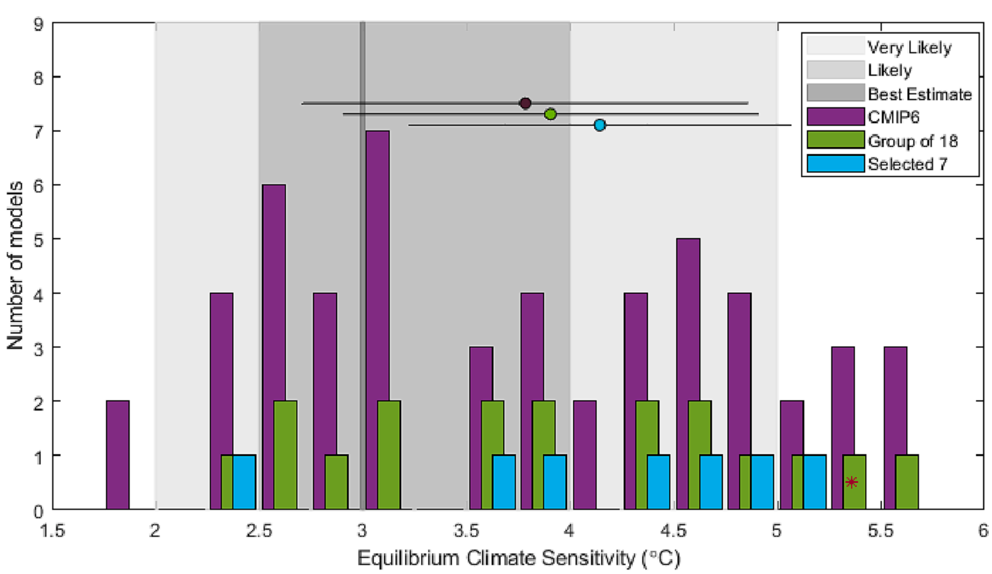


**Fig. 3.** The equilibrium climate sensitivity (ECS) of CMIP6 models, and subsets of models compared to the independent assessment reported in IPCC (2021) as grey shading. Bars show count of models (in 0.25 °C bins) for 53 CMIP6 models assessed by Zelinka (2022) as a reflection of the whole ensemble, the 18 models providing sub-daily data, and the final selected 7 models, circles and stems show the mean and standard deviation of the three model groups, the extra 'low probability, high impact' case of UK-ESM1-0-LL is marked by a red star.

(2020). Projected changes are given as for mean change: for the eight 'cluster' regions for the SSP3-7.0 emissions pathway over the whole century (1950–1999 to 2050–2099).

## Results

### *Evaluation*

The four scores evaluating the mean climatology in seasonal SAT and PR (and one of MSLP) do not show clear groupings of models, but rather a spectrum of scores (Figs. 4–6). Spatial correlations of temperature over this large domain are all very high with little distinction between models, so model ranking is determined more by RMSE for this analysis (Fig. 4a). The model ranking for PR has only broad similarity to SAT, with some notable differences in the ordering (Fig. 4b). The S scores of annual SAT and PR (Fig. 5) and the M Statistic of SAT, PR and MSLP (Fig. 6) over Australian domains also show a spectrum of scores. There are some similarities in the model ranking from the evaluations over a large domain and those for S score and M statistics over Australia, but also some interesting contrasts between the two rankings. The differences are most likely reflecting the different spatial domains. For example, MIROC6 ranks highly over the large domain, but poorly over

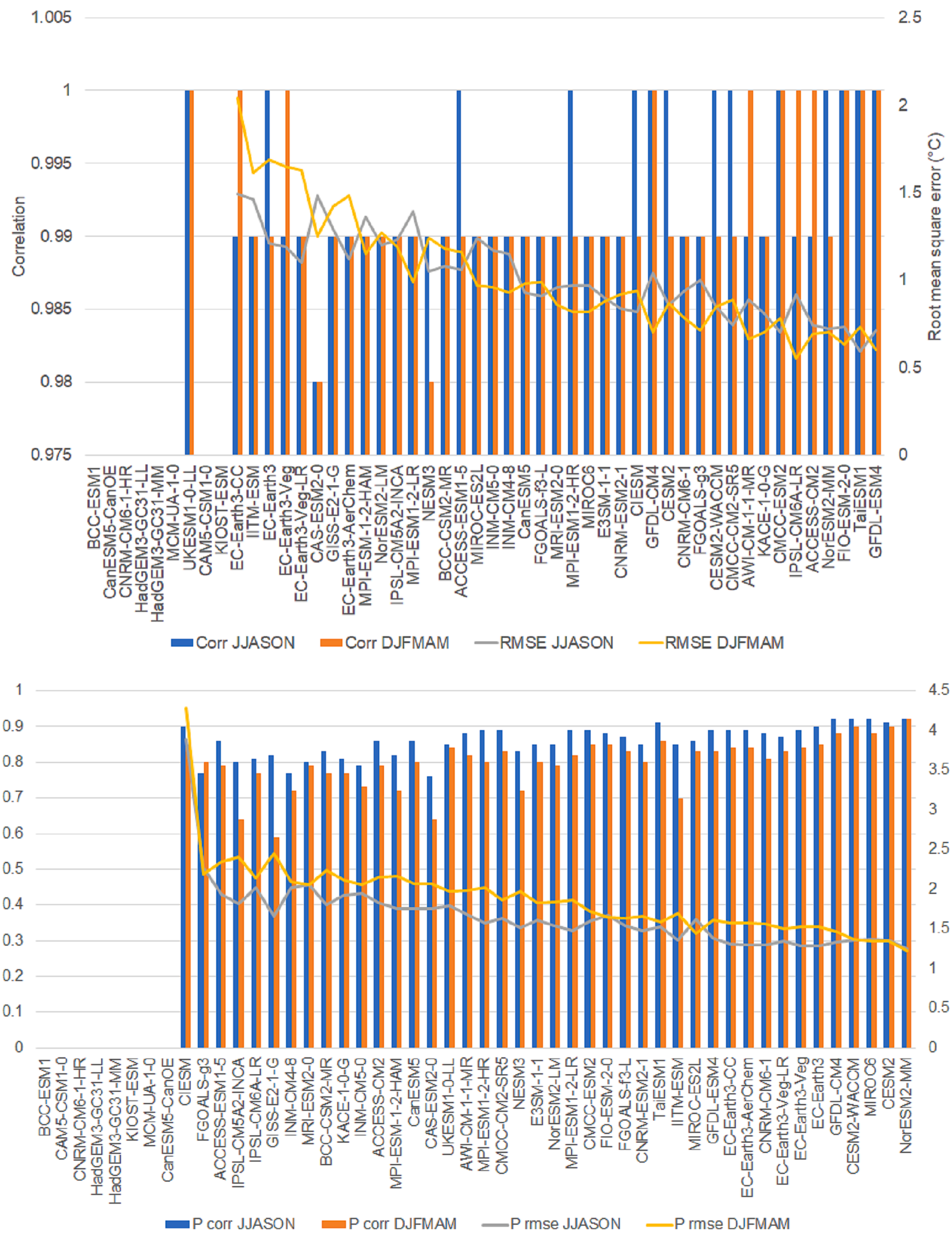


**Fig. 4.** Correlation and RMSE over the large Indo-Pacific domain (Fig. 2) in the two six-month seasons during the baseline period for TOP: surface air temperature, and BOTTOM: precipitation (an absence of bars/lines indicates data were insufficient to calculate). Models are ordered by overall performance across the four metrics, with the highest performers on the right. Note correlations are very high due to the large domain used.

Australia specifically. All scores are much higher for SAT than for PR, indicating the models' superior performance in simulating SAT. Model ranking for the M statistic of MSLP over a wider domain is broadly similar to the ranking from SAT and PR, with some scatter (Fig. 6). Some models had a lower ranking for MSLP evaluation despite high ranking for SAT (e.g., IPSL-CM5A-INCA, MRI-ESM2-0).

The ranking of evaluation of ENSO and IOD teleconnection to Australian rainfall (averaged for four regions) showed some similarity to mean climate evaluation, but some differences (Table S1). For example, some models were ranked highly for both (e.g., GFDL-ESM4, NotESM2-MM), some lower for both (e.g., INM-CM5-0, NESM3) but some models were in the top two thirds for mean climate but bottom third for this teleconnection (e.g., EC-Earth3-Veg). Results for the simulation of the subtropical jet and the storm track (Table S1), where the ranking of many models followed other scores (e.g., CAS-ESM2-0 scored poorly, the ACCESS models scored well), but with some exceptions (e.g., MRI-ESM2-0 was lower ranked). In addition, three models with a poor historical trend in SAM were flagged (CIESM, CNRM-CM6-1-HR, IPSL-CM6A-LR), and also four models with a simulation of the subtropical ridge that is outside observations (E3SM-1–1, EC-Earth3-Veg-LR, IPSL-CM5A2-INCA, MIROC6), see Table S1.

*Model rejection*

Of the tests performed in this paper, 16/50 models fall into the

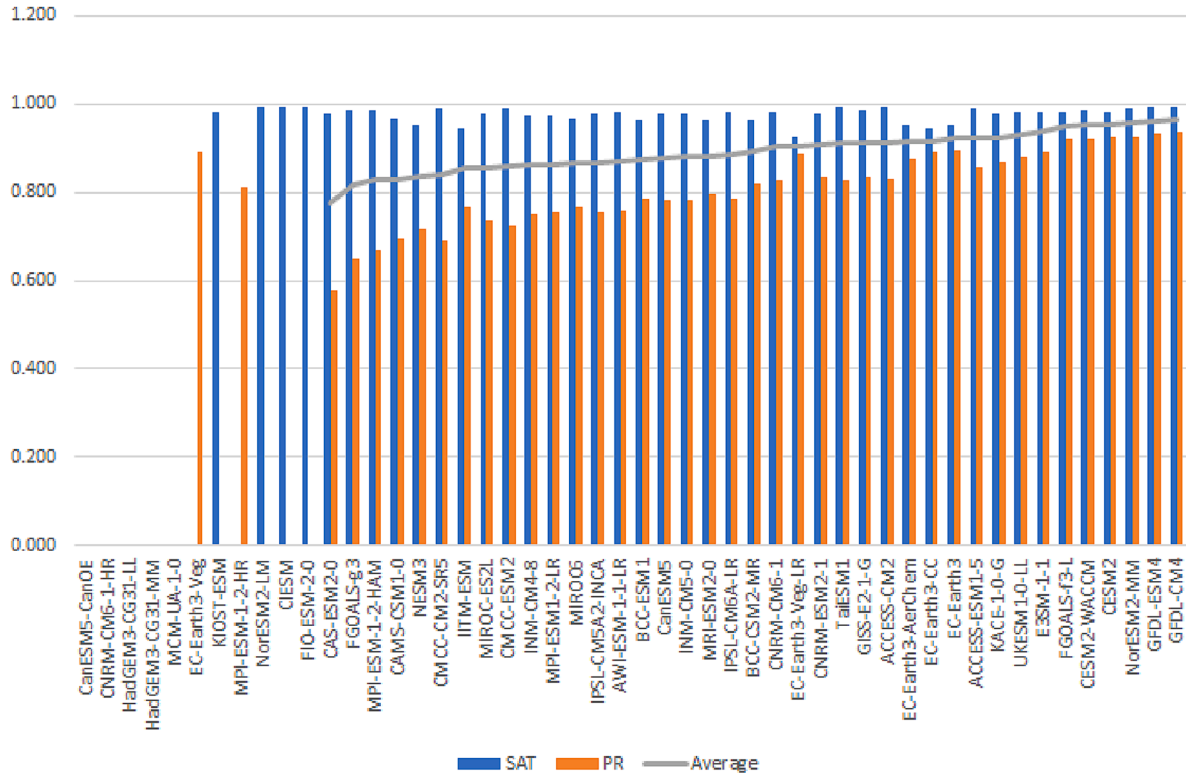


**Fig. 5.** The S scores for surface air temperatures and precipitation averaged over the Australian region (46°S-10°S, 111°E-158°E, Fig. 2), plotted in increasing order of the average score value.

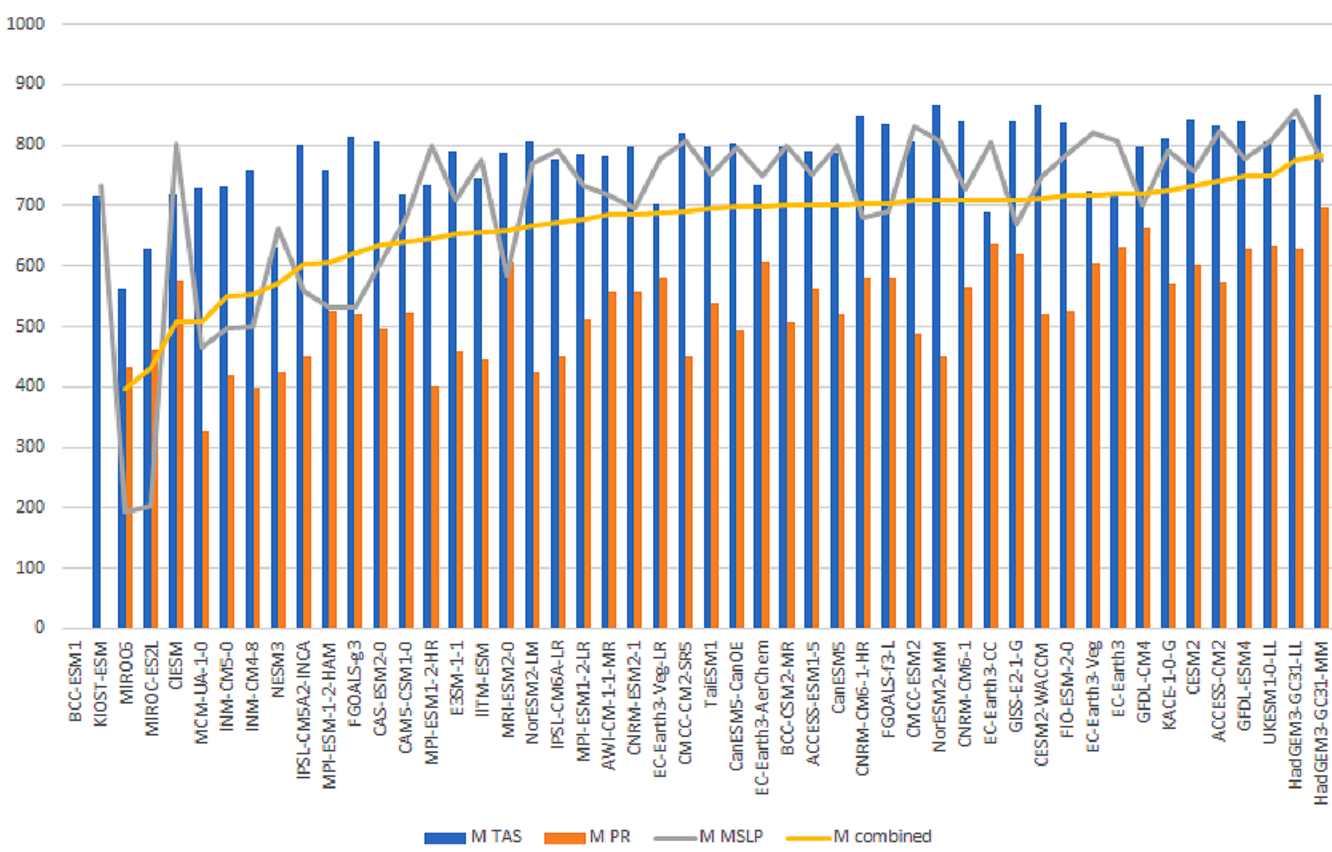


**Fig. 6.** As for Fig. 4, but showing the M Statistic for SAT, PR MSLP over the Australian region, and the average of scores.

bottom third of two or more out of four test categories for mean climate (TAS, PR and MSLP), and 16/50 fall in the bottom third of three or more out of seven test categories for circulation and driver evaluation tests or were above the chosen threshold for the test (see Table 1). Eleven models appear on both lists. Please note seven models are missing from the list due to unavailable data for some analyses. After incorporating other evidence and results from the literature, 17 out of 50 models are flagged for potential rejection, with a further 7 in the non-preferred list with at least one flag (plus some models with insufficient evaluations to assess fully).

Of the subset of 18 models with sub-daily data, seven models are rejected for use and two more are flagged as not preferred, leaving a group of up to 9 preferred models (Table 2). These groupings are found through the flags from mean climate (three models flagged), drivers and circulation (seven models flagged), then evidence from other studies and the literature. Flags from other studies come from Di Virgilio et al. (2022) and Syktus et al. (2022). Flags from other evidence include CanESM5 due to excessive past global temperature trends (Tokarska et al. 2020), MPI-ESM1-2-HR for the bias in the WPWP (Grose et al., 2020), MPI-ESM1-2-LR for SPCZ (Narsey et al., in press) as well as CanESM5 and BCC-CSM2-HR with low relative bias skill scores for heatwave metrics for Australasia specifically (Hirsch et al., 2021).

Several poor performers of mean climatology and daily distributions in SAT and PR within the Australian or sub-domains identified by Di Virgilio et al. (2022) or Syktus et al. (2022) are consistent with the findings here, including INM-CM4-8, MIROC6 and NESM3. However, some others are not consistent with this assessment, and flags are assigned to three additional models when both studies agree on poor performance: CMCC-CM2-SR5, IPSL-CM6A-LR, and NorESM2-LM. These models also scored poorly on some tests here (e.g., see Figs. 4–6), but weren't quite over the threshold set. Similarly, some models performed well in both the broader and Australian domains (e.g., GFDL-ESM4).

*Independence*

Of the 19 models with sub-daily data, 13 can be identified as falling into one of six families (Table 2). The 10 non-rejected and non-flagged models include the two EC-Earth models (EC-Earth3, EC-Earth3-Veg), two in the CESM family (CESM2 and NorESM2-MM), and two in the UK family (ACCESS-CM2 and UKESM1-0-LL). Only one from each group can be chosen as the first priority, so as to avoid models that are closely related. This results in a possibility of seven preferred models. Despite the same name, the two ACCESS models are independent above the threshold given in Brunner et al. (2020), see Mackallah et al. (2022) for details.

*Representativeness*

Restrictions of sub-daily data availability, evaluation and independence leaves a choice of up to only seven models. Representativeness is now examined to make the final selections and to illustrate the ensemble characteristics.

First the ECS value is considered. The INM-CM4-8 and INM-CM5-0 models with ECS below that are considered possible in Sherwood et al. (2021) are already rejected due to poor evaluation. Two models in the list of ten (Table 2) have an ECS above the *very likely* range (CESM2 and UKESM1-0-LL), and ACCESS-CM2 has high ECS, so downscaling of these models can be considered for the exploration of 'low probabilityhigh impact' outcomes of high warming. However, CESM2 in fact has a lower projected warming for Australia specifically than ACCESS-CM2 and other models (Fig. 7). CNRM-ESM2-1 alsohas high ECS (but within the *very likely* range) and strong warming in Australia. The ECS of the EC-Earth3 models are also above the *likely* range. NorESM2-MM has ECS at the low end of the *likely* range. Here, CESM2 and NorESM2-MM are selected to bracket the range of sensitivity, however ACCESS-CM2 and NorESM2-MM in fact bracket the range in Australian warming. UK-ESM1-0-LL is considered for a high warming case (CanESM5 is rejected).

The spread in the projected change in SAT and PR for SSP3-7.0 between 1950–1999 and 2050–2099 is broad for all Australian 'cluster' regions using all available ensemble members from CMIP6 (Fig. 7). The results are similar when considering 20-year periods 1995–2014 and 2080–2099, except there are more realisations available (EC-Earth3 simulations other than r1i1p1f1 start in 1970) and the results are noisier due to natural variability (Fig. S1).

The spread in projected change is the result of including different numbers of ensemble members from different models and the large numbers of models with high climate sensitivity (and some with very low), with equal weight as other simulations. For example, the distinct cluster of projections at the high end of the SAT range for SSP3-7.0 are the multiple realisations of CanESM5 and UKESM1-0-LL (Fig. 7), which we deliberately separate from the main selection. As mentioned above, ACCESS-CM2 is included to represent the high warming range, and NorESM2-MM the low end.

ACCESS-ESM1.5 is consistently the driest projection for every cluster region, so can be used as an illustrative example of a very dry storyline/scenario for Australia (Fig. 7). Interestingly, ACCESS-ESM1.5 is not much cooler than ACCESS-CM2, despite the difference in climate sensitivity due to regional enhancement of the warming signal related to rainfall reduction. The model was available to be re-run and sub-daily data saved for any ensemble member above r3, and we selected r6i1p1f1 as it is close to the ensemble mean change in all clusters (Fig. 7). This realisation also illustrates a dry scenario between the recent past and 2050 due to the strong drying signal together with natural variability, especially in eastern Australia (Fig. S2).

In contrast to ACCESS-ESM1.5, the two EC-Earth3 models produce a wet projection for most regions (Fig. 7). Either model could be used a representative case for a wet future, but both models can't be chosen due to independence reasons. EC-Earth3 has a more consistently wet projection than EC-Earth3-Veg through time and space, including for eastern Australian regions, so is selected in preference to the Veg version (Fig. S3). The first realisation (r1i1p1f1) of EC-Earth3 has the data required for downscaling and is in the wetter range of projected change from the 57 ensemble members (Fig. S3).

Together with the selection of four models to represent high and low warming, wet and dry projections, three other models are included in the subset (CMCC-ESM2, CNRM-ESM2-1, CESM2) that have projected changes in the mid-range of the ensemble and are not outliers for any region. This creates a selection of seven models.

As a check of how the selection represents the range from CMIP6 more broadly, we compare the projected change in these seven with the range in all models, excluding models with very high and very low ECS (shown as box plots in each panel in Fig. 7). The ranges broadly match in most instances, and the individual model points are distributed throughout the uncertainty space reasonably evenly and reach most corners. Some outlier cases are not sampled, including wet outliers in Monsoonal North and Wet Tropics. The spread in the subset is broadly representative of the whole ensemble for austral winter, after excluding models by ECS, excluding models rejected through evaluation, and averaging ensemble members from the same model first (Fig. S4). Similarly, the spread is similar for austral summer, except the subset doesn't include (and can't include) wet outlier KACE–1–0-G for Murray Basin and South and Southwestern Flatlands (Fig. S5). As for temperature change ranges, the range of ECS in the subset broadly matches CMIP6 and the group of 18 (Fig. 3), although importantly, the mean is still higher than the that from the independent assessment presented in IPCC (2021).

The subset includes models' representative of a relatively small poleward shift in the latitude of the monsoon shearline (NorESM2-MM, 0.1 °Lat under SSP5-8.5), those with a moderate poleward shift (EC-Earth3 at around 0.5 °Lat), and models with a larger shift (CMCC-ESM2,

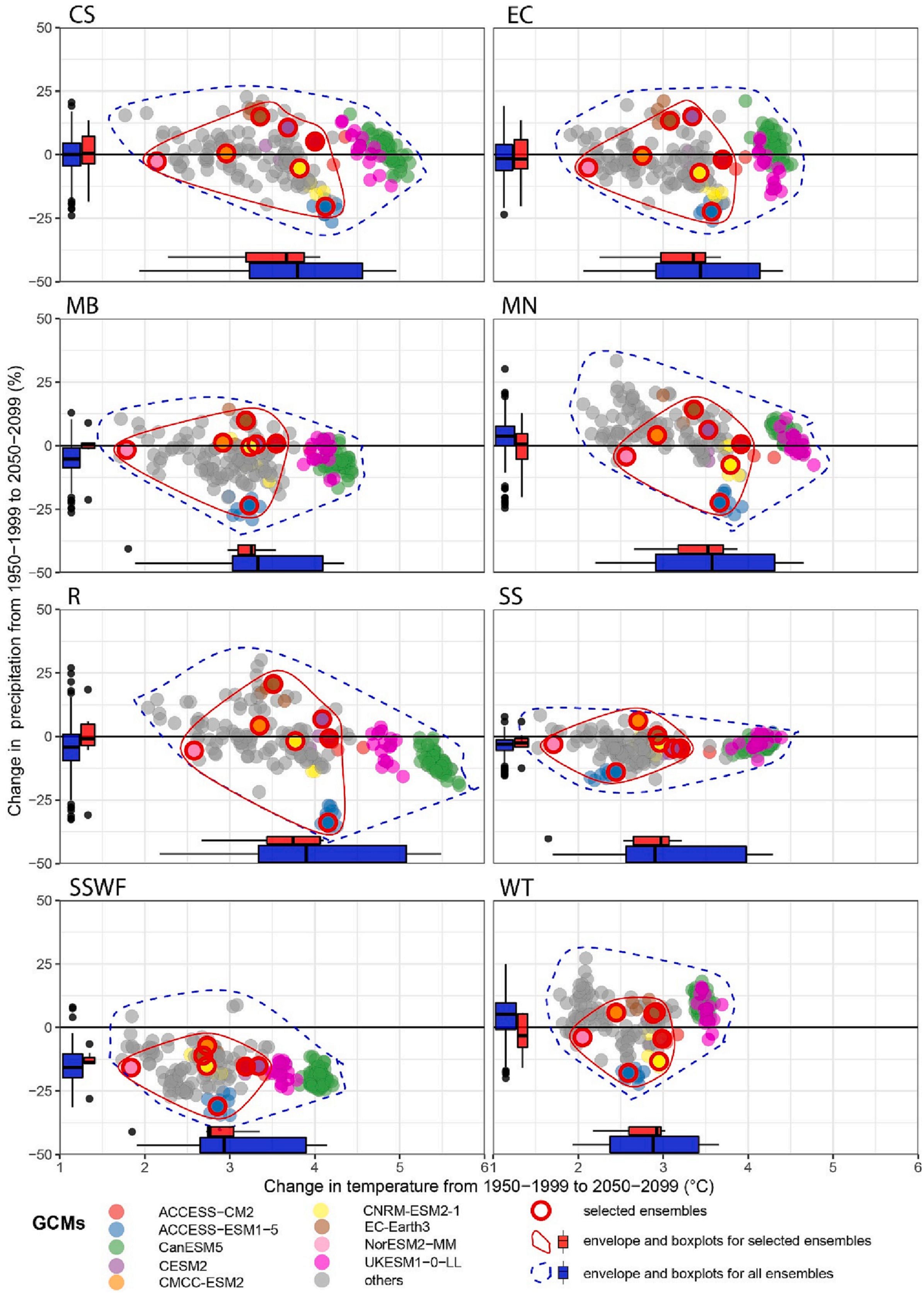


**Fig. 7.** Scatter plot of change in SAT and PR between 1950 and 2000 and 2050–2100 under SSP3-7.0 for all realisations from 35 models (grey circles), highlighting particular models and ensemble members, see legend. Markers circled in red indicate the realisation selected for each model (r4 for ACCESS-CM2, r6 for ACCESS-ESM1.5, otherwise r1). The spread of projected change from the whole ensemble and for the subset is as a boundary plot and boxplots below and to the left of the boundary plot show the mean and the 25–75% range and outliers in each model group.

around 1.5 °Lat). Models with larger poleward shifts than this, at >2 °Lat, are all previously rejected (INM-CM5, TaiESM1, CIESM), and no model shows an equator-ward shift. Similarly, the subset contains representatives of different magnitudes of poleward movement and strengthening of the Subtropical ridge, including models with a weaker projected strengthening of the ridge than the model mean (ACCESS-ESM1.5), strengthening near the model mean but a larger poleward movement (EC-Earth3) a stronger strengthening and poleward response than most models (NorESM2-MM), and those in the mid-range for both (ACCESS-CM2, CMCC-ESM2).

Further models within the group with sub-daily data available that could be downscaled include those with high climate sensitivity for an examination of 'low probability high impact' outcomes of high warming (CESM2, UKESM-1-0-LL), noting that these are not independent of those selected. Similarly, EC-Earth3-Veg could be downscaled as a wet case but is not independent from EC-Earth3. If the RCM doesn't require sub-daily data, then any models not rejected could be used, such as FGOALS-g3, GFDL-ESM4, CNRM-CM6-1-HR and KACE1-0-G (Table S1). Syktus et al. (2022) has selected some of these models for use in CORDEX Australasia (see Discussion).

The subset of models is also broadly representative of the projected change in RX1day rainfall relative to the change in the mean. In a subset of 17 CMIP6 models, virtually all show a projected change in RX1day rainfall that is more positive (wetter, or else less dry) than the change in the mean for all seven cluster regions (results not shown). This is broadly consistent with what could be expected from the thermodynamic

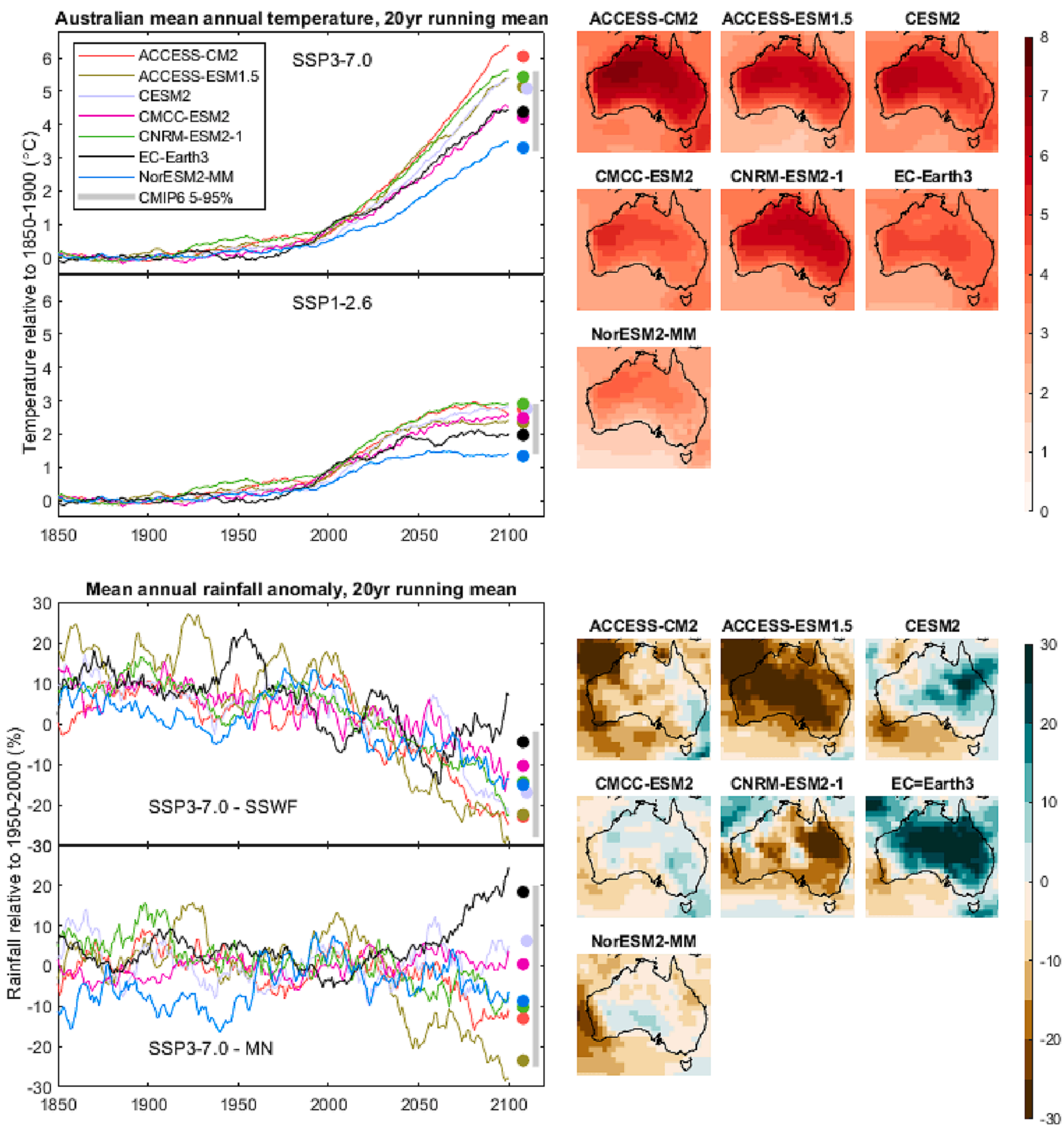


**Fig. 8.** Time series and map plots of change in the seven selected realisations from CMIP6, TOP: change in surface air temperature for Australia relative to 1850–1900, lines show 20-year running mean, filled dots and maps show the change between 1850 and 1900 and 2080–2099, grey bar shows the 5–95% range of all CMIP6 models. BOTTOM: as for top panels but showing change in mean annual precipitation relative to 1950–2000 (%), and line plots are given for two example 'cluster' regions of Australia: SSWF = Southern and Southwestern Flatlands, and MN = Monsoonal North.

response to warming along with changes to dynamic aspects such as the intensity or frequency of weather systems. Four of the selected models are part of this 17-member subset, and broadly represent the change in the 17, including smaller and larger enhancements (e.g., CNRCM-ESM2-1 and EC-Earth3 respectively in Monsoonal North), and containing rare exceptions where RX1day is more negative than the change in the mean (e.g., EC-Earth3 in the East Coast).

*The ensemble as a set of future 'storylines'*

This group of seven selected individual model realisations used as input to CORDEX Australasia are generally representative of the CMIP6 ensemble in terms of SAT and PR change in Australia, once rejected models and outliers in ECS are excluded (Fig. 7). This means their projections can be used in typical projections presentations plots such as plume plots and map plots with confidence they are broadly representative of the long-term projected trends in the 10–90% range from CMIP6 (Fig. 8). This includes a spread that covers the projected temperature range for Australia, the range of projected drying in the southwest (SSWF cluster) as well as the range of projected change in rainfall from significant decrease to increase for the north (the MN cluster). However, the subset doesn't correct or otherwise improve the ranges from CMIP6, such as the uneven spread of ECS (Fig. 3).

However, a subset of seven model simulations can't be a balanced and detailed sampling the probability distribution of the full ensemble of over 100 members. But the subset can be used to representatively sample the range of future climate changes in the ensemble, when thought of in broad categories such as wetter and drier. In this way, the seven simulations can be used to generate key climate futures that broadly encompass the range of possibilities, using some principles taken from a storyline approach (Shepherd et al. 2018), or the 'climate futures' approach (Whetton et al., 2012). These different physically plausible futures in temperature and rainfall (for a given SSP) can be summarised by short narratives (Table 3). The addition of the 'low probability high impact' outcome from models with ECS above the *very likely* range (also with very high Australian warming) could be thought of as another storyline.

Once the range of physical drivers behind the projected change in each model are understood in the host GCMs, these can be added as part of the narratives for each storyline. Also, once the effect of the regional modelling step is clear, including any 'added value' from the RCMs, this can also inform these narratives.

## Discussion

A large MME can potentially produce statistically balanced uncertainty assessments, but this at least requires a large ensemble size, such as the 78 EURO-CORDEX simulations analysed in Evin et al. (2021). It may not be appropriate to use a smaller MME, or arguably even a large ensemble, in this way. Instead, a limited modelling program may be thought of as illustrating details of plausible future climates using a storyline approach (Shepherd et al., 2018) rather than truly statistically balanced estimates. The storyline approach presents physically self-consistent future scenarios that can be used for event-based risk management. Storylines aren't given with formal probabilities and frequentist statistics of the ensemble (such as mean and standard deviation of the ensemble spread), but nevertheless provide a useful framework for decision making given the context.

There are several *a priori* limitations on this Australasian MME that are barriers to a balanced ensemble, including the limited GCM choice due to a lack of sub-daily data and the limited number of RCM simulations due to high costs. This means that the storyline approach is taken as the default, out of necessity. In terms of GCM selection, this means reducing the GCM ensemble to its most salient narratives, rather than aiming for a full statistical spread. Once RCM simulations are produced, these can then form the basis of the projections and input to storylines but are dependent on the GCM selection as part of a cascade of processes. So, while we can aim to make the outputs as balanced as practically possible, they will not sample the full range of epistemic uncertainty and won't be a probabilistic sample of future uncertainty. This problem is not specific to this program and is true of any limited ensemble. However, there is real explanatory power and utility in the storyline approach for the uses of climate projections and climate services, so this position is not in fact a poor one. The MME can use good practice such as rejecting unsuitable models, avoiding using models that are highly dependent, and representatively sampling the independently assessed uncertainty space of interest, within practical limitations.

A notable issue when using CMIP6 is the uneven spread of climate sensitivity, and therefore average warming amounts for the globe and for Australasia. For mean warming estimates in IPCC (2021), models were weighted and combined with an assessment of climate sensitivity and historical warming and the use of emulators, which appears warranted for regional projections too. The group of 18 models with sub-daily data available also features this problem (Fig. 3). Despite deliberately selecting models with a range of climate sensitivity and limiting the choice of models with ECS above the likely range, the spread of the selected set of seven highest priority models is still similar to CMIP6 as a whole and doesn't match the likely range from IPCC (2021) and Sherwood et al. (2020), see Fig. 3. This confirms that CMIP6, and this subset of CMIP6 models, shouldn't be used with equal weighting to produce a balanced estimate of mean warming. Rather, weighting or other techniques should be used. Alternatively, models can be used in a Global Warming Levels framework, where changes at particular levels of global warming since the pre-industrial era such as 1.5, 2, 3 and 4 °C are presented by 'time sampling' the simulations when they reach that level, which standardises for climate sensitivity, see presentation in the IPCC Atlas chapter by Gutiérrez et al. (2021). However, even without weighting or by using global warming levels, the models can be used in a storyline framework, illustrating plausible physically consistent future climates but without using frequentist statistics (e.g., Shepherd, 2021).

Some coordination and inter-comparability are very useful in an MME but competing interests and priorities from different modelling groups create practical barriers to central control or coordination. Also, there is value in diversity and advantages in using a 'sparse matrix' approach rather than a filled matrix. As part of the selection, the models assessed in the analysis for this paper were compared to the other two independent exercises in Di Virgilio et al. (2022) and Syktus et al. (2022). The studies made different choices regarding the spatial domain and the balance of evaluation metrics to use. Several models were identified as poor performers for this purpose and region regardless of the evaluation choices, including CIESM and IPSL-CM6A-LR. Some other models were rejected by the other studies but not this analysis, such as CMCC-CM2-SR5 (however, it was near the threshold for rejection). This

**Table 3**
Narrative descriptions of the projection from the seven selected models for Australia under SSP3-7.0 in terms of temperature and rainfall, also shown is the special case provided by UKESM1-0-LL. SWWA refers to Southwest Western Australia.

| Model | Description |
|---|---|
| ACCESS-CM2 | A much hotter future, and drier in most regions except the southeast |
| ACCESS-ESM1.5 | A hotter and much drier future |
| CESM2 | A hotter future, wetter in parts of the east and north |
| CMCC-ESM2 | A much warmer future with little change in mean rainfall (with regional exceptions) |
| CNRM-ESM2-1 | A much hotter future, much drier especially in the east, but wetter in the northwest |
| EC-Earth3 (or EC-Earth3-Veg) | A hotter and much wetter future for much of Australia (except SWWA) |
| NorESM2-MM | Lower warming, mid-range changes in rainfall |
| UKESM1-0-LL | Low probability, high impact case (high climate sensitivity, high Australian warming) |

suggests that, in part, the large-scale climatology of the Indo-Pacific region in terms of SAT and PR is acceptable, but that simulation within the Australasian domain is poorer. The differences are worth further investigation.

The other two independent exercises in Di Virgilio et al. (2022) and Syktus et al. (2022) have been completed in parallel and these results can be compared here to assess the status of the CORDEX Australasia sparse matrix from all Australian contributors (Table 4). There are 15 potential GCMs and 3 RCMs used, with 6 RCM configurations (CCAM-ACS, CCAM-Qld, CCAM-Coupled, BARPA, and two quite different configurations of WRF used in NARCLIM2). The generally dry case (ACCESS-ESM1.5), the generally wet case (EC-Earth3, or alternatively EC-Earth3-Veg) and the cool case (NorESM2-MM) are downscaled by all RCMs, providing a useful comparison for this dimension of RCM uncertainty, including identifying added value that is robust to the models used. A high warming case (ACCESS-CM2) is included from three models. Four GCMs that don't provide sub-daily data and so can only be downscaled by Syktus et al. (2022), meaning these models won't be part of a strict sparse matrix approach unless an additional RCM system is added, but provide useful alternative samples. Another dimension to the matrix is that Syktus et al. (2022) will run coupled RCM simulations for five simulations from four models. Two of these models will also be downscaled using the non-coupled version of the same model, and other RCMs as well (Chapman et al., submitted). This means the effect of different RCMs, and of coupling in the RCM, can be examined and distinguished. The 'low probability, high impact' case represented by UKESM1-0-LL, with climate sensitivity above the *very likely* range and high Australian warming, will be downscaled by one RCM, and will provide a useful illustration of this possible future.

By applying this selection strategy, the results will differ from other strategies in key respects. By applying an independence condition, considering climate sensitivity and considering representativeness for Australian climate change, the model list differs from the ranking of models by evaluation of the simulation of Australian climate. For example, if seven models were simply chosen by ranking of the combined M statistic (incorporating temperature, precipitation and surface pressure) then: 1) closely related models would be chosen (e.g., EC-Earth3 and EC-Earth3-Veg), 2) three of seven models would have high climate sensitivity (UKESM1-0-LL, ACCESS-CM2, CESM2), and the selection would not include some key 'Climate futures', such as the dry future in ACCESS-ESM1.5 (ranks at number eight). Also, if representativeness for a region other than Australia were considered (e.g., global changes, southern hemisphere, southeast Asia, Australasia as a whole) the results would also differ. Also, the results are likely to depend on the variables included in the representativeness assessment. Here we focussed on projected change in temperature and rainfall, with a basic check on the projection of major circulation changes, which likely yields different results than if the sub-setting included relative humidity, radiation and windspeed following Hayashi and Shiogama (2022).

The current strategy doesn't represent an idealised sparse matrix based on theory, so doesn't have a balanced number of RCMs downscaling a planned set of GCMs. However, it has many strong features that allow inter-comparison of different RCMs and analysis of added value. It also features good attributes in terms of evaluation, independence and

**Table 4**
The current status of the CORDEX-Australasia sparse matrix from participating Australian groups, with the GCM noted in rows and the RCM in columns, and the specific realisation used is noted in each cell. Regional climate modelling efforts are as follows: CCAM-Qld uses the CCAM model in non-nudged mode with bias correction of inputs and including some coupled simulations; NARCLIM2.0 is the third generation of the New South Wales and Australian Regional Climate Modelling project using two configurations of the WRF model (different physics parameterisations); CCAM is using the CCAM model in nudged mode without bias correction of inputs; BARPA is the Bureau of Meteorology Atmospheric Regional Projections for Australia. The orange highlight marks models with no sub-daily outputs available, so can only be used by CCAM-Qld, highlighted blue simulations are RCM run in ocean–atmosphere coupled mode.

| | CCAM-Qld | NARCLIM2.0 (2x WRF configurations) | CCAM | BARPA |
|---|---|---|---|---|
| **ACCESS-CM2** | r2i1p1f1oc | | r4i1p1f1 | r4i1p1f1 |
| **ACCESS-ESM1.5** | r6i1p1f1<br>r20i1p1f1oc<br>r40i1p1f1oc | r6i1p1f1 | r6i1p1f1 | r6i1p1f1 |
| **CESM2** | | | r11i1p1f1 | r11i1p1f1 |
| **CMCC-ESM2** | r1i1p1f1 | | r1i1p1f1 | r1i1p1f1 |
| **CNRM-CM6.1-HR** | r1i1p1f2<br>r1i1p1f2oc | | | |
| **CNRM-ESM2-1** | | | r1i1p1f2 | |
| **EC-Earth3** | r1i1p1f1 | | r1i1p1f1 | r1i1p1f1 |
| **EC-Earth3-Veg** | | r1i1p1f1 | | |
| **FGOALS-g3** | r4i1p1f1 | | | |
| **GFDL-ESM4** | r1i1p1f1 | | | |
| **GISS-E2-1-G** | r2i1p1f2 | | | |
| **MPI-ESM1-2-HR** | | r1i1p1f1 | | |
| **MPI-ESM1-2-LR** | r9i1p1f1 | | | |
| **MRI-ESM2-0** | r1i1p1f1 | | | |
| **NorESM2-MM** | r1i1p1f1<br>r1i1p1f1oc | r1i1p1f1 | r1i1p1f1 | r1i1p1f1 |
| **UKESM1-0-LL** | | r1i1p1f1 | | |

representativeness.

The results of this ensemble will be useful as a core data source but must always be placed in a larger context given the constraints involved. In particular, the use of two SSPs that bracket a range of plausible socio-economic developments (SSP1-2.6 and SSP3-7.0) is useful but limited, and the fuller range of plausible pathways need to be acknowledged, including a worst case (SSP5-8.5), a pathway consistent with 1.5 °C global warming (SSP1-1.9) and those outside of the Tier 1 SSPs including overshoot, carbon dioxide removal and more. The Global Warming Levels framework (time sampling regional climate as the world reaches 1.5, 2, 3 and 4 °C global warming) is a useful tool to standardise between different possibilities.

Also, the results should be put in the wider context of other models, including the host models themselves, other CMIP6 models, insights from running large ensembles of GCMs, further 'convective permitting' modelling to add further insights on regional climate change, as well as emulators and simple models. In addition, tools such as deep learning and machine learning (e.g., Wang et al., 2022) are now mature and should be used to provide additional insights.

## Conclusions

The future-focussed component of climate services requires a core data source of credible and locally relevant climate model simulations. While no modelling framework is perfect, the strategic use of downscaling using Regional Climate Models (RCMs) from selected Global Climate Models (GCMs) for a set of Shared Socio-economic Pathways (SSPs) in a coordinated program has several advantages. Such an ensemble uses a recognised and well-established source of global modelling, then provides regional scale 'added value' on top of this and also accounts for the effect of regionalisation in the 'cascade of uncertainty'.

Here we have described the strategy for producing a model ensemble of RCM simulations to provide climate services in Australia, with a focus on the appropriate selection of GCMs and structuring of the RCM ensemble. Climate simulations from a set of up to 15 global climate model simulations (seven selected here) downscaled by four or more regional models for two future scenarios (a very low and a high SSP) under an *ad hoc* 'sparse matrix' is proposed. The projections cannot be considered a probabilistic or balanced estimate of uncertainty given the limited ensemble size and underlying epistemic uncertainties. The ensemble can however be used in a 'climate futures' or 'storyline' approach to illustrate plausible future climates that broadly span the range of possibilities suggested by CMIP6, while producing added value at the regional scale. In this way, the ensemble can form a useful tool for informing climate change adaptation planning and motivating emissions mitigation. The seven GCMs proposed for selection here all simulate the current climate to an acceptable level, are relatively independent, and are representative of the projected range of change in temperature and rainfall over the 21st Century in CMIP6.

The strategic sampling of plausible future climates using selected GCMs that are then downscaled using RCMs in a coordinated program is a useful core data source for the current generation of climate services. However, the data should of course be consistent with the assessment of climate change from multiple lines of evidence. The results of this modelling are for two SSPs only, so should be put in wider context of other possibilities of SSPs and other potential pathways. Also, the results should be put in the wider context of other models, including CMIP6, large ensembles of GCMs, 'convective permitting' modelling, emulators, deep learning and machine learning and more.

## CRediT authorship contribution statement

**Michael R. Grose:** Conceptualization, Formal analysis, Writing – original draft. **Sugata Narsey:** Conceptualization, Data analysis, Writing – original draft. **Ralph Trancoso:** Formal analysis, Visualization. **Chloe Mackallah:** Data curation, Formal analysis. **Francois Delage:** Formal analysis, Visualization. **Andrew Dowdy:** Formal analysis, Writing – original draft. **Giovanni Di Virgilio:** Conceptualization, Writing – original draft. **Ian Watterson:** Formal analysis. **Peter Dobrohotoff:** Formal analysis. **Harun A. Rashid:** Formal analysis. **Surendra Rauniyar:** Formal analysis. **Ben Henley:** Formal analysis. **Marcus Thatcher:** Conceptualization, Writing – original draft. **Jozef Syktus:** Conceptualization. **Gab Abramowitz:** Conceptualization. **Jason P. Evans:** Writing – original draft. **Chun-Hsu Su:** Writing – original draft. **Alicia Takbash:** Data curation, Formal analysis.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

*Acknowledgements*

This work was supported by the Australian Climate Service (ACS). Valuable and constructive comments by the internal reviewers and anonymous external reviewers are greatly acknowledged.

NARCliM2.0 is supported by the New South Wales (NSW) Department of Planning and Environment with funding provided by the NSW Climate Change Fund.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cliser.2023.100368.

## References

Abramowitz, G., Herger, N., Gutmann, E., Hammerling, D., Knutti, R., Leduc, M., et al., 2019. Model dependence in multi-model climate ensembles: weighting, sub-selection and out-of-sample testing. Earth Syst. Dynam. 10 (1), 91–105. https://doi.org/10.5194/esd-10-91-2019.

Adler, R., Sapiano, M., Huffman, G., Wang, J.-J., Gu, G., Bolvin, D., Chiu, L., Schneider, U., Becker, A., Nelkin, E., Xie, P., Ferraro, R., Shin, D.-B., 2017. Global precipitation. Atmosphere 9 (4), 138.

Brunner, L., Pendergrass, A.G., Lehner, F., Merrifield, A.L., Lorenz, R., Knutti, R., 2020. Reduced global warming from CMIP6 projections when weighting models by performance and independence. Earth Syst. Dynam. 11, 995–1012.

Chapman S., Syktus J., Trancoso R., Thatcher M., Toombs N., Wong K.K. and Takbash A., (2023) Evaluation of dynamically downscaled CMIP6 models over Australia. Earth's Future (submitted).

Ciarlo, J.M., Coppola, E., Fantini, A., Giorgi, F., Gao, X., Tong, Y., Glazer, R.H., Alavez, J. A.T., Sines, T., Pichelli, E., Raffaele, F., 2020. A new spatially distributed added value index for regional climate models: the EURO-CORDEX and the CORDEX-CORE highest resolution ensembles. Climate Dyn. 57, 1403–1424.

Colman, R.A., Moise, A.F., Hanson, L.I., 2011. Tropical Australian climate and the Australian monsoon as simulated by 23 CMIP3 models. J. Geophys. Res.: Atmos. 116 (D10).

Su, C-H., Stassen, C., Howard, E., Ye, H., Bell, S., Pepler A., Dowdy, A.J., Tucker, S.O., Franklin, C. (2022), BARPA: New development of ACCESS-based regional climate modelling for Australian Climate Service, Bureau Research Report No. 069, ISBN: 978-1-925738-57-5, http://www.bom.gov.au/research/publications/researchreports/BRR-069.pdf.

CORDEX, 2021. CORDEX experiment design for dynamical downscaling of CMIP6. Technical report, May 2021: https://cordex.org/2021/05/24/experiment-protocol-rcms-is-published/.

CSIRO and Bureau of Meteorology (2015). Climate Change in Australia, Technical Report. Melbourne Australia. www.climatechangeinaustralia.gov.au.

Di Luca, A., Elía, R., Laprise, R., 2013. Potential for small scale added value of RCM's downscaled climate change signal. Climate Dyn. 40, 601–618.

Di Luca, A., Argüeso, D., Sherwood, S., Evans, J.P., 2021. Evaluating precipitation errors using the environmentally conditioned intensity-frequency decomposition method. J. Adv. Modeling Earth Systems, 13(7), e2020MS002447. doi: 10.1029/2020MS002447.

Di Virgilio, G., Evans, J.P., Di Luca, A., Grose, M.R., Round, V., Thatcher, M., 2020. Realised added value in dynamical downscaling of Australian climate change. Climate Dynamics 54 (11-12), 4675–4692.
Di Virgilio, G., Ji, F., Tam, E., Nishant, N., Evans, J.P., Thomas, C., Riley, M., Beyer, K., Grose, M., Narsey, S., Delage, F., 2022. Selecting CMIP6 GCMs for CORDEX dynamical downscaling: model performance, independence, and climate change signals. Earth's Future n/a(n/a): e2021EF002625.
Evans, A., Jones, D., Smalley, R., Lellyett, S., 2020. An enhanced gridded rainfall analysis scheme for Australia. Bureau Research Report 41. ISBN: 978-1-925738-12-4.
Evans, J.P., Ji, F., Lee, C., Smith, P., Argüeso, D., Fita, L., 2014. Design of a regional climate modelling projection ensemble experiment – NARCliM. Geosci. Model Dev. 7 (2), 621–629. https://doi.org/10.5194/gmd-7-621-2014.
Evin, G., Somot, S., Hingray, B., 2021. Balanced estimate and uncertainty assessment of European climate change using the large EURO-CORDEX regional climate model ensemble. Earth Syst. Dynam. Discuss. 2021, 1–40.
Eyring, V., Bony, S., Meehl, G.A., Senior, C.A., Stevens, B., Stouffer, R.J., Taylor, K.E., 2016. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geosci. Model Dev. 9, 1937–1958.
Grose, M.R., Timbal, B., Wilson, L., Bathols, J., Kent, D., 2015. The subtropical ridge in CMIP5 models, and implications for projections of rainfall in southeast Australia. Australian Meteorol. Oceanographic J. 65 (1), 90–106.
Grose, M.R., Risbey, J.S., Moise, A.F., Osbrough, S., Heady, C., Wilson, L., Erwin, T., 2017. Constraints on Southern Australian rainfall change based on atmospheric circulation in CMIP5 simulations. J. Climate 30 (1), 225–242.
Grose, M.R., Narsey, S., Delage, F.P., Dowdy, A.J., Bador, M., Boschat, G., Chung, C., Kajtar, J.B., Rauniyar, S., Freund, M.B., Lyu, K., Rashid, H., Zhang, X., Wales, S., Trenham, C., Holbrook, N.J., Cowan, T., Alexander, L., Arblaster, J.M., Power, S., 2020. Insights from CMIP6 for Australia's Future Climate. Earth's Future 8 (5) e2019EF001469.
Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol. 377 (1-2), 80–91.
Gutiérrez, J.M., R.G. Jones, G.T. Narisma, L.M. Alves, M. Amjad, I.V. Gorodetskaya, M. Grose, N.A.B. Klutse, S. Krakovska, J. Li, D. Martínez-Castro, L.O. Mearns, S.H. Mernild, T. Ngo-Duc, B. van den Hurk, and J.-H. Yoon, (2021). Atlas. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press.
Hausfather, Z., Marvel, K.D., Schmidt, G.A., Nielson-Gammon, J.W., Zelinka, M.D., 2022. Climate simulations: recognize the 'hot model' problem. Nature 605, 26–29. https://doi.org/10.1038/d41586-022-01192-2.
Hayashi, M., Shiogama, H., 2022. Assessment of CMIP6-Based Future Climate Projections Selected for Impact Studies in Japan. SOLA 18, 96–103.
Herger, N., Abramowitz, G., Knutti, R., Angélil, O., Lehmann, K., Sanderson, B.M., 2018. Selecting a climate model subset to optimise key ensemble properties. Earth Syst. Dynam. 9, 135–151. https://doi.org/10.5194/esd-9-135-2018.
Hersbach, H., et al., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Society 146 (730), 1999–2049.
Hirsch, A. L., N. N. Ridder, S. E. Perkins-Kirkpatrick, A. Ukkola, 2021. CMIP6 MultiModel Evaluation of Present-Day Heatwave Attributes. Geophysical Research Letters 48 (22): e2021GL095161.
IPCC, 2021. Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. In: MASSON-DELMOTTE, V., ZHAI, P., PIRANI, A., CONNORS, S. L., PÉAN, C., BERGER, S., CAUD, N., CHEN, Y., GOLDFARB, L., GOMIS, M. I., HUANG, M., LEITZELL, K., LONNOY, E., MATTHEWS, J. B. R., MAYCOCK, T. K., WATERFIELD, T., YELEKÇI, O., YU, R. & ZHOU, B. (eds.). Cambridge University Press.
Iturbide, M., Gutiérrez, J.M., Alves, L.M., Bedia, J., Cerezo-Mota, R., Cimadevilla, E., Cofiño, A.S., Di Luca, A., Faria, S.H., Gorodetskaya, I.V., Hauser, M., Herrera, S., Hennessy, K., Hewitt, H.T., Jones, R.G., Krakovska, S., Manzanas, R., Martínez-Castro, D., Narisma, G.T., Nurhati, I.S., Pinto, I., Seneviratne, S.I., van den Hurk, B., Vera, C.S., 2020. An update of IPCC climate reference regions for subcontinental analysis of climate model data: definition and aggregated datasets. Earth Syst. Sci. Data 12 (4), 2959–2970.
Jones, D.A., Wang, W., Fawcett, R., 2009. High-quality spatial climate data-sets for Australia. Australian Meteorol. Oceanographic J. 58 (4), 233–248.
Kim, Y.-H., Min, S.-K., Zhang, X., Sillmann, J., Sandstad, M., 2020. Evaluation of the CMIP6 multi-model ensemble for climate extreme indices. Weather and Climate Extremes 29, 100269.
Kotamarthi, R., Hayhoe, K., Mearns, L.O., Wuebbles, D., Jacobs, J., Jurado, J., 2021. Downscaling Techniques for High-Resolution Climate Projections: From Global Change to Local Impacts. Cambridge University Press, Cambridge.
Leduc, M., Laprise, R., De Elía, R., Separovic, L., 2016. Is institutional democracy a good proxy for model independence? J. Climate 29, 8301–8316. https://doi.org/10.1175/JCLI-D-15-0761.1.
Lloyd, E.A., Bukovsky, M., Mearns, L.O., 2021. An analysis of the disagreement about added value by regional climate models. Synthese 198 (12), 11645–11672.
Mackallah, C., Chamberlain, M.A., Law, R.M., Dix, M., Ziehn, T., Bi, D., Bodman, R., Brown, J.R., Dobrohotoff, P., Druken, K., Evans, B., Harman, I.N., Hayashida, H., Holmes, R., Kiss, A.E., Lenton, A., Liu, Y., Marsland, S., Meissner, K., Menviel, L., O'Farrell, S., Rashid, H.A., Ridzwan, S., Savita, A., Srbinovsky, J., Sullivan, A., Trenham, C., Vohralik, P.F., Wang, Y.-P., Williams, G., Woodhouse, M.T., Yeung, N., 2022. ACCESS datasets for CMIP6: methodology and idealised experiments. J. Southern Hemisphere Earth Systems Sci. 72 (2), 93–116.
McGinnis, S., Mearns, L., 2021. Building a climate service for North America based on the NA-CORDEX data archive. Climate Services 22, 100233.
Mearns, L.O., Arritt, R., Biner, S., Bukovsky, M.S., McGinnis, S., Sain, S., Caya, D., Correia, J., Flory, D., Gutowski, W., Takle, E.S., Jones, R., Leung, R., Moufouma-Okia, W., McDaniel, L., Nunes, A.M.B., Qian, Y., Roads, J., Sloan, L., Snyder, M., 2012. The North American regional climate change assessment program overview of phase i results. Bull. Am. Meteorol. Society 93 (9), 1337–1362.
Moalafhi, D.B., Evans, J.P., Sharma, A., 2016. Evaluating global reanalysis datasets for provision of boundary conditions in regional climate modelling. Climate Dyn. 47 (9–10), 2727–2745. https://doi.org/10.1007/s00382-016-2994-x.
Narsey, S. Y., J. R. Brown, R. A. Colman, F. Delage, S. B. Power, A. F. Moise, H. Zhang, 2020. Climate Change Projections for the Australian Monsoon from CMIP6 Models. Geophysical Research Letters 47(13): e2019GL086816.
Narsey, S., J. R. Brown, F. Delage, G. Boschat, M. R. Grose, R. Colman and S. B. Power (in press). Storylines of South Pacific Convergence Zone changes in a warmer world. J. Climate.
Pennell, C., Reichler, T., 2011. On the effective number of climate models. J. Climate 24 (9), 2358–2367.
Rayner, N.A., Parker, D.E., Horton, E.B., Folland, C.K., Alexander, L.V., Rowell, D.P., Kent, E.C., Kaplan, A., 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature 827 since the late nineteenth century. Journal of Geophysical Research: Atmospheres 108, 4407.
Riahi, K., et al., 2017. The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: an overview. Global Environmental Change 42, 153–168.
Rupp, D.E., Abatzoglou, J.T., Hegewisch, K.C., Mote, P.W., 2013. Evaluation of CMIP5 20th century climate simulations for the Pacific Northwest USA, J. Geophys. Res.: Atmos., 118, 10,884-810,906, 10.1002/jgrd.50843.
Seland, Ø., Bentsen, M., Olivié, D., Toniazzo, T., Gjermundsen, A., Graff, L.S., Debernard, J.B., Gupta, A.K., He, Y.C., Kirkevåg, A., Schwinger, J., Tjiputra, J., Aas, K.S., Bethke, I., Fan, Y., Griesfeller, J., Grini, A., Guo, C., Ilicak, M., Karset, I.H. H., Landgren, O., Liakka, J., Moseid, K.O., Nummelin, A., Spensberger, C., Tang, H., Zhang, Z., Heinze, C., Iversen, T., Schulz, M., 2020. Overview of the Norwegian Earth System Model (NorESM2) and key climate response of CMIP6 DECK, historical, and scenario simulations. Geosci. Model Dev. 13 (12), 6165–6200.
Shepherd, T.G., et al., 2018. Storylines: an alternative approach to representing uncertainty in physical aspects of climate change. Climatic Change 151 (3), 555–571.
Shepherd, T.G., 2021. Bringing physical reasoning into statistical practice in climate-change science. Climatic Change 169 (1), 2.
Sherwood S.C, et al., 2020. An assessment of Earth's climate sensitivity using multiple lines of evidence. Rev. Geophys., 58, e2019RG000678.
Sillmann, J., Kharin, V.V., Zhang, X., Zwiers, F.W., Bronaugh, D., 2013. Climate extremes indices in the CMIP5 multi-model ensemble. Part 1: Model evaluation in the present climate. J. Geophys. Res. Atmos. 118, 1716–1733. https://doi.org/10.1002/jgrd.50203.
Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Barker, D.M., Wang, W., Powers, J. G., 2008. A description of the Advanced Research WRF Version 3. NCAR Tech Note NCAR/TN-475+STR. NCAR. Boulder, CO.
Srikanthan, R., McMahon, T.A., 1986. Recurrence interval of long hydrologic events. J. Hydraul. Eng. 112 (6), 518–537. https://doi.org/10.1061/(ASCE)0733-9429(1986)112:6(518).
Syktus, J., Trancoso, R., Toombs, N., Wong, K., Thatcher, M., 2022. Dynamical downscaling of CMIP6 global models with a variable resolution climate model in the Australian region. International Conference on Southern Hemisphere Meteorology and Oceanography, ICSHMO 2022, Christchurch, New Zealand 2022.
Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. 106 (D7), 7183–7192.
Teichmann, C., Jacob, D., Remedio, A.R., Remke, T., Buntemeyer, L., Hoffmann, P., Kriegsmann, A., Lierhammer, L., Bülow, K., Weber, T., Sieck, K., Rechid, D., Langendijk, G.S., Coppola, E., Giorgi, F., Ciarlo', J.M., Raffaele, F., Giuliani, G., Xuejie, G., Sines, T.R., Torres-Alavez, J.A., Das, S., Di Sante, F., Pichelli, E., Glazer, R., Ashfaq, M., Bukovsky, M., Im, E.-S., 2021. Assessing mean climate change signals in the global CORDEX-CORE ensemble. Climate Dynamics 57 (5-6), 1269–1292.
Tokarska, K.B., Stolpe, M.B., Sippel, S., Fischer, E.M., Smith, C.J., Lehner, F., Knutti, R., 2020. Past warming trend constrains future warming in CMIP6 models. Sci. Adv. 6, eaaz9549.
Wang, F., Tian, D., Carroll, M., Customized Deep Learning for Precipitation Bias Correction and Downscaling, Geoscientific Model Development Discussions [preprint], doi: 10.5194/gmd-2022-213.
Watterson, I.G., 1996. Non-dimensional measures of climate model performance. Int. J. Climatol. 16 (4), 379–391.
Whetton, P., Hennessy, K., Clarke, J., McInnes, K., Kent, D., 2012. Use of Representative Climate Futures in impact and adaptation assessment. Climatic Change 115 (3–4), 433–442.
Zelinka, M., 2022. mzelinka/cmip56_forcing_feedback_ecs: Feb 28, 2022 Release (v2.1). Zenodo.