Received: 8 October 2023 | Revised: 12 May 2024 | Accepted: 17 May 2024

DOI: 10.1002/met.2217

RESEARCH ARTICLE

Meteorological Applications
Science and Technology for Weather and Climate
Open Access
RMetS

A multi-model likelihood analysis of unprecedented extreme rainfall along the east coast of Australia

Damien B. Irving[1] | James S. Risbey[1] | Dougal T. Squire[1] | Richard Matear[1] | Carly Tozer[1] | Didier P. Monselesan[1] | Nandini Ramesh[2] | P. Jyoteeshkumar Reddy[1] | Mandy Freund[1]

[1]Environment, Commonwealth Scientific and Industrial Research Organisation, Hobart, Australia

[2]Data61, Commonwealth Scientific and Industrial Research Organisation, Hobart, Australia

Correspondence
Damien B. Irving, Environment, Commonwealth Scientific and Industrial Research Organisation, Hobart, Tasmania 7000, Australia.
Email: damien.irving@csiro.au

Funding information
Australian Climate Service

Abstract

A large stretch of the east coast of Australia experienced unprecedented rainfall and flooding over a two-week period in early 2022. It is difficult to reliably estimate the likelihood of such a rare event from the relatively short observational record, so an alternative is to use data from an ensemble prediction system (e.g., a seasonal or decadal forecast system) to obtain a much larger sample of simulated weather events. This so-called 'UNSEEN' method has been successfully applied in several scientific studies, but those studies typically rely on a single prediction system. In this study, we use data from the Decadal Climate Prediction Project to explore the model uncertainty associated with the UNSEEN method by assessing 10 different hindcast ensembles. Using the 15-day rainfall total averaged over the river catchments impacted by the 2022 east coast event, we find that the models produce a wide range of likelihood estimates. Even after excluding a number of models that fail basic fidelity tests, estimates of the event return period ranged from 320 to 1814 years. The vast majority of models suggested the event is rarer than a standard extreme value assessment of the observational record (297 years). Such large model uncertainty suggests that multi-model analysis should become part of the standard UNSEEN procedure.

KEYWORDS

Australia, climate extremes, climate risk, CMIP6, decadal forecast, extreme rainfall

1 | INTRODUCTION

Extreme rainfall and significant flooding affected the east coast of Australia from 22 February to 9 March 2022. The heavy rains began in south-east Queensland and north-east New South Wales during the last week of February and continued further south into eastern New South Wales in early March. Numerous multi-day rainfall records were broken over the two-week period and many rivers peaked at record (in north-east New South Wales especially) or near-record levels (Australian Bureau of Meteorology, 2022). Twenty-two people are known to have died during the disaster and as of August 2023 it had caused AU$6 billion in insured damages, surpassing the 1999 Sydney hailstorm as the costliest extreme weather event in Australia's history (Insurance Council of Australia, 2023).

In the aftermath of unprecedented extreme weather, there is often interest from policymakers, contingency planners and insurers in understanding the probability of experiencing such an event. Our observational records typically only span a century or so, which is well short of the many thousands of years of data required to adequately constrain likelihood estimates (Risbey et al., 2023). In an attempt to address this small sample problem, the research community has turned to large model ensembles. The UNprecedented Simulated Extremes using ENsembles (UNSEEN) approach has been used to assess the likelihood of unprecedented events in a wide range of contexts including monthly wintertime rainfall (Thompson et al., 2017), daily summertime rainfall (Kent et al., 2022) and summertime heat (Kay et al., 2020) in the United Kingdom; monthly summertime heat in south-east China (Thompson et al., 2019); drought in the north-east farming region of China (Kent et al., 2019); summer monsoon rain in India (Jain et al., 2020; Jain & Scaife, 2022); severe water stress across major global maize producing regions (Kent et al., 2017); flooding in the Amazon River basin (Kelder, Wanders, et al., 2022); concurrent extreme drought and fire weather in south-east Australia (Squire et al., 2021); wet season rain in northern Brazil (Kay et al., 2022); El Niño and La Niña events (Merryfield & Lee, 2023); and sudden stratospheric warming in the Southern Hemisphere (Wang et al., 2020). It has also been used to look at trends in 3-day rainfall extremes in western Norway (Kelder et al., 2020) and growing season temperatures in South Africa (Bradshaw et al., 2022).

Now that the UNSEEN approach has been applied by a number of different authors in a variety of different contexts, it is possible to identify the common analysis steps involved (Kelder, Marjoribanks, et al., 2022). The process starts by defining the event/metric of interest and selecting a large ensemble dataset to analyse. Initialised ensemble predictions (i.e., forecast or hindcast datasets) are usually preferred because they typically involve a large number of ensemble members and simulate recent decades many times over (often using multiple initialisation dates per year). In contrast, CMIP-style historical experiments typically have fewer ensemble members and extend back to the mid-to-late 1800s, which can be problematic if there is a strong anthropogenic trend in the variable of interest. Forecast systems including the Met Office Decadal Prediction System (DePreSys3; Dunstone et al., 2016), European Centre for Medium-Range Weather Forecasts seasonal forecast system (SEAS5; Johnson et al., 2019) and Commonwealth Science and Industrial Research Organisation Climate Analysis Forecast Ensemble (CAFE; O'Kane et al., 2021a, 2021b) have all been used in UNSEEN studies. The type of event may inform the choice of forecast dataset and vice versa; for instance, an event of short duration over a small spatial area lends itself to a higher resolution dataset (i.e., a seasonal forecast system), whereas drought events need longer simulations to avoid issues of dependence (i.e., a decadal forecast system). Once an appropriate metric has been defined and calculated from the large ensemble dataset, an evaluation of the independence, stability and fidelity of that metric is conducted and (if necessary) steps are taken to resolve any detected issues (Kelder, Marjoribanks, et al., 2022). Depending on the situation, these steps might include removing early lead times to resolve independence or stability issues, bias correction to resolve fidelity issues, or a decision not to proceed with the analysis if the issues cannot be satisfactorily resolved.

While multi-model analysis is common-place in climate projections science, most UNSEEN studies to-date have analysed a single large ensemble. If they do use multiple models, samples from all models have been pooled to produce a single super ensemble (Jain & Scaife, 2022; Kent et al., 2022). This means the structural model uncertainty (Knutti et al., 2010) associated with likelihood estimates derived from UNSEEN analysis is largely unknown. The Decadal Climate Prediction Project (DCPP; Boer et al., 2016) contribution to the Coupled Model Intercomparison Project Phase 6 (CMIP6; Eyring et al., 2016) provides a new opportunity to explore structural model uncertainty. With over a dozen modelling centres participating in the DCPP, it could be a valuable resource for the UNSEEN community. In this study, we apply the UNSEEN method to estimate the likelihood of the extreme rainfall event along the east coast of Australia in late February and early March 2022. We use the DCPP dataset to perform a multi-model analysis in an attempt to understand how likelihood estimates can vary depending on the modelling system used. The paper is structured according to the steps involved in an UNSEEN analysis: Section 2 defines the extreme event of interest, Section 3 describes the selected large ensembles, Section 4 presents the evaluation of those large ensembles and Section 5 presents the results of the likelihood analysis.

2 | EVENT DEFINITION

Over the 15-day period from 23 February to 9 March 2022, significant rainfall totals were observed east of the Great Dividing Range all the way from the Sunshine Coast in Queensland southwards to the border between Victoria and New South Wales. In order to capture the full temporal and spatial extent of the event in one simple metric, we decided to aggregate daily precipitation data by calculating the annual (September to August) maximum consecutive 15-day total averaged over the river catchment areas impacted by the event. Using the topographic drainage divisions and river regions defined by the Australian Hydrological Geospatial Fabric (Atkinson et al., 2008), this represented all river regions in the South East Coast (NSW) drainage division and the regions in the North East Coast division south of (and including) the Burrum River and Fraser Island regions (Figure 1a). Not only do these river regions closely match the area of highest rainfall totals during the event of interest, but also they are highly relevant to the widespread flooding that was experienced. The use of a non-standard September to August annual calendar reduced the chances of splitting a significant 15-day rainfall event over 2 years, since that is the driest time of year averaged over the area of interest. Following the nomenclature used in the climate extremes literature, we abbreviate the metric to Rx15day.

The Australian Gridded Climate Data (AGCD) dataset (formerly called the Australian Water Availability Project dataset) is the Australian Bureau of Meteorology's official dataset for climate analyses. The daily rainfall data are available on a national 5-km spatial grid from 1900 to present day (Jones et al., 2009). Calculation of the Rx15day metric using AGCD data reveals that the 2022 Rx15day event of 410 mm was indeed unprecedented in the observational record, far exceeding the previous record of 289 mm in February 2020 (Figure 2a). The event occurred during the most common time of the year for Rx15day events (February/March; Figure 2b). Using the Met Office Hadley Centre's sea ice and sea surface temperature (HadISST; Rayner, 2003) dataset to calculate the monthly Niño3.4 (5 N–5S, 170–120 W) anomaly for the month corresponding to the final day of each Rx15day event, it appears that large Rx15day totals have historically occurred during both La Niña and neutral phases of the El Niño Southern Oscillation (ENSO) but not during the El Niño phase (Figure 2c).

The extreme rainfall over the 15-day period of interest was associated with a blocking high-pressure system over New Zealand, which assisted the formation of a series of slow-moving low-pressure systems within a trough that fed a large volume of warm moist air from the Coral and Tasman seas into eastern Australia. The subsequent development of a series of deep low-pressure systems delivered intense rain to east and south-east New South Wales (Australian Bureau of Meteorology, 2022). We see the blocking high and onshore flow clearly (Figure 1b) in

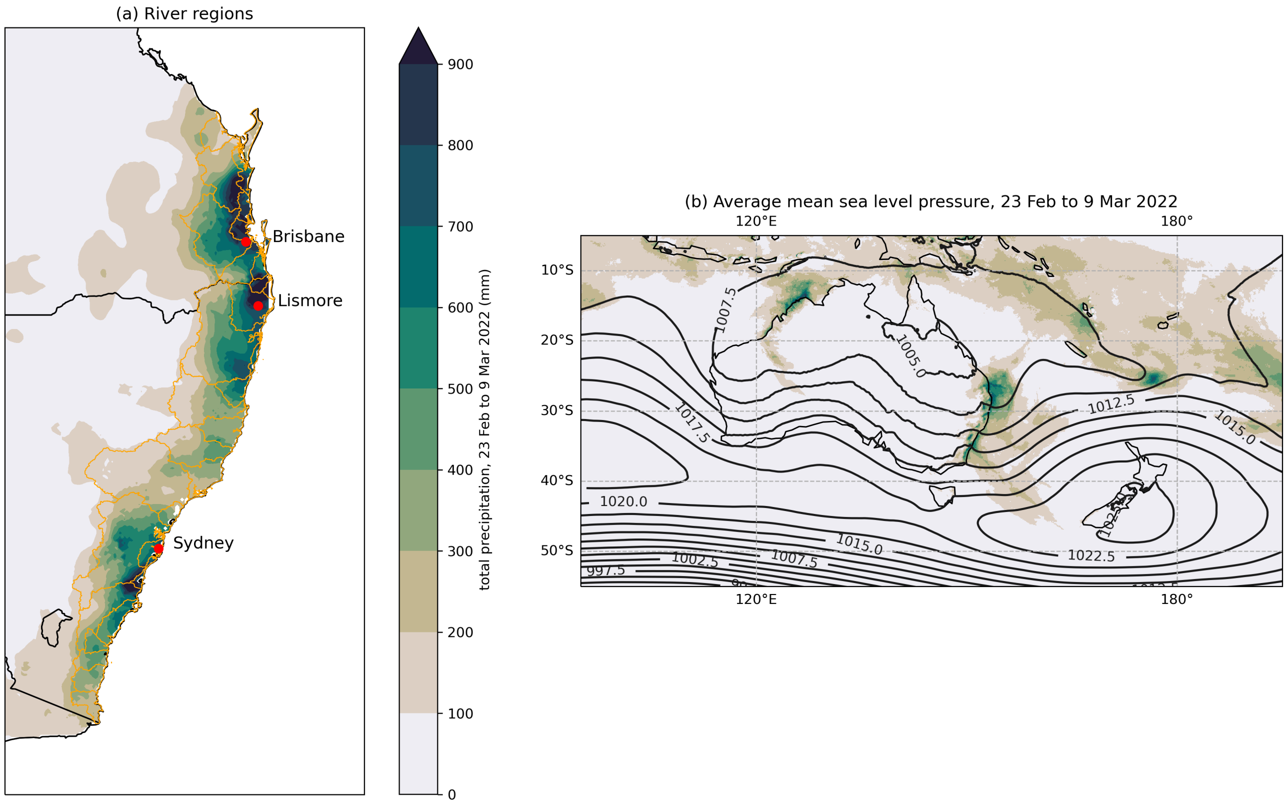

FIGURE 1 Rainfall totals and average circulation over the 15-day period 23 February to 9 March 2022. The river regions used in calculating the Rx15day metric are shown in orange (panel a). Data sources are AGCD (panel a) and BARRA-R2 (panel b) respectively.

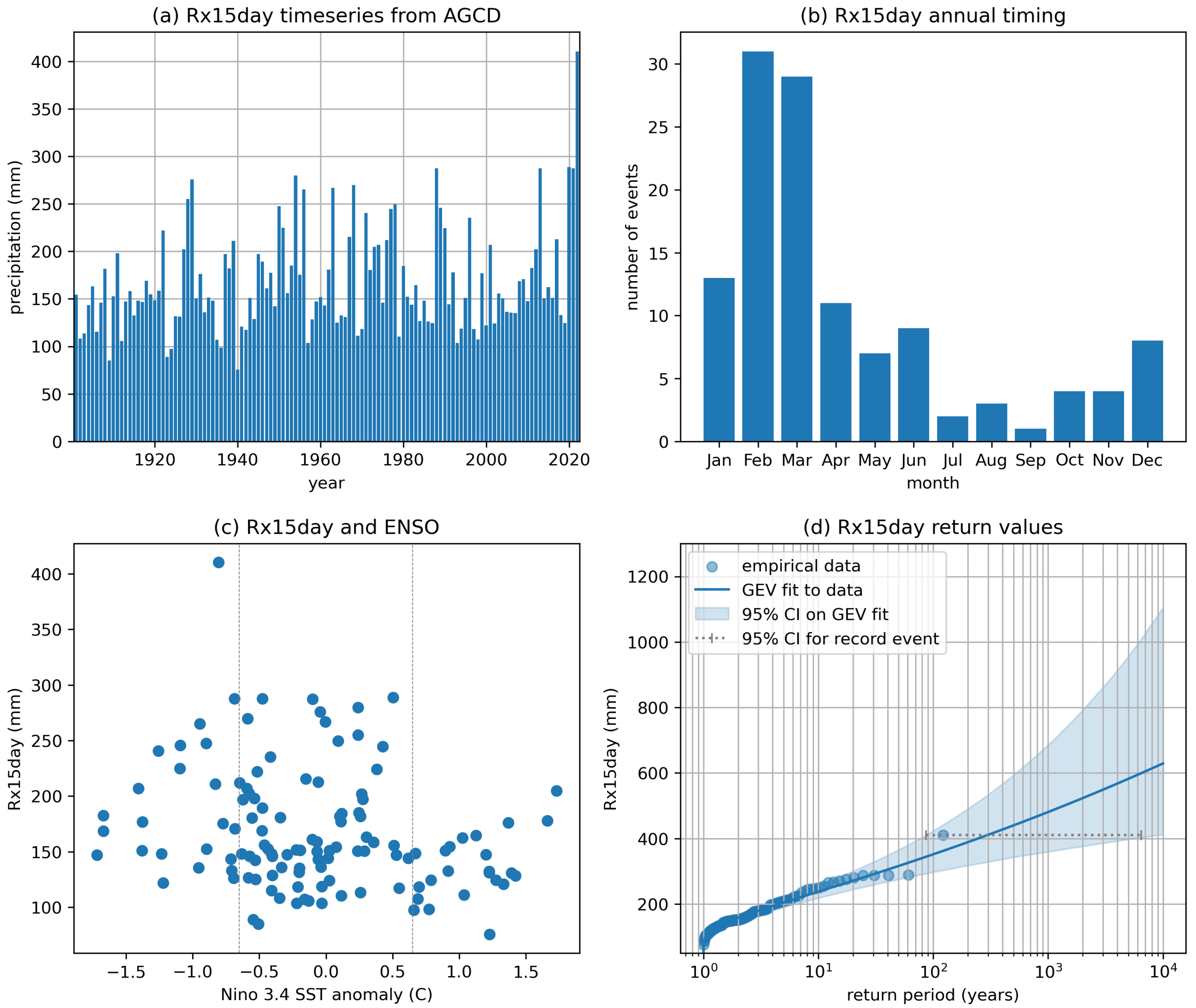

FIGURE 2 Rx15day in the 122 year (1901–2022) AGCD observational record. Niño 3.4 values above 0.65C and below −0.65C are commonly used as thresholds to designate a El Niño and La Niña events respectively (panel c). The 95% confidence interval (CI) on the return period curve was derived from 10,000 bootstraps of the observational record (panel d). It spans 87 to 6451 years for the 410-mm event observed in February/March 2022.

data from the new second version of the Bureau of Meteorology Atmospheric high-resolution Regional Reanalysis for Australia (BARRA-R2; Su et al., 2019, 2022). Those general features (block and trough) are also present for the several highest Rx15day events in the observational record (not shown).

3 | LARGE ENSEMBLE DATA

Given that complete years of data are required to calculate an annual metric like Rx15day, the most appropriate data source for the study was decadal (as opposed to seasonal) forecast data. At the time of writing, a total of nine DCPP models (Table 1) had submitted daily precipitation data (required to calculate the Rx15day metric) for the 'dcppA-hindcast' experiment (Boer et al., 2016). Each submission typically involved running a 10–20-member ensemble of 10-year hindcasts initialised once annually from 1960 to the present day (i.e., around 2017 when most modelling groups performed the DCPP experiments). The spatial resolution of the DCPP models ranges from approximately 0.6 degrees of latitude by 0.8 degrees of longitude (HadGEM3-GC31-MM) to 2.5 degrees of latitude and longitude (CanESM5). The model resolutions in Table 1 are expressed as the number of model grid cells that overlap (by 10% of the grid cell area or more) with the region of interest, which were the cells selected in spatially aggregating the data to calculate Rx15day (the coarser resolution models have a smaller number of

TABLE 1 Model characteristics.

Model	Sample size	Resolution[a]
CAFE	34,944	11
CanESM5	10,260	6
CMCC-CM2-SR5	3600	24
EC-Earth3	7830	51
HadGEM3-GC31-MM	5310	53
IPSL-CM6A-LR	5130	12
MIROC6	5310	18
MPI-ESM1-2-HR	5310	31
MRI-ESM2-0	2400	25
NorCPM1	9440	9

[a]Resolution refers to the number of grid cells that overlap with the analysis region.

cells). The sample size (i.e., number of Rx15day values) for each DCPP model ranged from 2400 to 10,260 depending on the precise number of ensemble members, hindcast length, initialisation years and the results of the independence test (see below). In order to include a much larger ensemble in the analysis, we also used data from the CAFE hindcast dataset which comprises 10-year long, daily forecasts, each with 96 ensemble members, initialised at the beginning of every May and November over the period 1995–2020 (for an ensemble size of 34,944).

4 | EVALUATION

In order to provide a reliable estimate of the likelihood of the unprecedented 2022 event, the Rx15day samples from a given model must be stable, independent and realistic estimates of the real world (Kelder, Marjoribanks, et al., 2022). Stability here refers to an absence of systematic changes in the estimates of Rx15day as the forecasts progress in time (i.e. an absence of model drift; Irving et al., 2021). The time elapsed since the beginning of a forecast is known as the 'lead time' and stability is necessary for pooling samples across lead times. With respect to independence, any dependence between model samples inflates the sample size without adding new information and can arise at short lead times because the ensemble forecasts are initialised from similar initial conditions. Finally, the realism (or fidelity) of a simulation refers to its ability to simulate Rx15day events both in terms of their observed statistical properties and also the relevant physical processes (e.g., the meteorological features typically associated with such events).

It was also necessary to assess the stationarity of the Rx15day metric because if there were strong trends in time (as opposed to lead time; e.g., due to anthropogenic forcing), it might be necessary to limit the likelihood analysis to model years close to 2022 (i.e., by discarding the earlier decades from DCPP submissions that start around 1960) or de-trend the data prior to use. Strong trends might also mean that fidelity tests (i.e., comparing observational and model data) need to be performed over a common time period. Fewer model samples are available for calendar forecast years towards the start and end of each model forecast period (because these years have fewer lead times available), so if the data are highly non-stationary, it can be necessary to also limit fidelity and likelihood assessments to model years that are equally sampled (Squire et al., 2021).

4.1 | Stability and stationarity

In the context of UNSEEN analysis, the most common method of assessing model stability is to check that the probability density functions for each lead time (for the metric of interest) are similar and that the extreme value distributions of the individual lead times fall within the uncertainty range of the distribution for all lead times pooled together (e.g., Kelder et al., 2020). We find that the CAFE and DCPP models pass these simple checks (see Figure 3a,b for a representative model and Figures S1–10a,b for all models), meaning that the model simulations are sufficiently stable to allow the pooling of data across all lead times without any de-drifting or removal of problematic lead times. A similar approach can be used to assess model stationarity, grouping the data by time (e.g., by decade) as opposed to by lead time. In the case of our Rx15day metric, all of the models were stationary with respect to time (Figures 3c,d and S1–10c,d). Not only is this finding useful for deciding whether de-trending or the use of restricted time periods is required, but it also suggests that climate change has not substantially altered the likelihood of extreme Rx15day values over the historical period. This is consistent with global gridded analyses of the commonly used Rx5day metric, which show no significant trend for grid points along the east coast of Australia in observations (Dunn et al., 2020) or in the CMIP6 ensemble until much later this century in the higher emission scenarios (Almazroui et al., 2021).

4.2 | Independence

To determine the lead time at which the ensemble members can be considered independent, we follow Squire et al. (2021) and test whether the correlation between

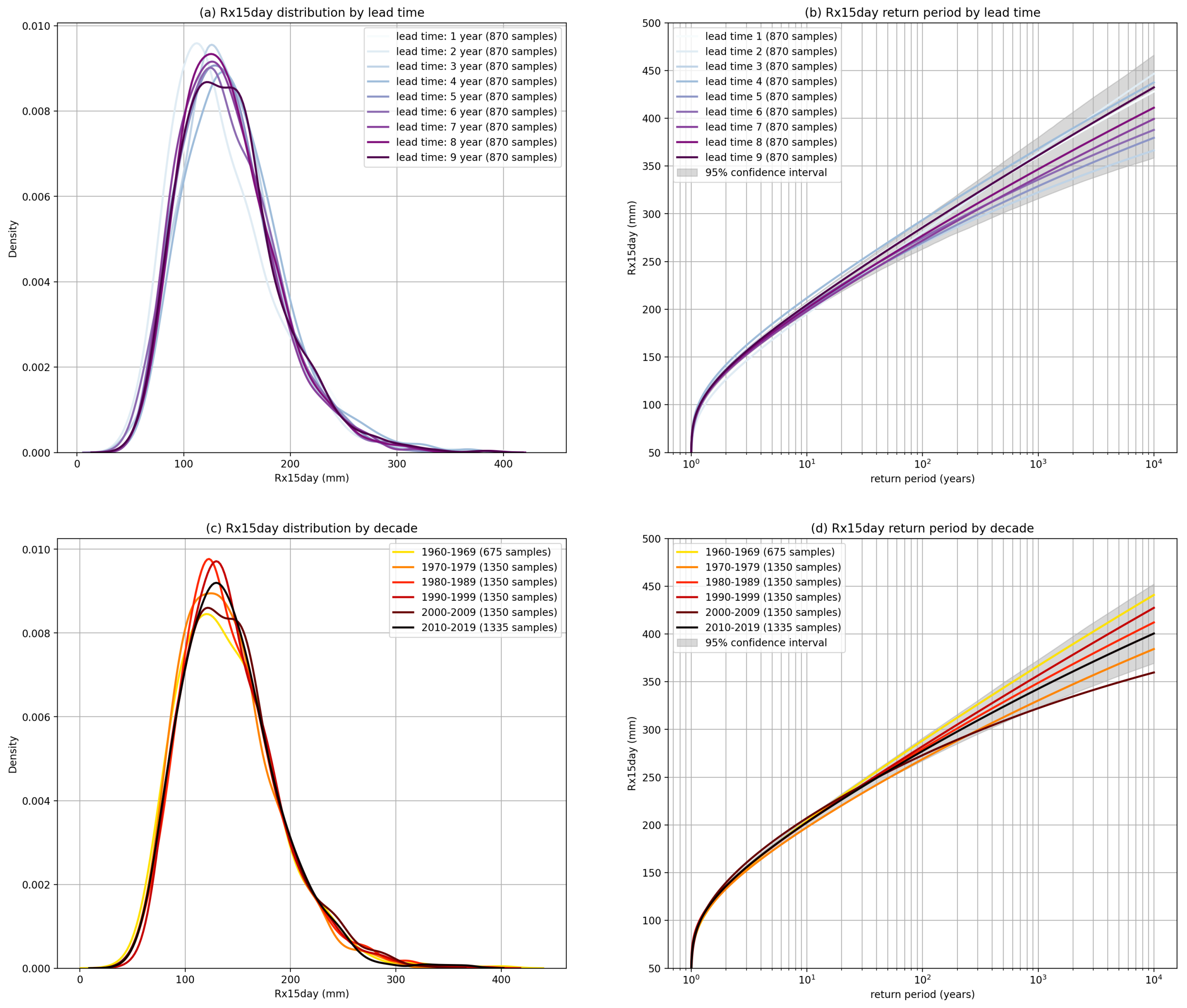

FIGURE 3 Stability and stationarity evaluation for the EC-Earth3 model. Return period curves (panels b and d) were derived from a GEV fit to the data. Grey shading illustrates the 95% confidence intervals of the distribution of the pooled lead times, bootstrapped to timeseries of similar length to the individual lead times with $n = 1000$.

ensemble members at a given lead time is sufficiently close to zero. At each lead time, the HadGEM3-GC31-MM submission to DCPP (for instance) provides 10 (members), 59-year timeseries of Rx15day (spanning, e.g., 1961–2019 at 1-year lead, or 1965–2023 at 5-year lead). We define our test statistic, ρ_t, for each lead time as the mean Spearman correlation in time between all combinations of the 10 ensemble members (of which there are 45: member 1 with 2, member 1 with 3, etc). Significance of ρ_t is estimated using a permutation test, whereby 10,000 sets of 10×59 points are randomly drawn from the complete model dataset to produce 10,000 estimates of the mean Spearman correlation. Because these estimates are constructed from randomly drawn data, they represent the distribution of mean correlation values for uncorrelated data (i.e., the null distribution). Ensemble members are considered to be dependent (i.e., the null hypothesis of independence is rejected) at a given lead time if ρ_t falls outside of the 95% confidence interval calculated from the randomly sampled distribution. Samples from dependent lead times (see Figure S11) were removed prior to fidelity and likelihood assessment.

4.3 | Fidelity

A number of statistical tests are used in the UNSEEN literature to assess how well a forecast ensemble simulates the event/metric of interest. The most common is the so-called bootstrap or moments test, whereby the model data are bootstrapped into a large number of (e.g., 10,000) series of equal length to the observed

Meteorological Applications
Science and Technology for Weather and Climate
Open Access
RMetS

timeseries and the empirical moments of each series (mean, standard deviation, skewness and kurtosis) are calculated (e.g., Kelder et al., 2020; Thompson et al., 2017). If the moments of the observed timeseries fall within the 95% confidence intervals for the statistics derived from the bootstrapped series, the model is considered to have passed the test. In addition to these four basic empirical moments, some authors have also calculated the shape, location and scale parameters from a Generalised Extreme Value (GEV) distribution fit (using maximum likelihood estimation of the distribution parameters) to the data (Kelder et al., 2020; Kent et al., 2022). In order to avoid issues associated with multiple testing (Wilks, 2011), other authors prefer a single test score comparing the modelled and observed data. The Kolmogorov–Smirnov test (e.g., Squire et al., 2021) and Anderson-Darling test (e.g., Kent et al., 2022) have been used to assess how likely it is that the observed and model samples were drawn from the same (but unknown) probability distribution. A test *p*-value of greater than 0.05 is typically taken to indicate that the null hypothesis (that the two samples are from the same population) cannot be rejected, meaning the model data are sufficiently similar to observations to be used in likelihood analysis. Each of these tests/approaches can give slightly different insights, so we apply all of them to our 10 models.

We find that all 10 models fail the moments test in a similar way, with the observed mean and standard deviation lying beyond the upper limit of the bootstrapped 95% confidence interval (see Figure 4 for a representative model and Figures S12-21 for all models). In other words, the distribution of modelled Rx15day values in all models was too dry and lacked variance, which meant the observed GEV location (closely linked to the mean) and scale (linked to standard deviation) parameters also lay beyond the 95% confidence interval. In fact, only one model (MIROC6) simulated an Rx15day value greater than the 410 mm observed in 2022. All the models bar one (CMCC-CM2-SR5) failed the Kolmogorov–Smirnov and Anderson-Darling tests (Table S1). Aside from the general underestimation of the observed mean and variance, the shape of the model Rx15day distribution was relatively well simulated, with all models passing the moments tests relating to skewness, kurtosis and the GEV shape parameter (Figures 4 and S12-21).

The most common bias correction method used in the UNSEEN literature to overcome model bias in extreme precipitation metrics is simple multiplicative mean scaling, whereby the model data (i.e., the calculated metric of interest; in this case Rx15day) is multiplied by the ratio of the average observed and modelled metric values. Since the model simulations were found to be stable and there was no significant trend in the observations over time, it was appropriate to calculate the average observed value over the entire observational record (i.e., 1901–2022) and average model value over all initialisation dates, lead times and ensemble members. When the ratio of those two values was used to correct the modelled Rx15day values, all of the models passed the Kolmogorov–Smirnov and Anderson-Darling tests (Table S2). Performance on the mean and standard deviation components of the moments test was also generally much improved, although the standard deviation of a few of the models (CMCC-CM2-SR5, IPSL-CM6A-LR, MIROC6, NorCPM1) still lay beyond the upper bound of the 95% confidence interval (Figures 4 and S12-21). The acceptable shape of the model distribution was also retained after the bias correction and most of the models simulated at least one (and in some cases several) Rx15day values greater than 410 mm. The corrected and uncorrected distributions for each model are shown in Figure 5a-j.

Going beyond statistical tests, some UNSEEN studies also conduct a process-based analysis to see whether the models simulate the basic meteorology and other characteristics (e.g., seasonality) associated with metric of interest (e.g., Kelder, Wanders, et al., 2022). Similar to the observed unprecedented event in 2022 (Figure 2b), the most extreme Rx15day events in all the models we assessed were associated with an onshore flow along the coast of New South Wales and south-east Queensland (Figure 6a–j). Many of the models also simulated the observed blocking surface high near New Zealand (Figure 6b,d,e,g,h,j), while a few simulated a deep surface low off the coast of Queensland (Figure 6c,e,j) that was not a feature of the observed 2022 event. Most of the DCPP models only archived surface variables, so it was not possible to conduct a detailed analysis of the upper-level circulation features associated with Rx15day events. With respect to annual timing, all the models do a reasonably good job of capturing the tendency for Rx15day events to occur during the summer months (Figure S22). Most of the models show a tendency for Rx15day events to be slightly more extreme during La Niña events and less extreme during El Niño events (Figure S23), which is broadly consistent with the observations (Figure 2c).

5 | LIKELIHOOD ANALYSIS

The ultimate aim of any UNSEEN analysis is to compare and contrast a likelihood estimate for an unprecedented event obtained from the (relatively short) observational record with that obtained from a large model ensemble/s. Such estimates are typically calculated by fitting a GEV

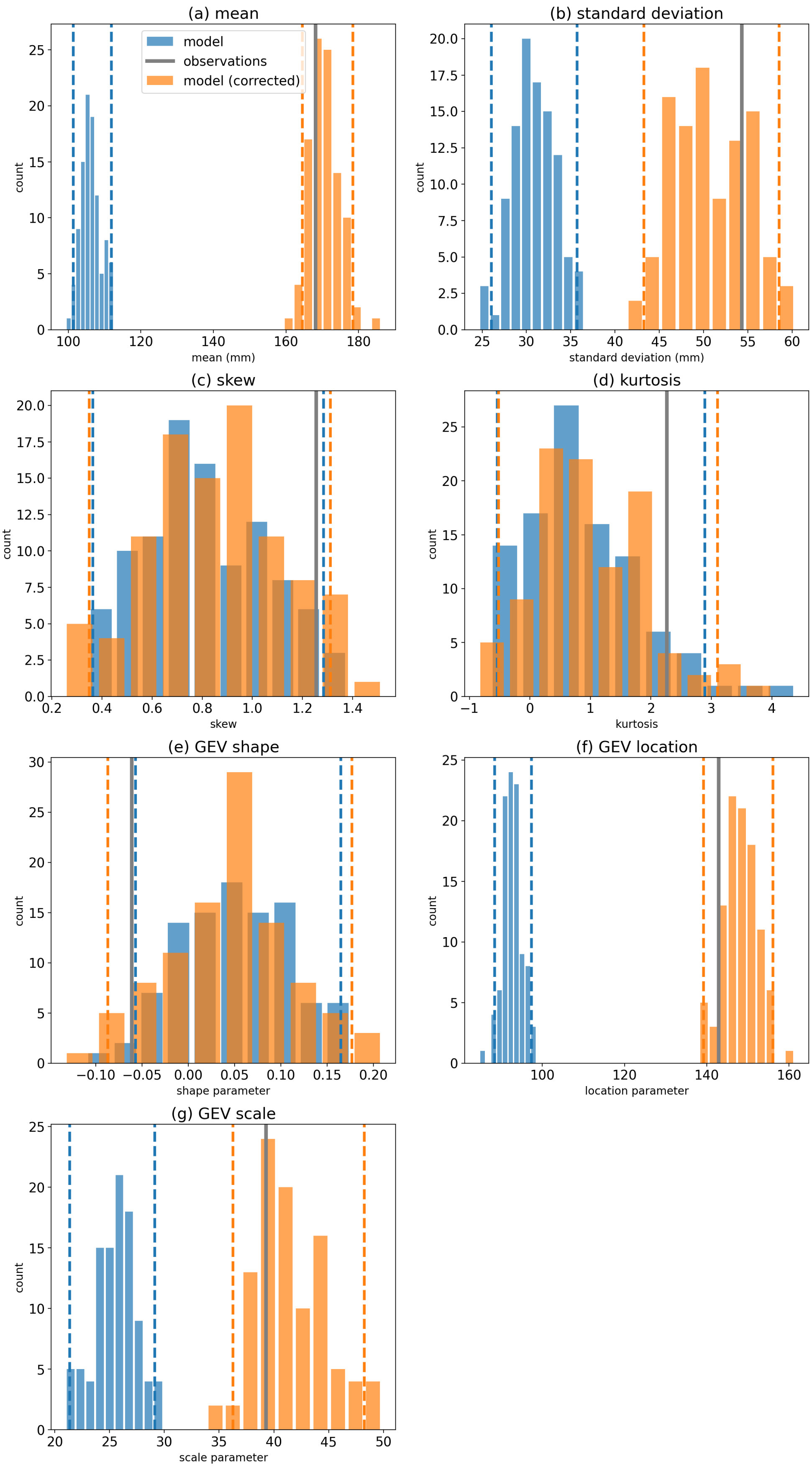

FIGURE 4 Moments tests for the MRI-ESM2-0 model. Each panel shows the histogram (bars) and 95% confidence interval (dashed lines) derived from 1000 bootstrap samples of 122 years (the length of the observational Rx15day record) from the model data. The sign convention for the GEV shape parameter is such that negative and positive values correspond to a Frechet and reversed Weibull distribution respectively.

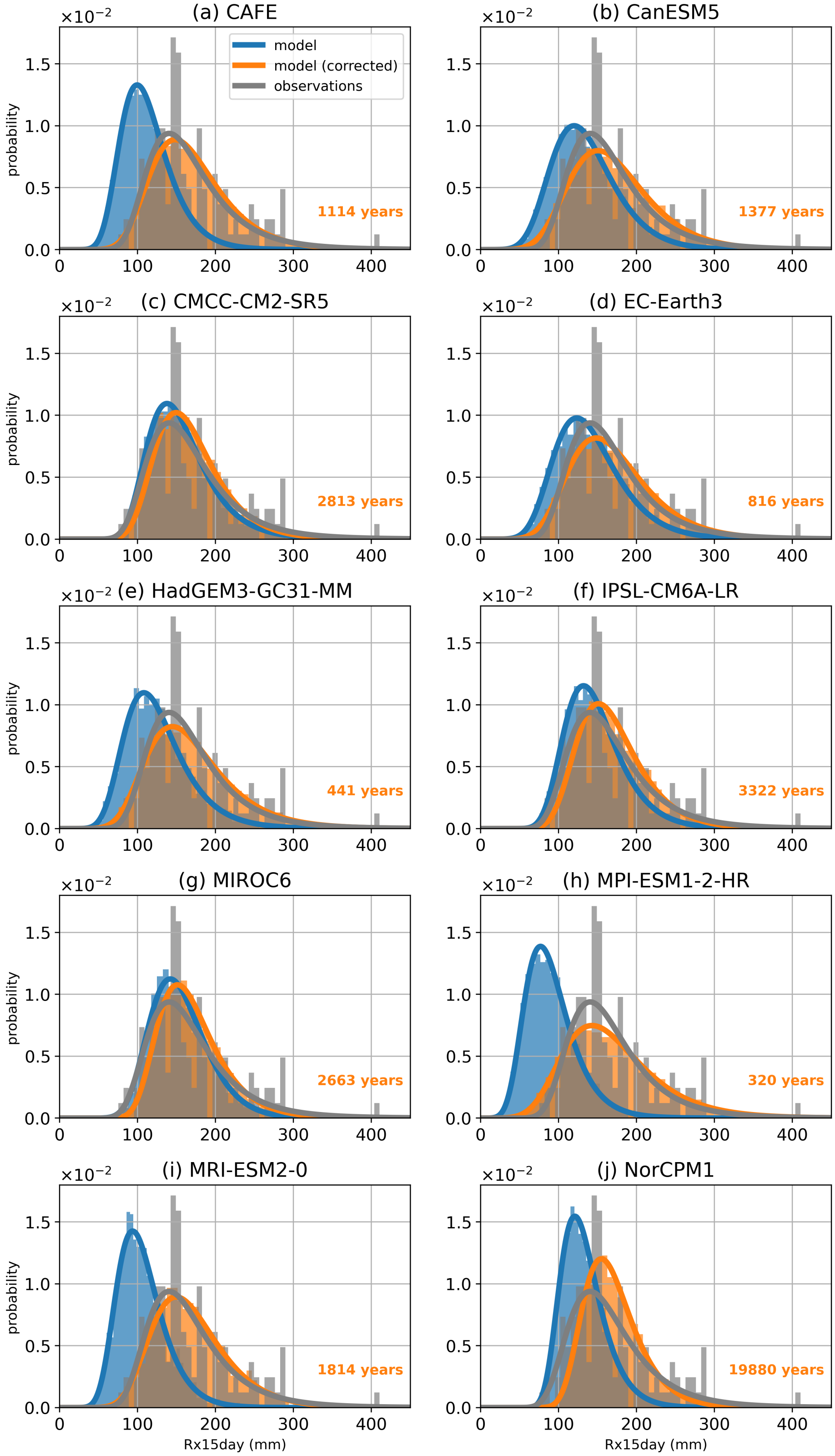

FIGURE 5 Probability distributions (histograms) and associated GEV fits (solid curves) for uncorrected model data, multiplicative bias-corrected model data and the full 122-year observational record. The number printed in each panel corresponds to the estimated return period for a 410-mm event from the GEV fit to the bias-corrected model data (see Figure S24 for the uncertainty bounds on those return period estimates).

to the data; the survival function corresponding to the GEV fit can be used to obtain an estimate of the event probability. That probability is typically communicated as a return period, which is the estimated average time between events. For instance, an event threshold that is only exceeded in 5 out of 100 annual samples has a probability of 0.05 or a return period of 20 years. When working with small samples like the observational record, estimation of the return period for an unprecedented event (i.e., an event way out in the tail of the distribution)

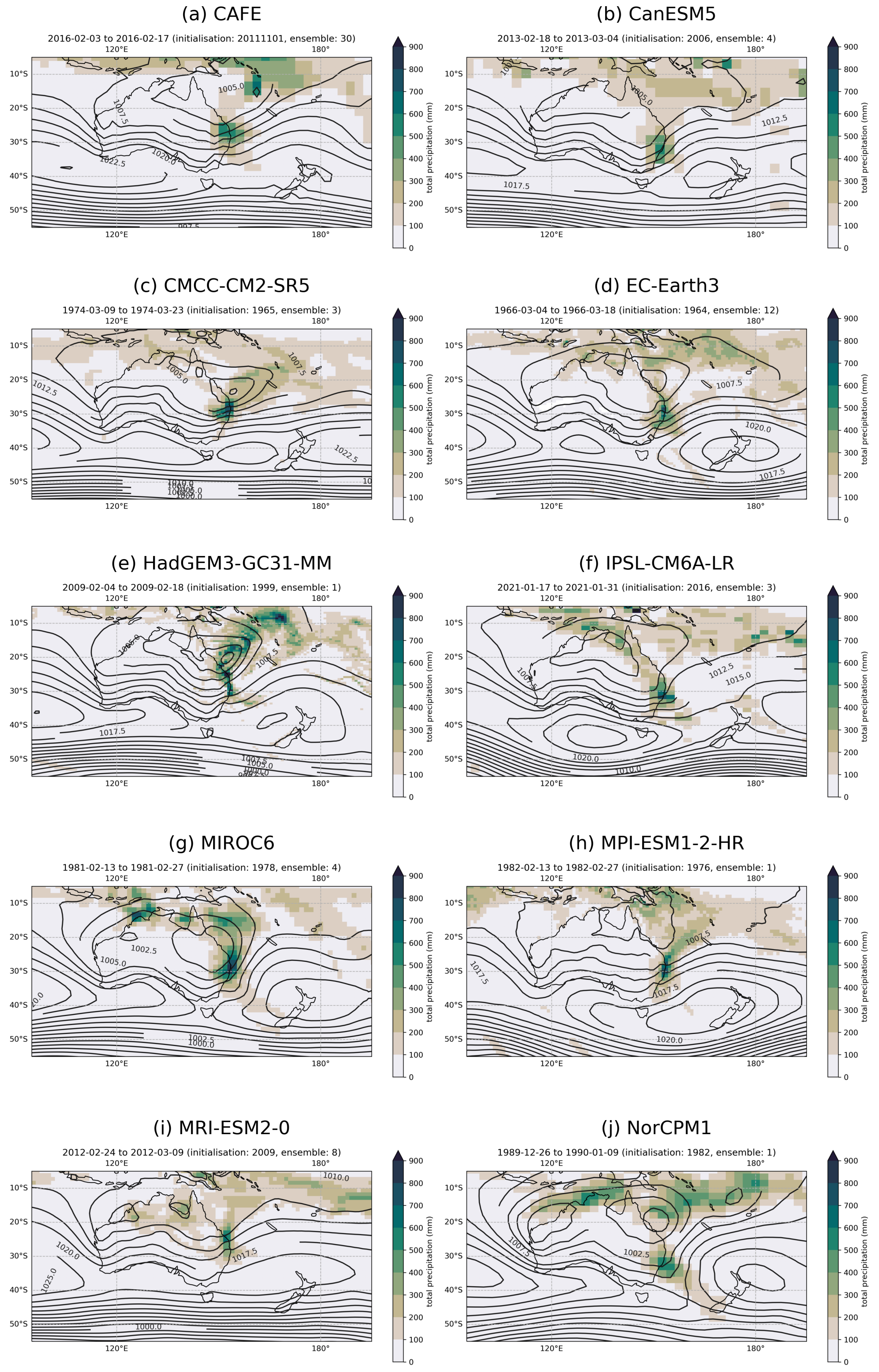

FIGURE 6 15-day mean sea level pressure for the most extreme Rx15day event from each model. Precipitation values have not been bias-corrected.

is complicated by the decision of whether or not to include the record event in the sample (Miralles & Davison, 2023). Including the unprecedented event will tend to cause an overestimation of the probability (underestimation of the return period) of the event and leaving out the unprecedented event tends to produce an underestimation of the probability (overestimation of the return period). In the case of the Rx15day values calculated from the AGCD dataset, a GEV fit to the entire 122-year record produces a likelihood estimate of 297 years for the 410-mm event that occurred in February/March 2022 (Figure 2d). A fit to a 121-year sample that excludes the record value from 2022 produces an estimate of 757 years.

For the model comparison, the fidelity evaluation guides the selection of appropriate data. It is clear from the various statistical tests we performed that it would not be appropriate to use uncorrected model data due to

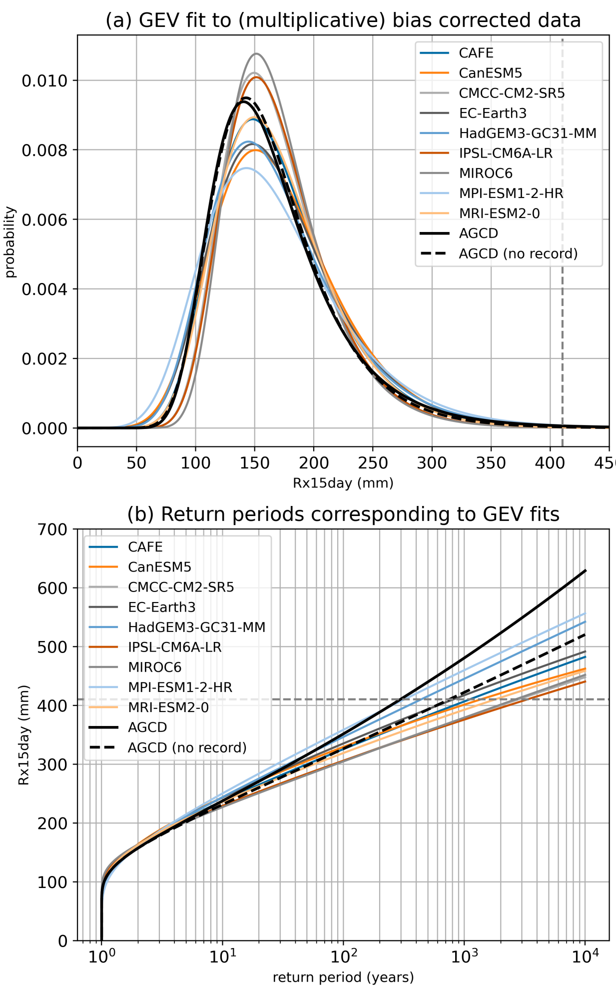

FIGURE 7 GEV fits to (multiplicative) bias-corrected Rx15day model data (panel a) and corresponding return periods (panel b). The grey dashed line indicates the observed record event of 410 mm.

substantial underestimation of the mean and variance of the Rx15day metric. For the multiplicative bias-corrected data, the four models that fail the moments test due to low variance (i.e. CMCC-CM2-SR5, IPSL-CM6A-LR, MIROC6 and NorCPM1) would be expected to produce an inflated return period, so estimates obtained from those data could be excluded or at least treated with scepticism. Further model elimination could be considered on the basis of the process-based fidelity assessment, but in this case, the models did a relatively good job of capturing the Rx15day seasonality, relationship with ENSO and characteristic mean sea level pressure pattern. For the remaining six models, the return period for a 410 mm event obtained from a GEV fit to the multiplicative bias-corrected data ranges from 320 to 1814 years, with a mean of 980 years (Figures 5 and S24). All the models produce longer return period estimates than the GEV fit to the full observational record, and all but two produce longer estimates than the 121 year observational sample (Figure 7).

6 | DISCUSSION

The UNSEEN approach is becoming an increasingly common method for estimating the likelihood of unprecedented climatic events, such as the extreme rainfall that occurred along the east coast of Australia from 22 February to 9 March 2022. The UNSEEN studies to-date have tended to rely on a single initialised ensemble. We make use of the DCPP dataset to perform a multi-model UNSEEN assessment of the extreme east coast rainfall and find a large spread in likelihood estimates across 10 different hindcast ensembles. The standard UNSEEN model fidelity assessment process allowed for the potential elimination of four of the 10 models, with the likelihood estimates from the remaining models ranging from a return period of 320 to 1814 years (with a mean of 980 years). The vast majority of the models suggest that the unprecedented 2022 event is rarer than a standard extreme value assessment of the 122-year observational record would suggest, which ranges from 297 to 757 years depending on whether the record event is included in the sample. A caveat on the model-derived estimates relates to the fact that all the models were biased in a similar way (they all underestimate the observed mean and variance), which may have manifested as a common over/underestimation of the return period. A detailed analysis of the impact of different bias correction methodologies and decisions will be a focus of future work, but for now suffice to say that a higher degree of confidence would be attached to estimates derived from a collection of models with a range of different bias profiles or better still models that did not require any bias correction whatsoever.

Attempts have been made in recent years to define a standard process for UNSEEN analysis (Kelder, Marjoribanks, et al., 2022). The large model uncertainty demonstrated in our analysis suggests that it would be worthwhile to consider adding multi-model analysis as part of that standard process, just as it is in other areas such as climate projections science (Knutti et al., 2010). The fidelity tests applied in UNSEEN studies can be used as a basis for including or excluding (or at least treating with scepticism) particular models, although different objective tests can lead to different outcomes. For instance, in our analysis of bias-corrected model data all 10 models passed the Kolmogorov–Smirnov and Anderson-Darling tests, but only six passed the moments test. There is an even greater degree of subjectivity involved with making model selections on the basis of physically based tests of model fidelity such as assessing the simulated meteorology (Kelder, Wanders, et al., 2022). Given this subjectivity, we decided not to exclude any models unless their physical representation of Rx15day events was very clearly and obviously lacking. Besides reporting the range of model likelihood estimates, it may be desirable to provide a central estimate such as the multi-model mean. In the climate projections space, equally weighted multi-model means tend to outperform means calculated by weighting each model on some measure of skill/fidelity (Weigel et al., 2010). An alternative would be to pool data from all the models and calculate the return period from a single super ensemble (Jain & Scaife, 2022; Kent et al., 2022). The impact of mixing different model distributions into one is unclear and the differing sample sizes between models would represent a form of weighting, so we decided not to analyse a super ensemble.

Initiatives like the DCPP that make a large selection of forecast ensembles easily accessible to the research community greatly facilitate multi-model UNSEEN analysis. The application of such datasets to unprecedented events in different geographic locations using different model variables and over different temporal and spatial scales will help to further understand the model uncertainty associated with the UNSEEN approach to likelihood estimation.

AUTHOR CONTRIBUTIONS

Damien B. Irving: Conceptualization (lead); data curation (lead); formal analysis (lead); methodology (lead); software (lead); visualization (lead); writing – original draft (lead); writing – review and editing (lead). **James S. Risbey:** Conceptualization (supporting); funding acquisition (lead); methodology (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Dougal T. Squire:** Conceptualization (supporting); data curation (supporting); methodology (supporting); software (supporting); writing – review and editing (supporting). **Richard Matear:** Conceptualization (supporting); funding acquisition (lead); methodology (supporting). **Carly Tozer:** Conceptualization (supporting); methodology (supporting); writing – review and editing (supporting). **Didier P. Monselesan:** Conceptualization (supporting); methodology (supporting); writing – review and editing (supporting). **Nandini Ramesh:** Conceptualization (supporting); methodology (supporting); writing – review and editing (supporting). **P. Jyoteeshkumar Reddy:** Conceptualization (supporting); methodology (supporting); writing – review and editing (supporting). **Mandy Freund:** Conceptualization (supporting); methodology (supporting); writing – review and editing (supporting).

ACKNOWLEDGEMENTS

We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modelling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access and the multiple funding agencies who support CMIP6 and ESGF.

FUNDING INFORMATION

This work was supported by the Australian Climate Service.

CONFLICT OF INTEREST STATEMENT

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 STATEMENT

With respect to the decadal forecast data used in this study, the DCPP dataset is publicly available through the distributed CMIP6 data archive developed and operated by the Earth System Grid Federation. The CAFE dataset is available via Australia's National Computational Infrastructure (NCI); for details see the online CAFE documentation (Squire, 2022) at https://cafef6.readthedocs.io. With respect to observational data, the AGCD and BARRA-R2 datasets are available via the NCI Data Catalogue (Bureau of Meteorology, 2023, 2024). HadISST1 data are available from the Met Office Hadley Centre at https://hadleyserver.metoffice.gov.uk/hadisst/data/download.html. The code written to process these datasets and produce the results presented in the paper is available at https://github.com/AusClimateService/east-coast-rain.

ORCID

Damien B. Irving https://orcid.org/0000-0003-1258-5002

REFERENCES

Almazroui, M., Saeed, F., Saeed, S., Ismail, M., Ehsan, M.A., Nazrul Islam, M. et al. (2021) Projected changes in climate extremes using CMIP6 simulations over SREX regions. *Earth Systems and Environment*, 5, 481–497. Available from: https://doi.org/10.1007/s41748-021-00250-5

Atkinson, R., Power, R., Lemon, D., O'Hagan, R., Dovey, D. & Kinny, D. (2008) The Australian hydrological geospatial fabric – development methodology and conceptual architecture. Technical Report URL https://publications.csiro.au/rpr/download?pid=procite:5126351f-b297-409b-b472-654d3534e3ae&dsid=DS1

Australian Bureau of Meteorology. (2022) Special Climate Statement 76: Extreme rainfall and flooding in south-eastern Queensland and eastern New South Wales. Technical report URL http://www.bom.gov.au/climate/current/statements/scs76.pdf

Boer, G.J., Smith, D.M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B. et al. (2016) The decadal climate prediction project (DCPP) contribution to CMIP6. *Geoscientific Model Development*, 9, 3751–3777. Available from: https://doi.org/10.5194/gmd-9-3751-2016

Bradshaw, C.D., Pope, E., Kay, G., Davie, J.C.S., Cottrell, A., Bacon, J. et al. (2022) Unprecedented climate extremes in South Africa and implications for maize production. *Environmental Research Letters*, 17, 084028. Available from: https://doi.org/10.1088/1748-9326/ac816d

Bureau of Meteorology. (2023) *Australian Gridded Climate Data (AGCD) v1.0.1/ Australian Water Availability Project (AWAP)* [Data set]. NCI Australia. Available from: https://doi.org/10.25914/HJQJ-0X55

Bureau of Meteorology. (2024) *Bureau of Meteorology Atmospheric high-resolution Regional Reanalysis for Australia – Regional Version 2 (BARRA-R2)* [Data set]. NCI Australia. Available from: https://doi.org/10.25914/90RQ-D839

Dunn, R.J.H., Alexander, L.V., Donat, M.G., Zhang, X., Bador, M., Herold, N. et al. (2020) Development of an updated global land in situ-based data set of temperature and precipitation extremes: HadEX3. *Journal of Geophysical Research-Atmospheres*, 125, e2019JD03226. Available from: https://doi.org/10.1029/2019JD032263

Dunstone, N., Smith, D., Scaife, A., Hermanson, L., Eade, R., Robinson, N. et al. (2016) Skilful predictions of the winter North Atlantic oscillation one year ahead. *Nature Geoscience*, 9, 809–814. Available from: https://doi.org/10.1038/ngeo2824

Eyring, V., Bony, S., Meehl, G.A., Senior, C.A., Stevens, B., Stouffer, R.J. et al. (2016) Overview of the coupled model Intercomparison project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9, 1937–1958. Available from: https://doi.org/10.5194/gmd-9-1937-2016

Insurance Council of Australia. (2023) Catastrophe 221: NSW and South East Queensland floods. URL https://insurancecouncil.com.au/news-hub/current-catastrophes/catastrophe-221-storms-in-south-east-qld-and-northern-nsw/

Irving, D., Hobbs, W., Church, J. & Zika, J. (2021) A mass and energy conservation analysis of drift in the CMIP6 ensemble. *Journal of Climate*, 34, 3157–3170. Available from: https://doi.org/10.1175/JCLI-D-20-0281.1

Jain, S. & Scaife, A.A. (2022) How extreme could the near term evolution of the Indian summer monsoon rainfall be? *Environmental Research Letters*, 17, 034009. Available from: https://doi.org/10.1088/1748-9326/ac4655

Jain, S., Scaife, A.A., Dunstone, N., Smith, D. & Mishra, S.K. (2020) Current chance of unprecedented monsoon rainfall over India using dynamical ensemble simulations. *Environmental Research Letters*, 15, 094095. Available from: https://doi.org/10.1088/1748-9326/ab7b98

Johnson, S.J., Stockdale, T.N., Ferranti, L., Balmaseda, M.A., Molteni, F., Magnusson, L. et al. (2019) SEAS5: the new ECMWF seasonal forecast system. *Geoscientific Model Development*, 12, 1087–1117. Available from: https://doi.org/10.5194/gmd-12-1087-2019

Jones, D., Wang, W. & Fawcett, R. (2009) High-quality spatial climate datasets for Australia. *Australian Meteorological and Oceanographic Journal*, 58, 233–248 URL http://www.bom.gov.au/jshess/docs/2009/jones_hres.pdf

Kay, G., Dunstone, N.J., Smith, D.M., Betts, R.A., Cunningham, C. & Scaife, A.A. (2022) Assessing the chance of unprecedented dry conditions over North Brazil during El Niño events. *Environmental Research Letters*, 17, 064016. Available from: https://doi.org/10.1088/1748-9326/ac6df9

Kay, G., Dunstone, N., Smith, D., Dunbar, T., Eade, R. & Scaife, A. (2020) Current likelihood and dynamics of hot summers in the UK. *Environmental Research Letters*, 15, 094099. Available from: https://doi.org/10.1088/1748-9326/abab32

Kelder, T., Marjoribanks, T.I., Slater, L.J., Prudhomme, C., Wilby, R.L., Wagemann, J. et al. (2022) An open workflow to gain insights about low-likelihood high-impact weather events from initialized predictions. *Meteorological Applications*, 29, e2065. Available from: https://doi.org/10.1002/met.2065

Kelder, T., Müller, M., Slater, L.J., Marjoribanks, T.I., Wilby, R.L., Prudhomme, C. et al. (2020) Using UNSEEN trends to detect decadal changes in 100-year precipitation extremes. *Npj Climate and Atmospheric Science*, 3, 47. Available from: https://doi.org/10.1038/s41612-020-00149-4

Kelder, T., Wanders, N., van der Wiel, K., Marjoribanks, T.I., Slater, L.J., Wilby, R.l. et al. (2022) Interpreting extreme climate impacts from large ensemble simulations—are they unseen or unrealistic? *Environmental Research Letters*, 17, 044052. Available from: https://doi.org/10.1088/1748-9326/ac5cf4

Kent, C., Dunstone, N., Tucker, S., Scaife, A.A., Brown, S., Kendon, E.J. et al. (2022) Estimating unprecedented extremes in UK summer daily rainfall. *Environmental Research Letters*, 17, 014041. Available from: https://doi.org/10.1088/1748-9326/ac42fb

Kent, C., Pope, E., Dunstone, N., Scaife, A.A., Tian, Z., Clark, R. et al. (2019) Maize drought Hazard in the northeast farming region of China: unprecedented events in the current climate. *Journal of Applied Meteorology and Climatology*, 58(10), 2247–2258. Available from: https://doi.org/10.1175/JAMC-D-19-0096.1

Kent, C., Pope, E., Thompson, V., Lewis, K., Scaife, A.A. & Dunstone, N. (2017) Using climate model simulations to assess the current climate risk to maize production. *Environmental Research Letters*, 12(5), 054012. Available from: https://doi.org/10.1088/1748-9326/aa6cb9

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J. & Meehl, G.A. (2010) Challenges in combining projections from multiple climate models. *Journal of Climate*, 23(10), 2739–2758. Available from: https://doi.org/10.1175/2009JCLI3361.1

Merryfield, W.J. & Lee, W.-S. (2023) Estimating probabilities of extreme ENSO events from Copernicus seasonal hindcasts. *Asia-Pacific Journal of Atmospheric Sciences*, 59, 479–493. Available from: https://doi.org/10.1007/s13143-023-00328-2

Miralles, O. & Davison, A.C. (2023) Timing and spatial selection bias in rapid extreme event attribution. *Weather and Climate Extremes*, 41, 100584. Available from: https://doi.org/10.1016/j.wace.2023.100584

O'Kane, T.J., Sandery, P.A., Kitsios, V., Sakov, P., Chamberlain, M.A., Collier, M.A. et al. (2021a) CAFE60v1: a 60-year large ensemble climate reanalysis. Part I: system design, model configuration and data assimilation. *Journal of Climate*, 34, 5153–5169. Available from: https://doi.org/10.1175/JCLI-D-20-0974.1

O'Kane, T.J., Sandery, P.A., Kitsios, V., Sakov, P., Chamberlain, M.A., Squire, D.T. et al. (2021b) CAFE60v1: a 60-year large ensemble climate reanalysis. Part II: evaluation. *Journal of Climate*, 34, 5171–5194. Available from: https://doi.org/10.1175/JCLI-D-20-0518.1

Rayner, N.A. (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, 108(D14), 4407. Available from: https://doi.org/10.1029/2002JD002670

Risbey, J.S., Irving, D.B., Squire, D.T., Matear, R.J., Monselesan, D.P., Pook, M.J. et al. (2023) A large ensemble illustration of how record-shattering heat records can endure. *Environmental Research: Climate*, 2, 035003. Available from: https://doi.org/10.1088/2752-5295/acd714

Squire, D.T., Richardson, D., Risbey, J.S., Black, A.S., Kitsios, V., Matear, R.J. et al. (2021) Likelihood of unprecedented drought and fire weather during Australia's 2019 megafires. *Npj Climate and Atmospheric Science*, 4(1), 64. Available from: https://doi.org/10.1038/s41612-021-00220-8

Squire, D.T., Chamberlain, M.A., Chapman, C.C., Collier, M.A., Fiedler, R., Irving, D.B., et al. (2022) *Documentation of the CAFE-f6 decadal climate forecasts (Version v1.0) [Computer software]*. Zenodo. Available from: https://doi.org/10.5281/ZENODO.7077359

Su, C.-H., Eizenberg, N., Steinle, P., Jakob, D., Fox-Hughes, P., White, C.J. et al. (2019) BARRA v1.0: the Bureau of Meteorology Atmospheric high-resolution regional reanalysis for Australia. *Geoscientific Model Development*, 12, 2049–2068. Available from: https://doi.org/10.5194/gmd-12-2049-2019

Su, C.-H., Rennie, S., Dharssi, I., Torrance, J., Smith, A., Le, T. et al. (2022) BARRA2: Development of the next-generation Australian regional atmospheric reanalysis. Technical report, Australian Bureau of Meteorology. URL http://www.bom.gov.au/research/publications/researchreports/BRR-067.pdf

Thompson, V., Dunstone, N.J., Scaife, A.A., Smith, D.M., Hardiman, S.C., Ren, H.-L. et al. (2019) Risk and dynamics of unprecedented hot months in south East China. *Climate Dynamics*, 52, 2585–2596. Available from: https://doi.org/10.1007/s00382-018-4281-5

Thompson, V., Dunstone, N.J., Scaife, A.A., Smith, D.M., Slingo, J.M., Brown, S. et al. (2017) High risk of unprecedented UK rainfall in the current climate. *Nature Communications*, 8, 107. Available from: https://doi.org/10.1038/s41467-017-00275-3

Wang, L., Hardiman, S.C., Bett, P.E., Comer, R.E., Kent, C. & Scaife, A.A. (2020) What chance of a sudden stratospheric warming in the southern hemisphere? *Environmental Research Letters*, 15, 104038. Available from: https://doi.org/10.1088/1748-9326/aba8c1

Weigel, A.P., Knutti, R., Liniger, M.A. & Appenzeller, C. (2010) Risks of model weighting in multimodel climate projections. *Journal of Climate*, 23(15), 4175–4191. Available from: https://doi.org/10.1175/2010JCLI3594.1

Wilks, D.S. (2011) Introduction. In: *Statistical methods in the atmospheric sciences*, 3rd edition. Oxford: Academic Press.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Irving, D. B., Risbey, J. S., Squire, D. T., Matear, R., Tozer, C., Monselesan, D. P., Ramesh, N., Reddy, P. J., & Freund, M. (2024). A multi-model likelihood analysis of unprecedented extreme rainfall along the east coast of Australia. *Meteorological Applications*, *31*(3), e2217. https://doi.org/10.1002/met.2217