Weather and Climate Extremes 34 (2021) 100384

Contents lists available at ScienceDirect

Weather and Climate Extremes

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

Improvements in subseasonal forecasts of rainfall extremes by statistical postprocessing methods

Ming Li [a,*], Huidong Jin [b], Quanxi Shao [a]

[a] CSIRO Data61, 26 Dick Perry Avenue, Kensington, 6151, Australia
[b] CSIRO Data61, GPO BOX 1700, Canberra, ACT, 2601, Australia

ARTICLE INFO

Keywords:
Seasonal rainfall forecast
Daily rainfall
Statistical postprocessing
Extreme rainfall indices
Forecast verification

ABSTRACT

End users of seasonal rainfall forecasts demand not only skilful forecasts of rainfall totals but also skilful forecasts of rainfall extremes. For better forecasts of rainfall extremes, the copula-based postprocessing (CPP) method, which was originally designed for forecasting rainfall totals, is modified with a hybrid probability distribution to model low-to-medium and heavy rainfall separately and to allow the forecast of extreme rainfall events that have never occurred in observed records. A case study for 17 rainfall stations in Queensland, Australia is carried out to test the forecast performance of the modified CPP to postprocess the raw Australian Community Climate and Earth-System Simulator (ACCESS) Seasonal model (ACCESS-S) seasonal rainfall forecast at a daily scale. The modified CPP improves the overall skill of forecasting 12 rainfall indices from the raw forecast and outperforms two quantile mapping based methods in most cases. The use of the hybrid distribution leads to more promising forecast skill for heavy and very heavy rainfall related indices. The forecast skill decreases with longer lead times and the modified CPP leads to neutral forecasts (i.e. forecasts with skill similar to climatology forecasts) for most rainfall indices beyond 0-month lead time. The skill improvement has been found in all selected climate regions and initialisation dates (from the 1st day of each month), though more substantial improvement is observed in the rainfall stations within the tropical zone where the raw forecast is particularly unskilful.

1. Introduction

Desired seasonal to subseasonal (S2S) rainfall forecasts at a daily scale from end user perspectives should offer useful information on not only forecasts of rainfall totals but also forecasts of rainfall extremes. As rainfall is highly variable and climate changes have already caused an increasing trend in the frequency of extreme rainfall events [Allan and Soden, 2008; Donat et al., 2016; Papalexiou and Montanari, 2019], end users demand skilled S2S rainfall forecasts with emphasis on rainfall extremes to make smart decisions to match expected rainfall conditions [Curtis et al., 2017; Feng et al., 2018; Gonzalez et al., 2019].

Nowadays seasonal rainfall forecasts at a daily scale are most often available from general circulation models (GCMs), which are primarily designed to simulate large-scale climatic features for climate change studies. Despite continuous development of dynamical models, GCMs at present are yet unable to effectively simulate and forecast local daily rainfall conditions beyond two weeks because of several reasons, such as oversimplified model parameterization, imperfect initial conditions, and inadequate spatial resolution [Merryfield et al., 2020]. Statistical postprocessing has been used with any dynamic forecasting system before the end users can use seasonal rainfall forecasts from GCMs for real applications [Hawkins et al., 2013; Ines and Hansen, 2006; Maraun, 2013; Maraun and Widmann, 2018; Olsson et al., 2015; Schepen and Wang, 2014; Schepen et al., 2014; Veenhuis, 2013; Wilks, 2015].

The advances of GCM development promise more accurate prediction of extreme rainfall from longer lead times. Most extreme rainfall events are due to strong convective processes with horizontal scales up to a few kilometres, which currently cannot be directly simulated from any operational GCM. Nevertheless, GCMs provide a wealth of useful information on the atmospheric circulation and skilful predictions on large-scale climate features which have been related to climate extremes [King et al., 2020]. For example, Cowan et al. [2019] found that one ensemble member of Australian Community Climate and Earth-System Simulator (ACCESS) Seasonal model (ACCESS-S) [Hudson et al., 2017] did manage to forecast ~85% of the magnitude of record-breaking, not just extreme, rainfall associated with the northern Queensland floods of February 2019 [Bureau of Meteorology, 2019)].

To date, the extent to which statistical postprocessing can improve

* Corresponding author.
E-mail address: Ming.Li@data61.csiro.au (M. Li).

https://doi.org/10.1016/j.wace.2021.100384
Received 12 February 2021; Received in revised form 29 August 2021; Accepted 5 September 2021
Available online 6 September 2021

Table 1
Definition of extreme rainfall indices used for Australian purposes.

Short Name	Long Name	Definition	Unit
PRCPTOT0	Total precipitation	The total precipitation	mm
WD	Wet days	The count of days with daily precipitation ≥1 mm	day
R10mm	Heavy precipitation days	The count of days with daily precipitation ≥10 mm	day
R30 mm	Very heavy precipitation days	The count of days with daily precipitation ≥30 mm	day
Rx1day	Maximum 1-day precipitation	The maximum 1-day precipitation total	mm
Rx5day	Maximum 5-day precipitation	The maximum consecutive 5-day precipitation total	mm
R95p	Very wet day precipitation	The total precipitation when daily precipitation >95th percentile	mm
R99p	Extremely wet day precipitation	The total precipitation when daily precipitation >99th percentile	mm
PRCPTOT	Total wet day precipitation	The total precipitation on wet days (daily precipitation ≥1 mm)	mm
SDII	Simple daily intensity	The total precipitation divided by the number of wet days (daily precipitation ≥1 mm)	mm/day
CDD	Consecutive dry days	Maximum number of consecutive days with daily precipitation <1 mm	day
CWD	Consecutive wet days	Maximum number of consecutive days with daily precipitation ≥1 mm	day

the forecasts of rainfall extremes from the raw GCM outputs is not well understood. Traditional statistical postprocessing methods (such as simple bias correction and quantile matching (QM)) correct the overall discrepancy (such as overall biases and distribution mismatches) between observed and forecast rainfall. Because the forecast of extreme rainfall requires the conditional forecast distribution in the very far tail, traditional statistical postprocessing may not necessarily improve the forecast of rainfall extremes. Mylne et al. [2002] showed that statistical postprocessing may improve the forecasts of nonextreme events but degrade the forecasts of extreme events. Pioneering studies have attempted to develop statistical postprocessing methods with emphasis on forecasting extreme events by employing heavy-tailed distributions in replacement of a Gaussian distribution [Friederichs, 2010; Friederichs et al., 2018].

Li et al. [2020] developed a copula-based postprocessing (CPP) method for seasonal climate forecasts at a monthly scale. Inspired by model output statistics (MOS) [Glahn and Lowry, 1972], CPP employs copulas to derive the forecast distribution conditioning on the raw GCM outputs and other relevant predictors. More specifically, CPP firstly applies the probability integral transform to transfer original data (e.g. observations and raw forecasts) to "pseudo" data following a standard uniform distribution and then characterises the dependence structure of observations, raw GCM forecasts and other relevant predictors in the pseudo data space as a joint uniform distribution by copulas. To post-process seasonal rainfall forecasts at a daily scale, *Li and Jin* [2020]

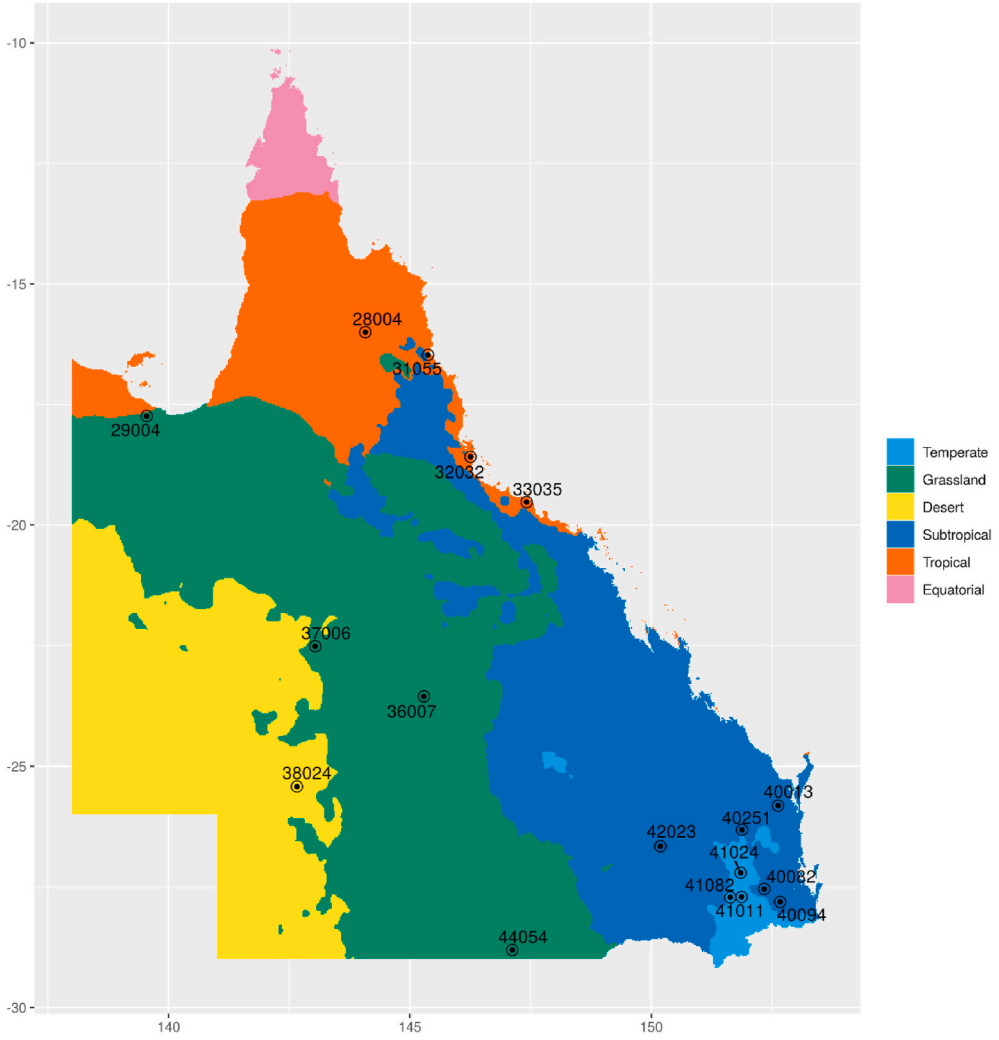

Fig. 1. The location of 17 rainfall stations in Queensland Australia labelled with station code in a map of Australian climate zones.

Table 2
A summary of 17 test rainfall stations, including station code, station name, latitude, longitude, annual climatology rainfall (calculated from 1961 to 1990) and daily maximum rainfall during the test period (calculated from 1990 to 2012).

Station code	Station name	Lat	Lon	Climate region	Annual rainfall (mm)	Daily maximum rainfall (mm)
28004	PALMERVILLE	−16.00	144.08	Tropical	1 021.7	307.0
29004	BURKETOWN POST OFFICE	−17.74	139.55	Grassland	761.1	366.3
31055	MOSSMAN SOUTH ALCHERA DRIVE	−16.47	145.37	Tropical	2 227.2	343.6
32032	MACKNADE SUGAR MILL	−18.59	146.25	Tropical	2 102.8	339.8
33035	KALAMIA ESTATE	−19.52	147.42	Tropical	1 061.8	200.7
36007	BARCALDINE POST OFFICE	−23.55	145.29	Grassland	488.7	121.2
37006	BLADENSBURG	−22.52	143.04	Grassland	378.5	181.0
38024	WINDORAH EVAP	−25.42	142.66	Desert	269.6	113.7
40013	BAUPLE	−25.82	152.62	Subtropical	1 103.2	253.6
40082	UNIVERSITY OF QUEENSLAND GATTON	−27.54	152.34	Subtropical	838.9	182.0
40094	HARRISVILLE POST OFFICE	−27.81	152.67	Subtropical	892.4	135.0
40251	WONDAI POST OFFICE	−26.32	151.88	Subtropical	792.0	102.0
41011	CAMBOOYA POST OFFICE	−27.71	151.87	Temperate	741.7	103.6
41024	DOCTORS CREEK	−27.21	151.85	Temperate	756.3	126.4
41082	PITTSWORTH	−27.72	151.63	Temperate	730.4	123.7
42023	MILES POST OFFICE	−26.66	150.18	Subtropical	667.9	170.4
44054	MULGA DOWNS	−28.81	147.12	Grassland	420.9	93.2

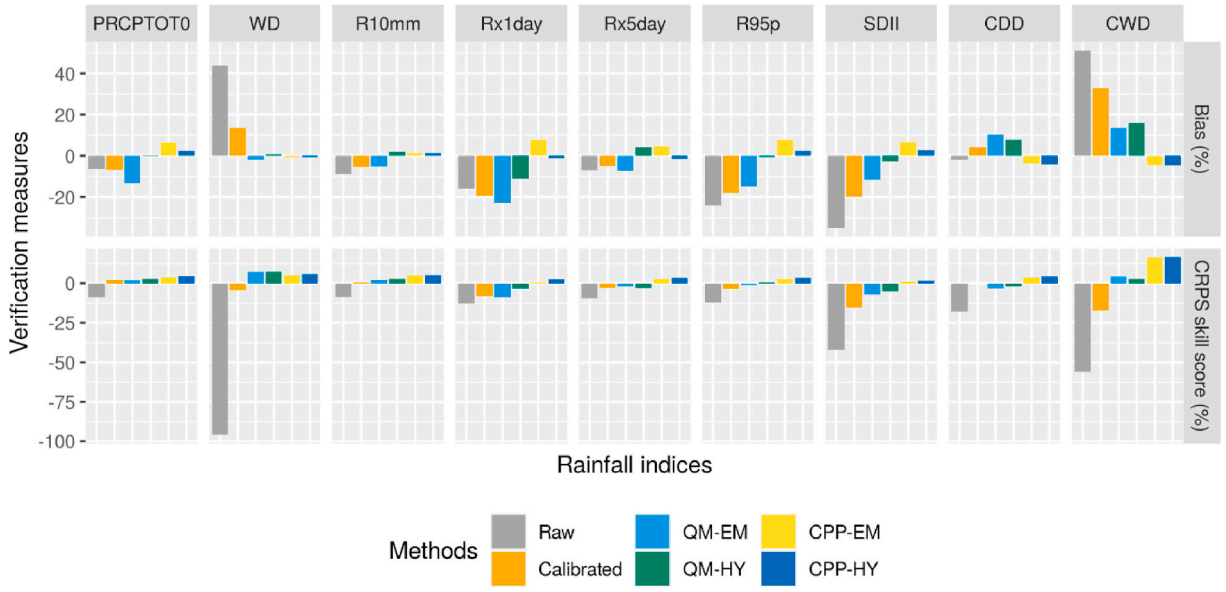

Fig. 2. The overall forecast bias (based on ensemble median) and CRPS skill score of 9 extreme rainfall indices from the raw ACCESS-S1, the calibrated ACCESS-S1 and four trialled postprocessing methods (including QM-EM, QM-HY, CPP-EM and CPP-HY). The verification measures are calculated from all events pooled from initial dates on the 1st day of each month and all stations.

treated daily rainfall as a left-censored random variable and revised the original CPP with a new likelihood estimation and simulation procedure to overcome the problem from daily rainfall with numerous zero occurrences. From several case studies in Australia [Li and Jin, 2020; Li et al., 2020], CPP has been shown to lead to more skilful seasonal rainfall forecasts than simple postprocessing methods such as bias correction and quantile mapping at both monthly and daily scales. In addition to the improvement of forecast skill, CPP, designed as a fully calibrated method, can further improve forecast reliability. A review of the existing CPP postprocessing method can be found in Appendix A.2.

However, the existing development of CPP does not take extreme events into consideration. The use of copulas in CPP requires the estimation of the probability distribution of each climate variable. In the current implementation of CPP, the probability distribution of daily rainfall is estimated nonparametrically by an empirical distribution function that takes values only from the range of samples. As a result, the current CPP cannot forecast any extreme event that has not been observed in the training period. As both the occurrence and magnitude of extreme rainfalls are expected to increase with a changing climate in the future [Stott, 2016], the use of an empirical distribution function to fit a daily rainfall distribution is a potential hurdle for CPP to further improve the forecast of extreme rainfall and to forecast events that have never been recorded (i.e. record-breaking).

To adapt CPP to forecast extreme rainfall, we need to find a better way to characterise the daily rainfall distribution. Although daily rainfall can be classically modelled by a single theoretical probability distribution, such as Gaussian after transformation [Cecinati et al., 2017; Cressie, 2015], exponential [Stern and Coe, 1984; Todorovic and Woolhiser, 1975] and gamma [Wilks, 2020], we follow a more popular approach to use a mixture of probability distributions (also known as a hybrid distribution) to characterise low to medium and extreme rainfall separately. Hybrid probability distributions have been used to represent the full range distribution of daily rainfall for many applications, such as simulating rainfall [Li et al., 2012], climate projection [Cannon et al., 2015] and rainfall intensity modelling [Naveau et al., 2016]. These hybrid rainfall distributions are commonly taken from a mixture of two theoretical probability distributions, such as two exponential distributions [Wilks, 1998; 1999; Woolhiser and Roldan, 1982], one gamma and

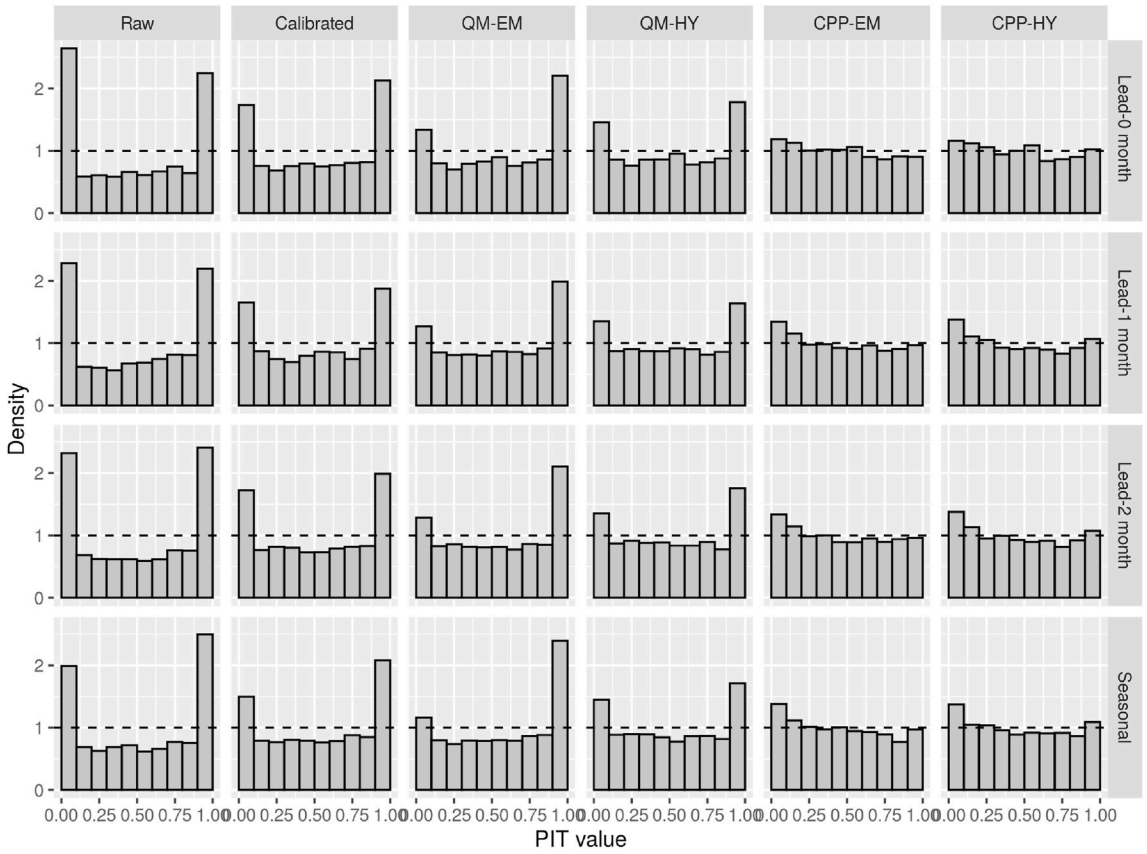

Fig. 3. The PIT histogram for rainfall total (PRCPTOT0) forecasts from the raw ACCESS-S1, the calibrated ACCESS-S1 and four trialled postprocessing methods (including QM-EM, QM-HY, CPP-EM and CPP-HY).

one generalized Pareto distribution [Furrer and Katz, 2008; Hundecha et al., 2009; Vrac and Naveau, 2007], and one exponential and one generalized Pareto distribution [Li et al., 2012]. Extreme value distributions have been used in the design of the above-mentioned hybrid distribution for rainfall, and the generalized Pareto distribution is a popular choice to fit the distribution of extreme rainfall [Koutsoyiannis, 2004a; b]. This is because it is the theoretical limiting distribution of rainfall exceeding a certain threshold [Coles, 2001] to fit the upper tail of the rainfall distribution in the peak-over-threshold framework. Shao et al. [2004] introduced the extended three-parameter Burr XII (EBXII) distribution as the theoretical distribution of data following the generalized Pareto distribution (GPD) after applying a power transformation. As a result, the EBXII distribution offers more flexibility and is less sensitive to model exceedances over thresholds than the GPD in a finite sample size (e.g. 30-year daily observations) [Shao et al., 2004]. The EBXII distribution has been used in different domains such as hydrology [Shao et al., 2004], civil engineering [Liu et al., 2018], reliability analysis [Jia et al., 2018] and operation research [Abouelmagd et al., 2017] to model extreme value over a threshold. In this study, we keep using the same empirical distribution as the existing CPP to represent low to medium rainfall because this arrangement allows most of the pseudo data in the training period to perfectly follow a uniform distribution and leads to robust model parameter estimation. Given the relatively small sample size of the training dataset in this study, we choose the EBXII distribution to characterise extreme rainfall.

The remainder of this manuscript is organised as follows. In Section 2, the modification of CPP with a hybrid distribution is discussed. In Section 3, a case study of postprocessing ACCESS-S rainfall forecasts to 17 rainfall stations across the northeast Australian state of Queensland is described. Forecast skills on overall performance and rainfall extremes are reported in Section 4. Discussion and conclusions are made in Sections 5 and 6, respectively.

2. Methods

2.1. Daily rainfall distribution and a modified CPP

The daily rainfall distribution plays an important role in CPP to transfer data in the original space to standard uniform variables on [0,1]. In the current implementation of CPP, an empirical distribution F_0 is used to estimate the daily rainfall distribution from training data $x_1, x_2, \ldots, x_n$

$$F_0(x) = \frac{1}{n}\sum_{i=1}^{n} I(x_i \le x), \tag{1}$$

where I is an indicator function and n is the sample size of training data. By treating rainfall as a left-censored variable, the pseudo training data, i.e. the training data after applying a probability integral transform (PIT) based on empirical distributions in the training period, perfectly follows a standard uniform distribution which fulfils the model assumptions imposed by CPP. However, empirical distributions have two undesired properties: i. $F_0(x) = 1$ *for any* $x \ge \max(x_1, x_2, \ldots, x_n)$, ii. $F_0^{-1}(y) \le \max(x_1, x_2, \ldots, x_n)$ for any $0 \le y \le 1$.

Based on the simulation procedure given by (A4) and (A5) in Appendix A.2, the implications of these two unfavourable properties on CPP include: (a) Raw forecasts exceeding the maximum of raw forecasts in the training data based on the ACCESS-S1 hindcasts cannot be distinguished by CPP and will lead to identical conditional forecast distributions and thereby the same CPP forecasts; (b) CPP can only issue forecasts within the range of historical rainfall observations and cannot forecast any extreme rainfall events that have never happened in the training period. Due to the facts that (1) the training period is often short

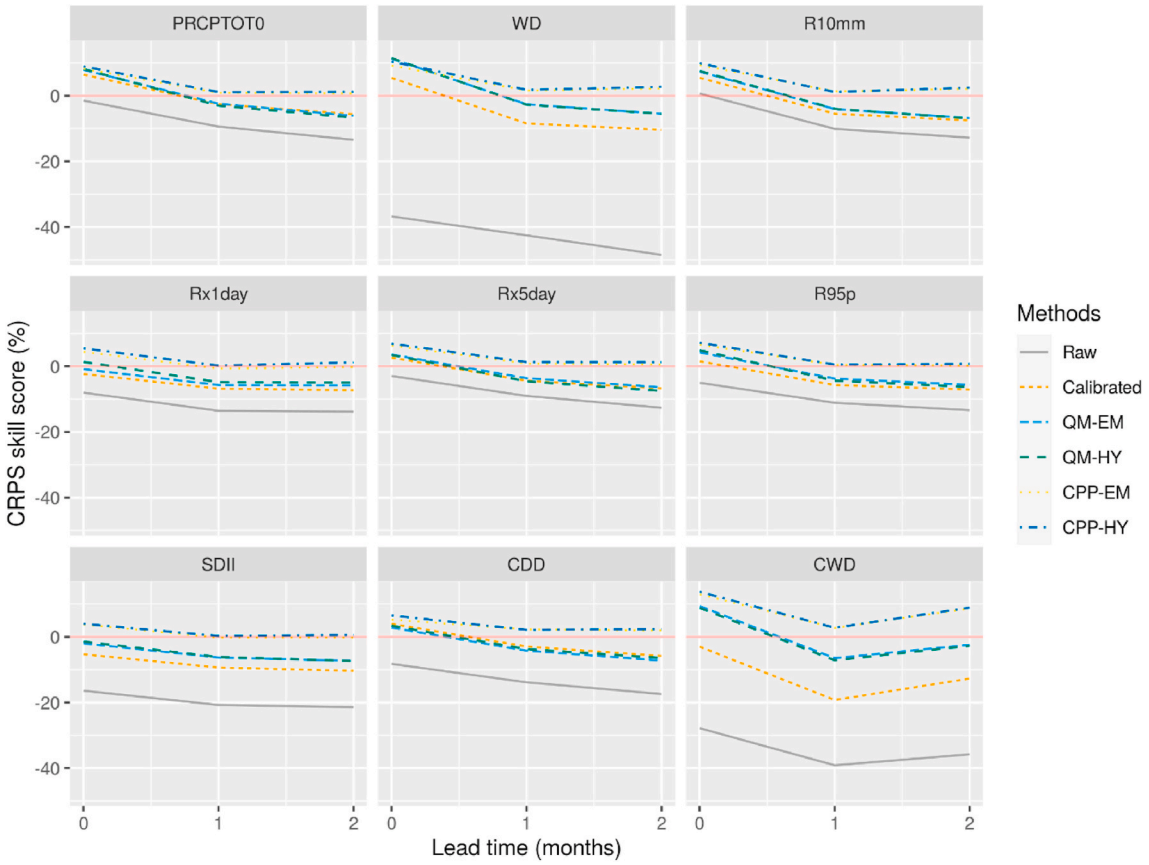

Fig. 4. CRPS skill scores of the forecasts of the selected nine extreme rainfall indices from the raw ACCESS-S1, the calibrated ACCESS-S1 and four trialled post-processing methods (including QM-EM, QM-HY, CPP-EM and CPP-HY) as a function of lead time.

and (2) the magnitude of extreme rainfall has become different due to multi-decadal variability, CPP may not be skilled at forecasting extreme rainfall.

To overcome the drawback of using an empirical distribution in CPP, we must consider a better way to fit the daily rainfall distribution. For this reason, this study proposes a hybrid probability distribution to model the distribution of low to medium and extreme rainfall separately. An empirical distribution is used for low to medium rainfall and daily rainfall above a large threshold is fitted by the EBXII distribution. Please see Appendix A.3for more technical details. The hybrid rainfall distribution is a mixture of discreet and continuous components and may look 'awkward' for a small sample size (e.g. ~700 observations in this study). However, with a small sample size, a hybrid distribution is more favourable than a pure empirical distribution or a pure parametric distribution, which often causes even more problems (such as failure to capture the long-tail or lack of overall fitting). We implemented some practical solutions to minimise the impact of a small sample size on the CDF estimation by using as many data points as possible (see Section 3) and applying a more robust parameter estimation (see Section 5.1).

A choice of $y_{ts} = 90\%$ is made for the case study described in Section 3 and this means that daily rainfall above the 90% historical quantile is fitted by an EBXII distribution. Rather than this empirical choice of the threshold, many formal procedures developed for estimating the threshold for GPD (including the Gerstengarbe and Werner plot [Gerstengarbe and Werner, 1989], the mean residual life plot [Coles, 2001], goodness-of-fit-based methods [Choulakian and Stephens, 2001], Hill-assumption-based methods [Beirlant et al., 2006]) can be potentially useful for estimating the threshold for the EBXII distribution. Nevertheless, Shao et al. [2004] showed that the EBXII distribution is less sensitive to the value of threshold than the GPD. The use of a formal procedure to determine the exceeding threshold for the EBXII distribution is out of scope for this study and would be a nice statistical topic for future research.

To adapt CPP to forecast extreme rainfall, this hybrid distribution replaces the empirical distribution in the existing CPP. In particular, the distributions of daily rainfall observations and daily rainfall forecasts are both fitted by the hybrid distributions. For convenience, we denote the modified (with the hybrid rainfall distribution) and existing (with the empirical rainfall distribution) CPP by CPP-HY and CPP-EM in the rest of this manuscript. Two QM methods (technical details can be found in Appendix A.1), including QM based on the empirical distribution (denoted by QM-EM) and QM based on the same hybrid distribution (denoted by QM-HY), will be compared with two CPP methods in the case study (Section 4).

2.2. Ensemble forecast verification

Seasonal rainfall forecasts (i.e. over three months) are evaluated in terms of both the mean and the extreme tail of forecast rainfall distributions (refer as to mean rainfall forecast and extreme rainfall forecast, respectively). The predictive skill of the mean rainfall forecast is assessed by the quality of seasonal total rainfall. The predictive skill of the extreme rainfall forecast is indicated by the performance of the forecast of 11 extreme rainfall indices (Table 1) originally defined by Expert Team on Climate Change Detection and Indices (ETCCDI: Zhang et al. [2011]) and modified for Australian purposes by the Bureau of Meteorology (BoM) (http://www.bom.gov.au/climate/change/about/extremes.shtml). For convenience, the accumulated rainfall (for measuring the skill of the mean forecast) is treated as an additional rainfall index, PRCPTOT0 (Table 1), and this makes that 12 rainfall indices are considered in this study. All indices are calculated at a seasonal scale to demonstrate the performance of seasonal forecasts.

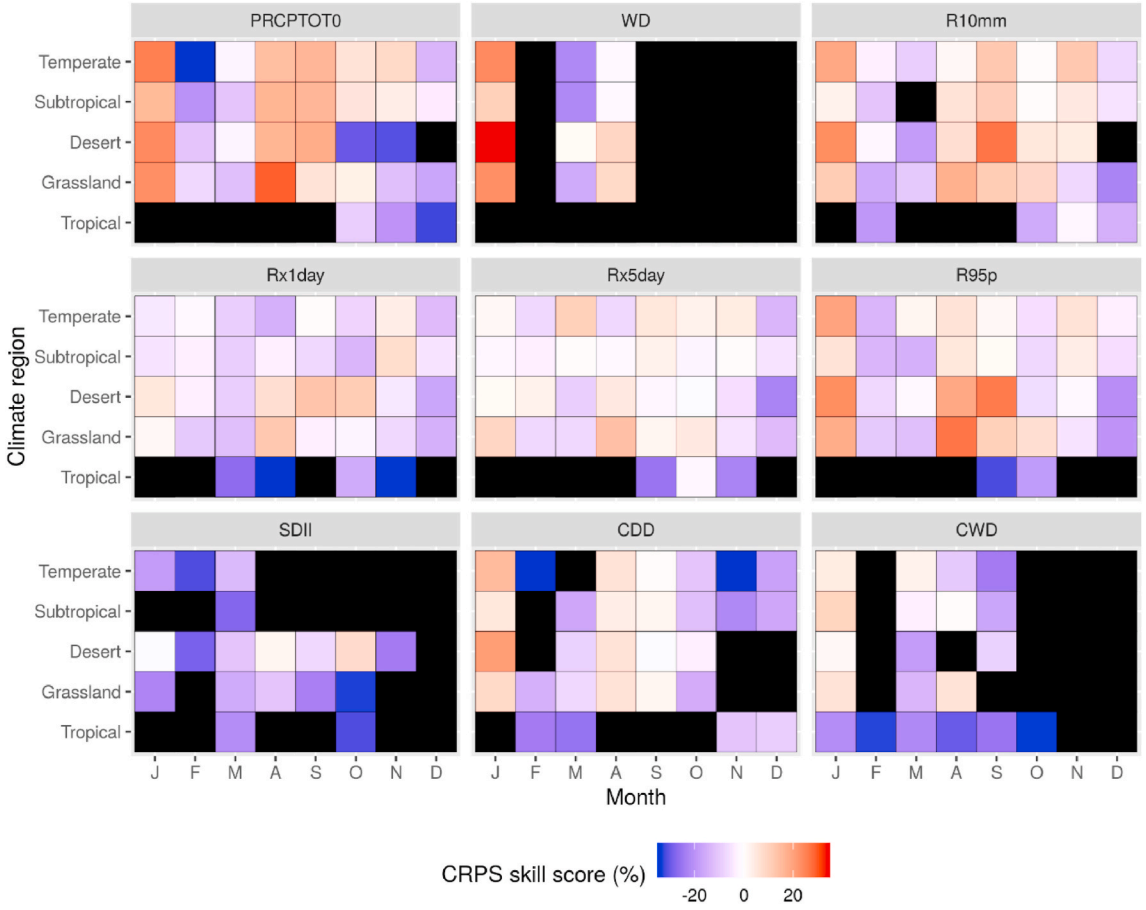

Fig. 5. CRPS skill scores of the lead-0 month forecasts of the selected nine rainfall indices from the raw ACCESS-S1 as a function of climate region and initialisation month. The numbers of rainfall stations used to calculate CRPS skill score are 3 (temperate), 5 (subtropical), 1 (desert), 4 (grassland) and 4 (tropical), respectively. Any CRPS skill score less than −35% is plotted in black. Note that dry months, including June, July and August, are not shown.

The skill of 12 rainfall index forecasts is evaluated in the framework of ensemble forecast verification. The overall skill of rainfall index forecasts is measured by the continuous ranked probability score (CRPS) [Grimit et al., 2006], which is the most popular proper score [Gneiting and Raftery, 2007] for ensemble forecast verification. CRPS is a highly aggregated measure and *Hersbach* [2000] suggested a decomposition of CRPS as accuracy, calibration and sharpness. To conveniently compare the forecast skill measured by CRPS, this study normalises CRPS as a CRPS skill score with respect to a reference forecast and makes it become unitless:

$$CRPS\ skill\ score = 1 - \frac{CRPS}{CRPS_{ref}}, \tag{2}$$

where $CRPS_{ref}$ is the CRPS calculated from the reference forecast. A CRPS skill score equal to 1 indicates a perfectly skilful forecast, and a positive CRPS skill score indicates a skilled forecast with respect to the reference forecast. In this study, the reference forecast is the 'cross-validation' version of the climatology forecast, which is formed by all observations from all years other than the forecast year. This reference forecast has been used for the same purpose in previous studies [Li and Jin, 2020; Li et al., 2020]. We will discuss a persistence forecast as an alternative reference forecast in Section 5.4.

To complement the skill verification by CRPS, three additional performance measures suggested by Engeland et al. [2010] are used to assess three individual forecast attributes including bias, calibration and sharpness. Bias is measured by the relative forecast bias calculated from ensemble mean, which is defined by the difference between ensemble mean and observations normalised by climatology mean. The sign of a relative forecast bias (negative or positive) suggests a tendency of underestimation or overestimation and the magnitude of this bias indicates the size of the deviation from the observations. Calibration refers to the statistical consistency between the assigned probability from an ensemble forecast and the frequency of an observed outcome and is measured by the uniformity of forecast probability integral transform (PIT) defined by $PIT = F_{ens}(x_0)$, where F_{ens} is the cumulative probability distribution from an ensemble forecast and x_0 is the observed value. In the case of perfectly calibrated forecasts, PIT follows a standard uniform distribution on [0,1]. The uniformity of the PIT histogram indicates whether an ensemble forecast is calibrated or not. Furthermore, calibration is quantified by the α-index [Hersbach, 2000] defined by

$$\alpha = 1 - \frac{2}{n}\sum_{t=1}^{n}\left|PIT_t^* - \frac{t}{n+1}\right|, \tag{3}$$

where PIT_t^* is the sorted PIT value in ascending order. The α-index is based on the distance between PIT values and the corresponding standard uniform variates, with the α-index being 1 suggesting a perfectly calibrated forecast. Sharpness is related to the spread of an ensemble forecast and an ideal ensemble forecast maximises the forecast sharpness subject to calibration [Gneiting et al., 2007]. In this study, sharpness is measured by the average width of the 90% confidence intervals (AWCI) and a smaller AWCI suggests better sharpness.

3. Stations and data

This study selects 17 rainfall stations in Queensland to test the performance of various seasonal forecasts during the test period of

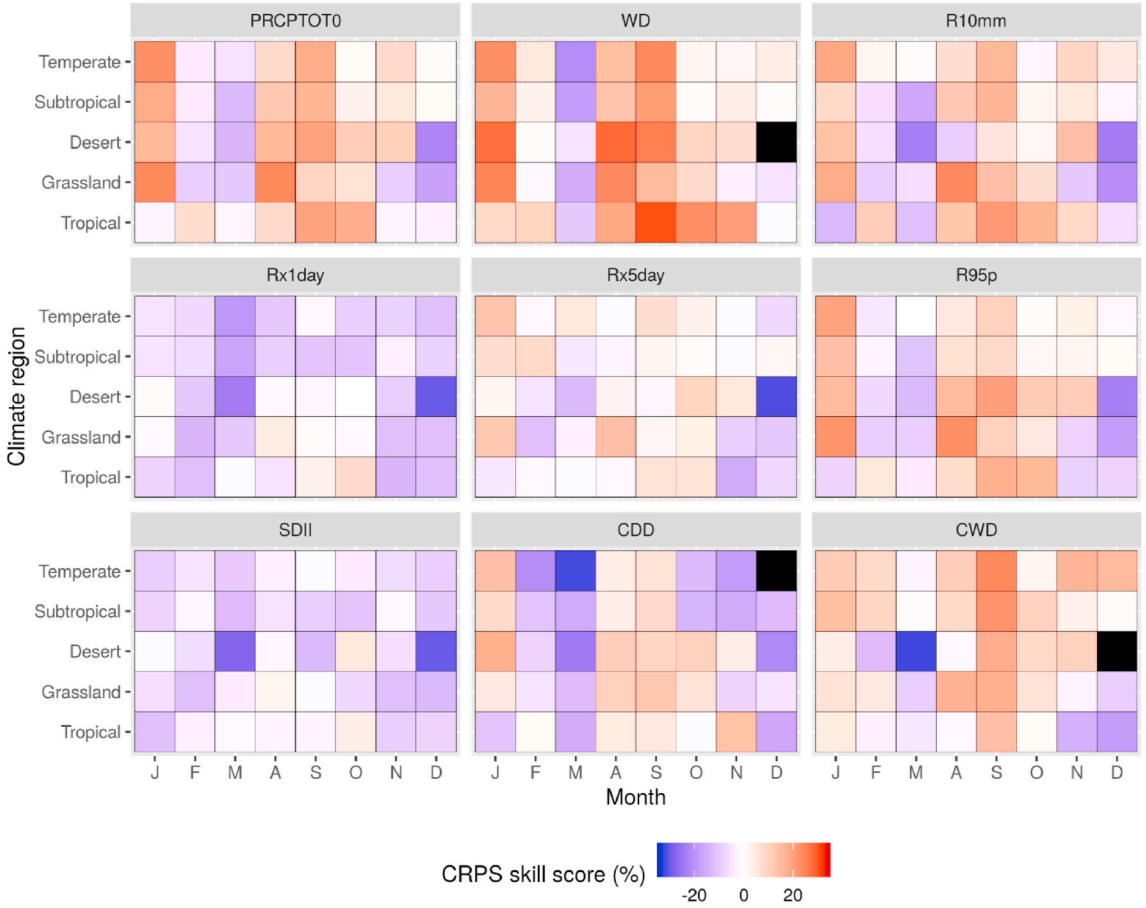

Fig. 6. As for Fig. 5, except for QM-EM.

1990–2012. Queensland rainfall varies greatly from one year to another and shows no clear long-term (1900–2010) trends in annual total rainfall [Queensland Government, 2012]. After drought years in the late 1990s and early 2000s, Queensland experienced one of the wettest years in history in 2010/2011. The test period covers December 2010 when heavy rainfall resulted in catastrophic flooding and multi-billion-dollar economic losses [Australian Business Roundtable for Disaster Resilience and Safer Communities, 2016]. Fig. 1 illustrates the location of these 17 rainfall stations in an official Bureau of Meteorology Australian climate zone map and Table 2 provides a summary of the test stations including the climatology annual rainfall (calculated from 1961 to 1990) and the maximum daily rainfall during the test period.

Raw seasonal rainfall forecasts at a daily step for these selected stations are obtained from the ACCESS-S version 1 (ACCESS-S1) at the corresponding nearest grid points. ACCESS-S1 is an advanced dynamical forecast modelling system developed by a collaboration between the BoM and the UK Meteorological Office. Hudson et al. [2017] showed that ACCESS-S1 obtains overall skill improvement on rainfall, maximum temperature (Tmax) and minimum temperature (Tmin) over Australia on multi-week scales compared to its predecessor Predictive Ocean Atmosphere Model for Australia (POAMA) [Hudson et al., 2017]. The hindcasts of ACCESS-S1 rainfall outputs are available at 60 km grid points and at the entire test period from 1990 to 2012. Each hindcast issued on the first day of a calendar month with up to 93-day lead times is considered in the forecast verification of this study. The SILO patched point data (available free of charge from Queensland Government, licensed under Creative Commons Attribution 4.0, at https://legacy.longpaddock.qld.gov.au/silo) are used as rainfall observations at stations. The entire ACCESS-S1 hindcasts from 1990 to 2012 are used in skill verification and the leave-one-year-out cross-validation is performed to check how the forecast skills will be generalized to an independent period without potentially overfitting the data in the hindcast period. When fitting the hybrid rainfall distributions, we use the historical observations (1961–2012) for the daily rainfall distribution and the hindcasts (1990–2012) for daily rainfall forecasts from ACCESS-S1 with the exclusion of one cross-validated year.

4. Results

4.1. Overall forecast performance

Fig. 2 presents the bias and CRPS skill score of seasonal rainfall index forecasts from the raw ACCESS-S1 and four trialled postprocessing methods. The verification measures in Fig. 2 are calculated from all events pooled from all the selected initial dates on the 1st day of each month and stations. For a better presentation, we leave the results for three indices (R30 mm, R99P and PRCPTOT) in the supplementary material. The raw ACCESS-S forecast leads to small biased forecasts (about 5% underestimation) of seasonal total rainfall (PRCPTOT0) but substantial overestimation of any indices related to light rainfall such as WD (~45%) and CWD (~50%) and substantial underestimation of indices related to heavy rainfall such as R30 mm (~30%), Rx1day (~15%), R95P (~25%) and R99P (~25%). The problem of the raw ACCESS-S1 to produce too many light and moderate precipitations but too few extreme precipitations is also found in many other GCMs [Dai, 2006; Kamiguchi et al., 2006; Stephens et al., 2010]. This problem is also known as the 'drizzle-day syndrome' that is related to the issue that GCMs calculate precipitation at grid points considerably larger than actual clouds and smooths out zero rainfall and extreme rainfall at a local scale. None of the seasonal rainfall indices from the raw

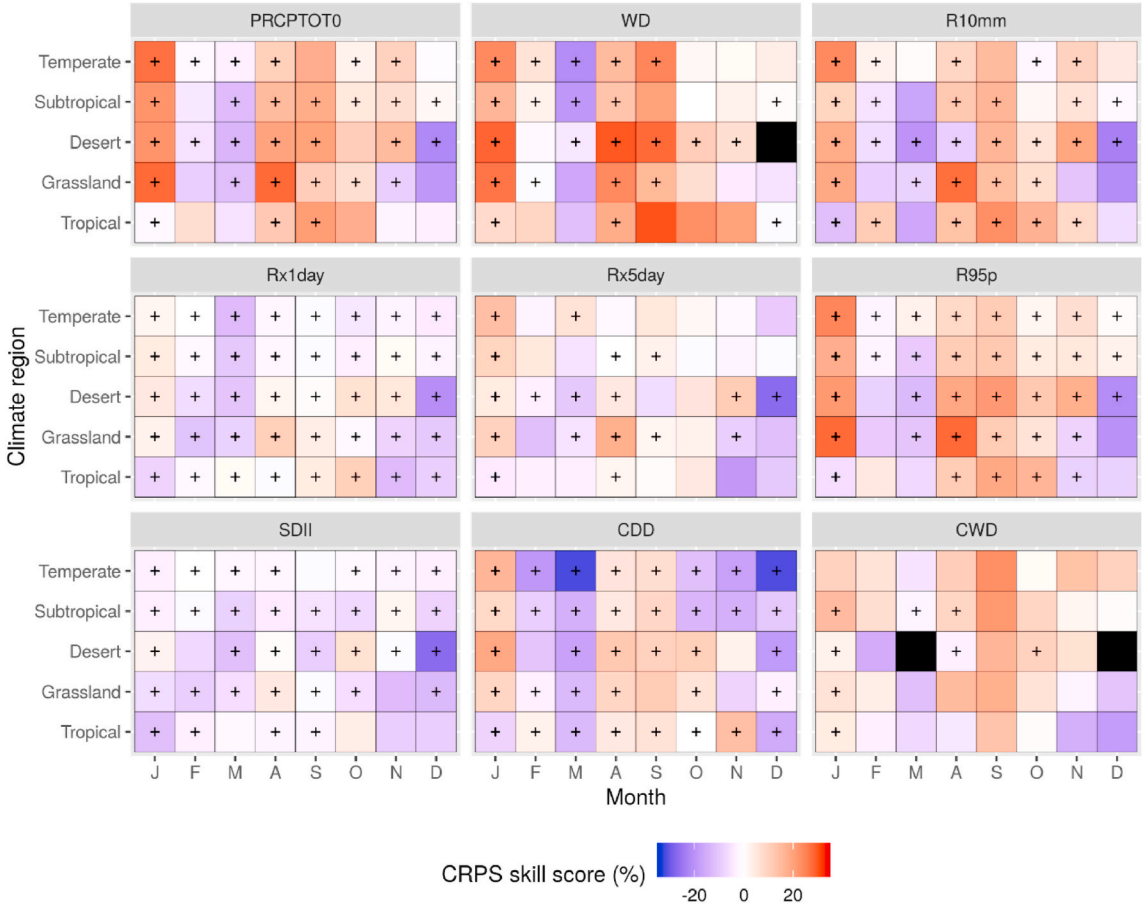

Fig. 7. As for Fig. 5, except for QM-HY. The "+" signs indicate the cases where QM-HY shows an improvement over QM-EM.

ACCESS-S1 are skilful with respect to the reference climatology forecast (suggested by all negative CRPS skill scores), partially due to the 'drizzle-day syndrome'.

All postprocessing methods trialled in this experiment effectively correct the 'drizzle-day syndrome' from the raw ACCESS-S1 and produce more realistic light and medium rainfalls. For light rainfall, the 45% bias of WD from the raw ACCESS-S1 is slashed to near zero by the four postprocessing methods. For medium rainfall, the bias of R10mm from the raw ACCESS-S1 is reduced at least by half when the postprocessing methods are used. Seasonal rainfall indices related to light and medium rainfall, including WD, R10mm and CWD, are overall skilful for all the postprocessing methods.

The improvement on heavy and very heavy rainfall is evident but generally smaller than the improvement on light and medium rainfall for postprocessing methods. All postprocessing methods indeed yield greater CRPS skill scores and thus better overall forecast performance than the raw ACCESS-S1 hindcasts for heavy and very heavy rainfall related indices (e.g. R30 mm, Rx1day, Rx5day, R95p and R99p). However, both QM based postprocessing methods hardly lead to positive CRPS skill scores for these heavy and very heavy rainfall indices, suggesting degraded forecasts relative to the reference climatology forecast on heavy and very heavy rainfall. Two CPP based postprocessing methods, instead, outperform the reference climatology forecast in terms of heavy and very heavy rainfall as indicated by almost all positive CRPS skill scores. It is evident that the two hybrid-distribution based methods lead to smaller biases and better forecast skill than the corresponding empirical-distribution based methods. The improvement from the hybrid distribution on R99p is more profound than that on R95p. This suggests that the hybrid distribution fits the far tail of the rainfall distribution better than the empirical distribution, as evidenced in this cross-validation analysis.

All the postprocessing methods improve the overall forecast skill of seasonal rainfall totals. The CRPS skill score of PRCPTOT0 turns positive when any trialled postprocessing method is applied. However, the forecast bias of PRCPTOT0 can be adversely increased by postprocessing if the rainfall distribution is only fitted by the empirical distribution. For example, QM-EM leads to greater negative biases for PRCPTOT0 as well as Rx1day and R99p. It is mainly caused by the empirical distribution underestimating any events in the validation period exceeding the maximum rainfall in the training period. The use of the hybrid distribution solves the problem of adverse bias increase and further improves the CRPS skill score. Another advantage of the CPP based methods is the improvement of forecast calibration. Fig. 3 presents the PIT histogram for seasonal rainfall totals from different methods. The U-shape PIT histogram for the raw ACCESS-S1 forecast clearly suggests a lack of calibration and under-dispersion. Unfortunately, the QM based methods, aiming for bias correction, keep the U-shape PIT histogram and therefore do not warrant improvement on forecast calibration. The CPP based methods lead to much more uniform PIT histograms and therefore better forecast calibration. The use of the hybrid distribution marginally improves the forecast calibration (more profound for QM than for CPP). Interestingly, the improvement of calibration is not at a cost of sharpness and Supplementary Fig. 1 shows that CPP-HY leads to the smallest average width of 90% forecast interval for seasonal rainfall totals. Although Fig. 1 in the supplementary materials shows that QM-EM compares quite well to CPP-HY in terms of sharpness and show a smaller spread than CPP-EM, QM-EM is not preferred as the sharper forecasts of QM-EM are caused by under-dispersion confirmed by Fig. 3. The longer-tailed boxplots for the postprocessing results in Fig. 1 in the supplementary materials suggest that the under-dispersion of raw

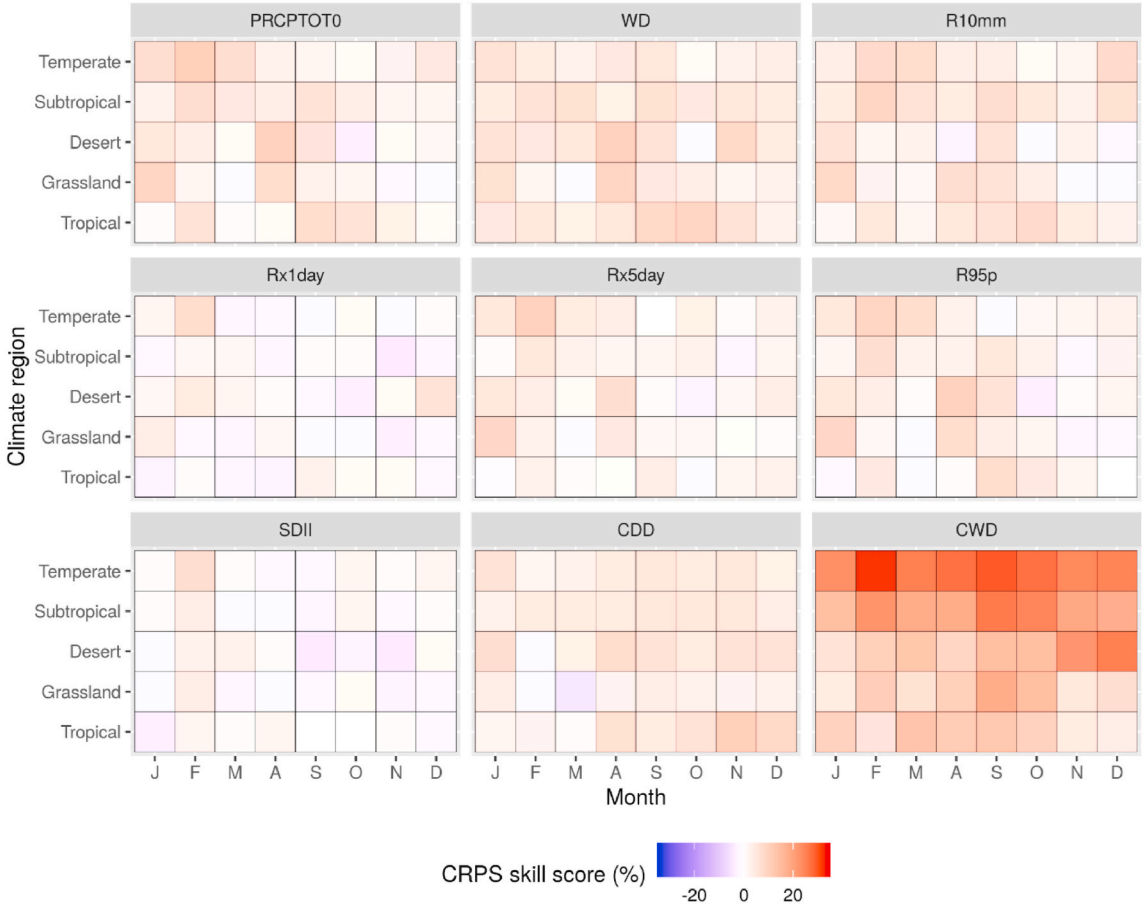

Fig. 8. As for Fig. 5, except for CPP-EM.

forecasts is somewhat corrected by the postprocessing methods.

4.2. The impact of lead time on forecast performance

Fig. 4 presents the CRPS skill scores calculated from the forecasts of 12 rainfall indices at a lead time of 0–2 months and illustrates how the forecast skill varies across different lead times. We have the following three main findings:

(1) It is a clear trend that the CRPS skill scores decrease with growing lead times for each method and in all the indices except CWD. This is very much expected as forecast skill/accuracy deteriorates with increasing lead times due to information loss. The small skill increases for CWD from lead time 1 month to lead time 2 month may be explained by: (a) CWD follows a long-tailed distribution and forecast skill can be sensitive to a single wet event with a number of consecutive wet days and (b) the verification period from lead time 1 month is Feb/1990 to Jan/2013, which is slightly different from that from lead time 2 month (i.e. Mar/1990 to Feb/2013).
(2) The CRPS skill scores for CPP-HY are positive at lead time one-month for all rainfall indices. To further explore this finding, we calculate the 5th and 95th percentiles of CRPS skill scores for CPP-HY at lead time 0-month with 2000 bootstrap sampling [Lim et al., 2011] (Supplementary Fig. 3). As the 5th percentiles for all rainfall indices are positive, CPP-HY is proven to be significantly skilful at lead time 0-month.
(3) The CRPS skill scores are negative (for two QM base methods) or nearly zero (for two CPP based methods) for almost all rainfall indices (except for WD, CDD and CWD) when lead times exceed one month. This finding supports that the subsequent post-processing methods are generally more skilful than climatology at short lead times (few weeks to one month), which also has been found in previous studies for mean forecast [Li and Jin, 2020; Li et al., 2020] and for extreme forecast [King et al., 2020]. CPP based methods provide a safety net of forecast performance (that is, at least as skilful as climatology) for rainfall totals as well as rainfall extremes at long lead times. As lead times increase, the effect of using the hybrid distribution becomes insignificant and forecast skill mainly depends on the parent postprocessing method (i.e. QM or CPP). Fig. 3 shows that forecast calibration, unlike forecast skill, is generally not affected by lead times.

4.3. The impact of climate region and initialisation date on forecast performance

To investigate how the performance of rainfall index forecasts are associated with locations and seasons, Fig. 5-Fig. 9 present the CRPS skill score of the lead-0 month rainfall index forecasts from the raw ACCESS-S1 and four postprocessing methods as a function of climate regions and initialisation month. The CRPS skill scores in these figures are calculated by aggregating all stations in the target region and calendar month. The discussion focuses on the lead-0 month forecasts as we have already shown in Section 4.2 that all trialled forecasts are generally either degraded or neutral beyond 1-month lead time. The performance for the raw ACCESS-S1 forecast, as shown by Fig. 5, is on average worse for the tropical region (particularly in summer) than for other regions. This poor performance is likely related to the challenge to simulate summer tropical cyclones by GCMs [Sperber and Palmer, 1996; Wang et al., 2004] and to the trouble with the representation of extreme

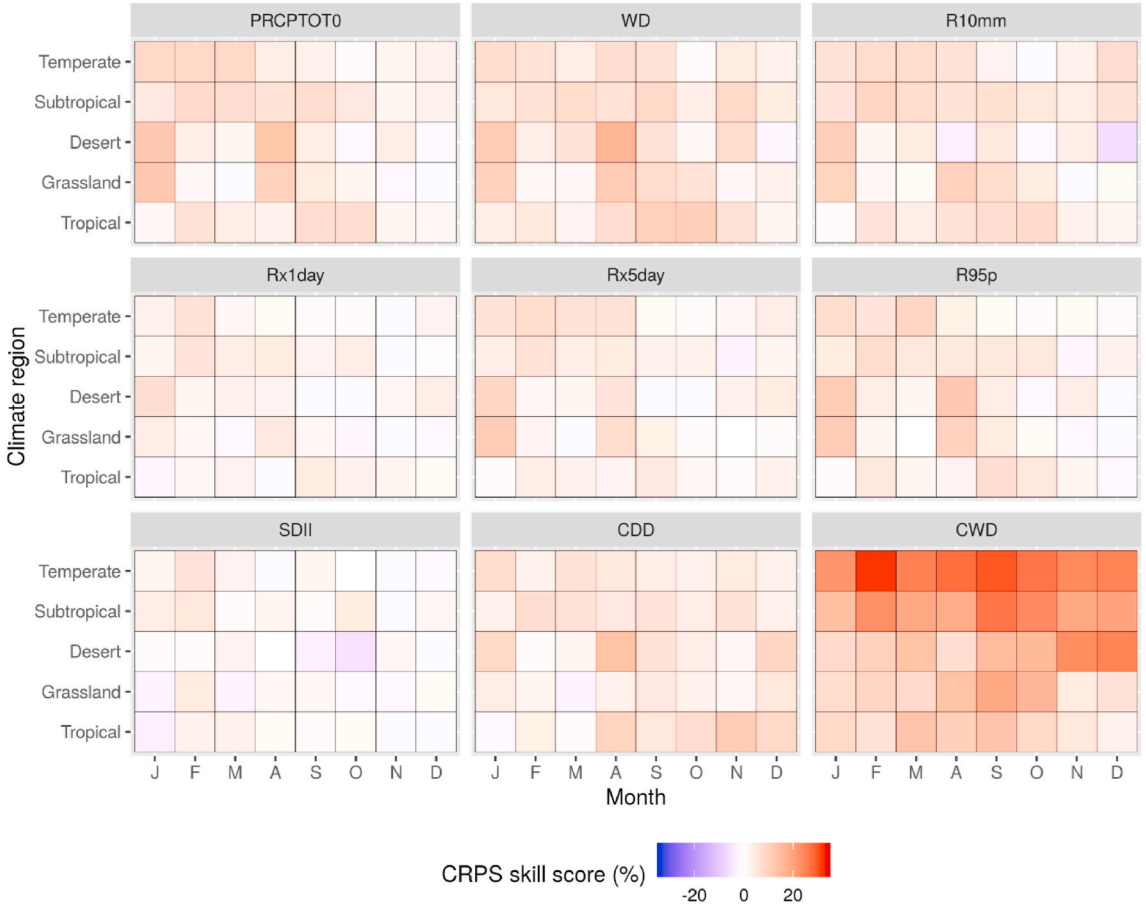

Fig. 9. As for Fig. 5, except for CPP-HY.

rainfall related to the active phase of the Madden-Julian oscillation (MJO) over northern Australia [Lim et al., 2019; Marshall and Hendon, 2015]. The raw model can issue skilful forecasts for total monthly rainfall (PRCPTOT0) relative to the climatology forecast from June to September except for the tropical region. In contrast, the skill of the raw model to forecast extreme rainfall is generally very limited and only can be found in some drier regions (such as desert and grassland) and months (such as July, June and August). Note that winter rainfall in July, June and August is uncommon for desert and grassland regions of Queensland. Our findings are consistent with those from King et al. [2020] that concluded that ACCESS-S1 tends to (1) perform less skilfully "in austral summer than other seasons" and (2) perform more skilfully in the "interior of the continent, particularly in the dry season".

When comparing Fig. 5 with Fig. 6, QM-EM significantly improves the forecast performance in the tropical region especially for total rainfall (indicated by PRCPTOT0) and light-to-medium rainfall (indicated by R10mm and WD). As foreshadowed by the overall performance assessment in Section 4.1, QM-HY can further improve the forecast performance especially for heavy and very heavy rainfalls (indicated by Rx1day, R95p and R99p) in almost all climate regions and months (Fig. 7). QM based methods generally have a greater improvement in the tropical region than in other regions.

Fig. 8 and Fig. 9 show that the forecast skill of CPP based methods is relatively robust and is consistently positive or small negative for different climate regions, months and rainfall indices. On one hand, CPP based methods do not always produce more skilful rainfall index forecasts than QM based methods. For instance, QM based methods exhibit better forecast skill (5%–10% greater CRPS skill score) than CPP based methods in forecasting R95p and R99p in desert and grassland regions in September and October, though CPP based methods perform better in the same regions in March and April. On the other hand, CPP based methods rarely issue degraded forecasts for both rainfall totals and rainfall extremes. It implies that the safety net of CPP based methods mentioned in the overall assessment is also applied to each individual region and month.

5. Discussion

To complement the main results, we provide further discussions on technical aspects of the proposed postprocessing methods in this section, which may be particularly interesting for further model developments.

5.1. *Fitting extreme rainfall distributions*

The main results in Section 4 have shown that skilful extreme rainfall forecasts require a proper model to characterise extreme rainfall. The EBXII distribution is chosen in this study to fit the distribution of rainfall exceeding a certain threshold, though extreme value distributions such as the Generalized Extreme Value (GEV) and Generalized Pareto Distributions (GPD) have been more commonly used to characterise the distribution of extreme rainfall [Koutsoyiannis, 2004a; b]. GEV does not serve the purpose of this study properly as it is often used within the block maxima framework (for example, for modelling the distribution of annual maximum rainfall). GPD, instead, could be another candidate parametric distribution to fit the upper tail of the rainfall distribution in the peak-over-threshold framework. Furthermore, the choice of a parameter estimation is as important as the choice of a probability distribution for the goodness-of-fit of the upper tail distribution. The maximum likelihood estimation (MLE) could replace the least square estimation as an alternative parameter estimation method.

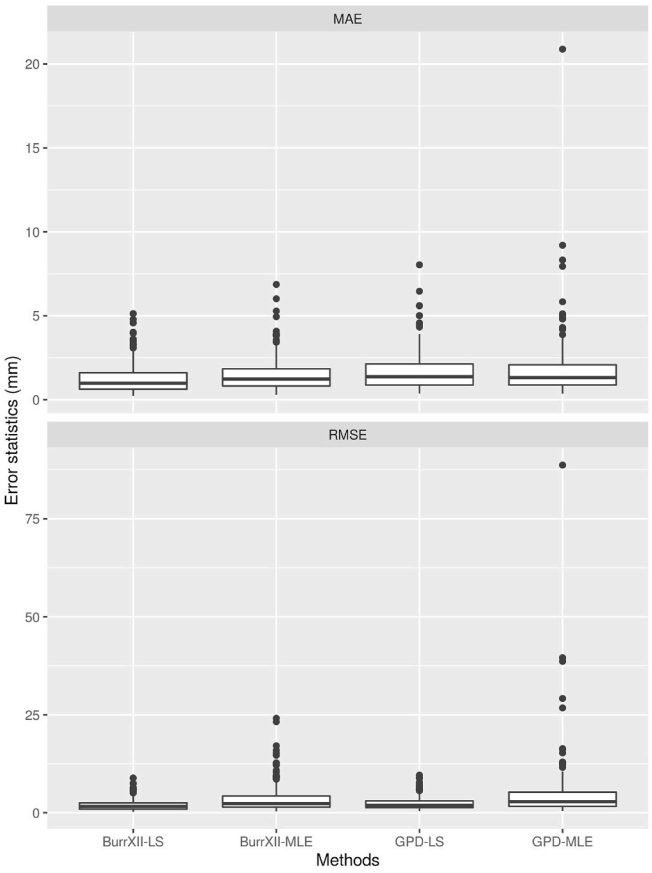

Fig. 10. A comparison of fitting the upper tail of rainfall distribution from two probability distributions (EBXII and GPD) and two parameter estimation (LS and MLE). Two error statistics, including mean absolute error (MAE) and root mean square error (RMSE), are calculated across 17 stations and 12 calendar months.

A small sample size of extreme rainfall makes us prefer EBXII and LS to GDP and MLE in this study. Because the rainfall distribution is fitted for each individual month and there are only 23-year observations available in the ACCESS-S1 hindcasts, the maximum number of observations exceeding the 90% quantile of historical rainfall is 23*31*10% = 71. In some dry months, zero value is observed in more than 90% of historical rainfall observations and this makes the effective sample size (which has to exclude zero rainfall) to fit the upper tail distribution even smaller. Fig. 10 presents a comparison of model fitting from two probability distributions (EBXII and GPD) and two parameter estimations (MLE and LS). Two error statistics, including mean absolute error (MAE) and root mean square error (RMSE), are calculated across 17 stations and 12 calendar months. EBXII and LS lead to smaller error statistics and thus better model fitting than GPD and MLE, respectively. In addition to the summary statistics, we compare the observed and fitted rainfall (by different distributions and with different estimation methods) for each individual station in the supplemental material and confirm that the EBXII distribution with LS estimated parameters fits best against observations. EBXII outperforms GPD because the additional parameter of EBXII provides a more flexible distributional shape to fit with a small sample size. LS naturally assigns more weight on very extreme observations in evaluating the objective function than MLE. The property of LS is favourable as it can fit very extreme observations (such as historical maximum rainfall) reasonably well. The best way to fit extreme rainfall distributions may be different for other regions. GPD and MLE might be appropriate when sufficient data are available. A similar analysis is recommended to check the model fitting from a specific probability distribution and parameter estimation before carrying out further postprocessing procedures. The new version of ACCESS-S model (ACCESS-S2) will provide more hindcasts covering from 1981 to 2018 and facilitate the estimation of daily rainfall distributions with a larger sample size.

5.2. *Postprocessing at grids or at stations*

Postprocessed forecasts at grid scales instead of station scales are often offered in place of raw forecasts for operational use. For example, the BoM produces a postprocessed version of the ACCESS-S1 hindcasts, which are calibrated to gridded observational daily rainfall at 5 km resolution by using the similar quantile-matching method described in Section 2.1. As shown in Fig. 2, the overall performance of the BoM calibrated forecast is, not surprisingly, better than the raw forecast for all rainfall indices with smaller biases and greater CRPS skill scores. Nevertheless, QM-EM (i.e. quantile mapping at stations) leads to marginally more skilful forecasts of extreme rainfall indices than the calibrated forecast (i.e. quantile mapping at grids). This suggests that postprocessing at stations can gain a little more confidence to forecast extreme rainfall at a station level than postprocessing at grids. Fig. 3 shows that the calibrated ACCESS-S1 has better forecast reliability than the raw counterpart but still suffers from under-dispersion. As postprocessing at stations may require substantially additional efforts, a decision whether to use postprocessed forecasts at grids directly from the model provider or not should be made based on a cost-benefit analysis.

5.3. *Impact of ensemble size*

It is understood that a large ensemble size is preferred to forecast extreme events as they happen so rarely. CPP estimates a conditional forecast distribution and can generate ensemble members as many as possible from this distribution. The convenience to generate a forecast with a desired ensemble size is an advantage of CPP, which also has been discussed in previous studies [Li and Jin, 2020; Li et al., 2020]. In this study, an ensemble size of 110 for CPP-HY and CPP-EM indeed contributes to the skill gain relative to the raw and QM forecasts, which only have 11 ensemble members. To understand the impact of ensemble size on the forecast performance, Fig. 11 presents the ensemble-size-adjusted CRPS skill scores [Ferro et al., 2008] of CPP-HY at lead time month 0 with five trialled ensemble sizes, including 11, 110, 242, 1 000 and 10000. The increase of the ensemble-size-adjusted CRPS skill scores is substantial when the ensemble size increases from 11 to 110 and becomes minor when the ensemble size further increases from 110. As larger ensemble sizes also require more storage space and more computing resources for further applications, an ensemble size of 110 looks a good option for CPP-HY similar to our previous studies [Li and Jin, 2020; Li et al., 2020]. Furthermore, combining hindcast members from different start dates on the same month increases ensemble size (for example, the new version of ACCESS model (ACCESS-S2) will have nine members per start date and six dates per month for the multi-week runs, as opposed to 11 members per start date and four dates per month in ACCESS-S1), and this may potentially benefit the improvement of forecast skill.

5.4. *Test against a persistence reference forecast*

A persistence forecast is an alternative reference forecast other than the climatology forecast, mainly for short-term forecasting, that is computationally inexpensive but may be sufficient for many applications. We generate a persistence forecast using the observations from the past 31 days prior to the initialisation date and compute CRPS skill scores of various forecasts with respect to this persistence forecast (Fig. 12). The raw forecast is not as skilful as the persistence forecast at lead time month 0, indicated by negative CRPS skill scores, except for R10mm and CDD. The QM and CPP based forecasts are obviously better

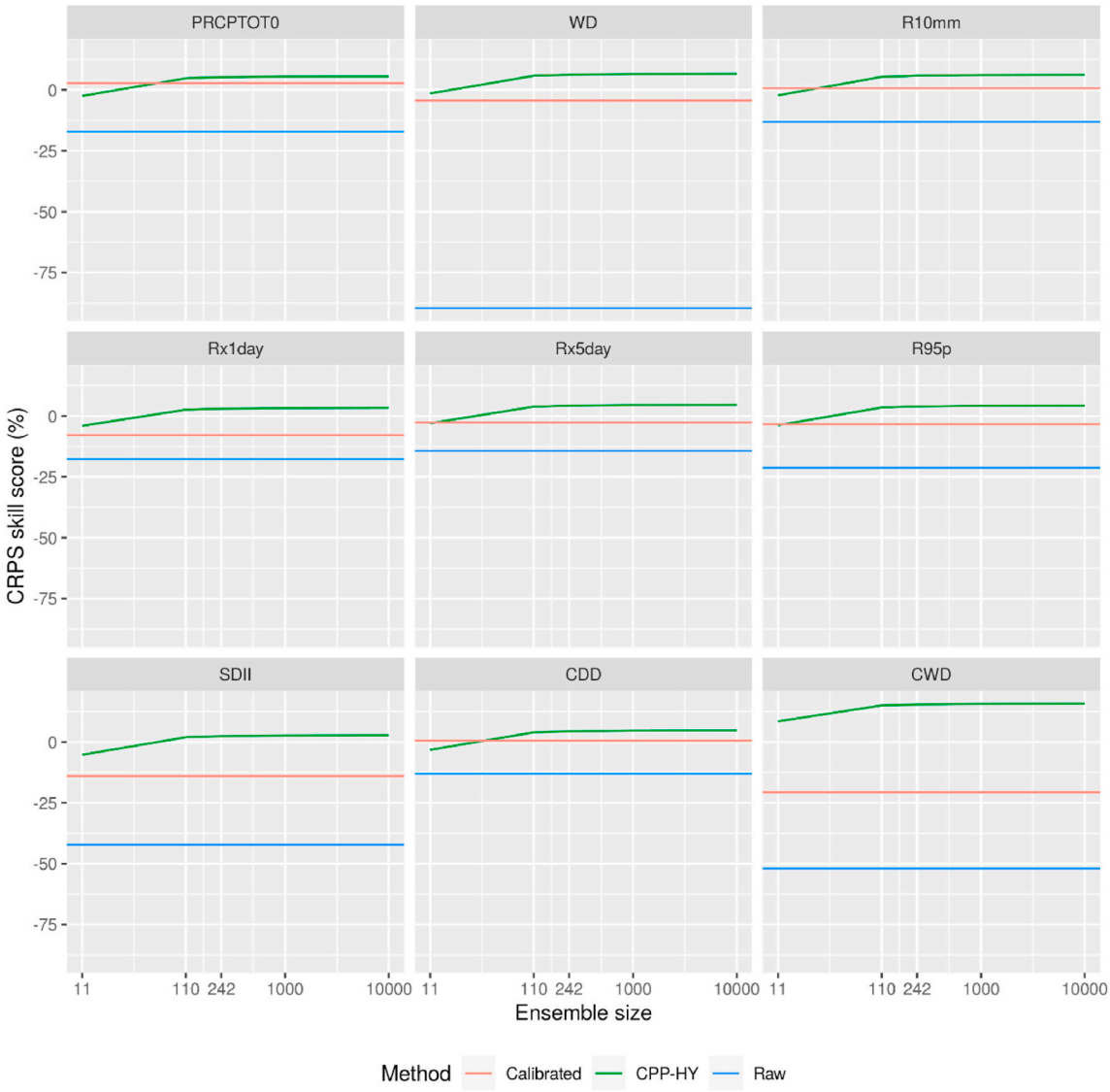

Fig. 11. CRPS skill score of nine selected extreme rainfall indices calculated from CPP-HY forecasts at lead time month 0 as a function of ensemble size. CRPS skill scores calculated from the raw and calibrated ACCESS-S1 forecasts are compared as references.

than the persistence forecast at lead time month 0. An increasing trend of CRPS skill scores with respect to the persistence forecast has been found. It is caused by the fact that this persistence forecast does not change with growing lead times and the actual performance of the persistence forecast decreases at longer lead times due to information loss. We would not recommend a persistence forecast for extreme events at any lead times.

5.5. *Skilful forecast lead times*

The motivation of this study is to compare the forecast skills of various postprocessing methods on subseasonal to seasonal rainfall extremes. QM and CPP based postprocessing methods improve skills with respect to the raw hindcast outputs at almost all lead times but have little (often <10% CRPS skill scores) or no skill with respect to the climatology reference forecast beyond the 0-month lead time (Fig. 4). The limited forecast skill is related to ACCESS-S1's skill in predicting rainfall (mean and extremes). Within the intra-seasonal scale, ACCESS-S1 can simulate climate drivers such as MJO, equatorial Rossby waves and Kelvin waves, and their simulated teleconnection to northern Australia. In contrast, beyond the intra-seasonal scale, ACCESS-S1's skill in predicting rainfall mainly comes from its ability to simulate El Niño-Southern Oscillation and the Indian Ocean Dipole. Therefore, ACCESS-S1 may not capture the teleconnection of these climate weather modes to northern Australia and no amount of postprocessing beyond 0-month lead time is likely to improve the raw model.

There is huge potential for postprocessing methods (such as CPP-HY) to be utilised to produce more skilful weekly rainfall forecasts for up to one month. This would be beneficial to end-users who can make quick decisions (i.e., moving cattle and possible early harvesting) if given forewarning of extremely heavy rainfall 2–3 weeks in advance. *Li and Jin* [2020] showed that CPP-EM can significantly improve the skill of daily rainfall forecast at the first two-week lead time. To assess the forecast skill of weekly extreme rainfall, we need to adapt extreme rainfall indices originally defined at the annual scale to the weekly scale. For example, CWD and CDD are not meaningful at the weekly scale, because continuous wet/dry days may exceed 7 days. Future research is recommended to validate the weekly extreme rainfall forecasts from postprocessing methods with appropriate extreme rainfall indices covering both extremely wet and dry events.

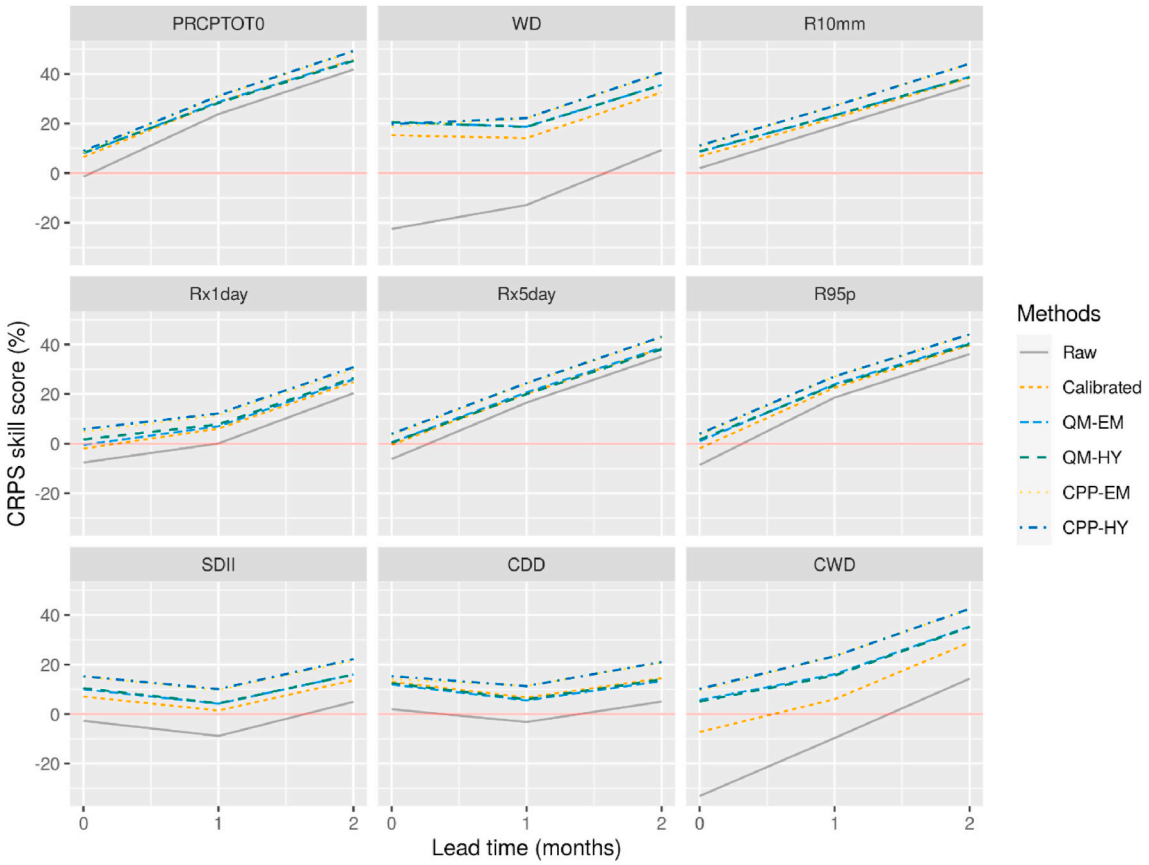

Fig. 12. As for Fig. 4, except for CRPS skill scores calcuated against a persistence forecast.

5.6. *Interpreting forecast skills*

Some difficult rainfall stations have been deliberately selected in this study. Burketown (29004) is an example, as this station is extremely dry (90% percentile being 0 mm) in the dry season but may experience extremely large rainfall (>300 mm) in the wet season. It is important to forecast extreme rainfall at such an arid or semi-arid station in the dry season although the probability is close to zero (only two extreme rainfall events with >300 mm/day during 1961–2010). The results for this case demonstrated that the proposed postprocessing method (CPP-HY) can provide a unified solution for difficult stations in the wet and dry seasons without separated modelling.

We also would reiterate that all the results are dependent on the selection of rainfall stations and recommend interpreting forecast skills in the context of model comparison. Because our case study is based on unevenly distributed rainfall stations and a limited study period, extra caution is needed in the case of generalizing our results to those rainfall stations that have not been tested in this study. We will test the proposed methods across the whole Australian continent and may develop a product for research use in coordination with the BoM.

5.7. *Future research*

We have shown that there is clear utility in applying the post-processing methods (particularly CPP-HY) to individual rainfall stations to boost the skill of predicting rainfall extremes out to one month. In future research, the postprocessed forecasts at representative stations can potentially be upscaled to a 5 km grid across Australia. This upscaling approach may be more applicable for rainfall indices with low natural variability (such as WD) but not for the others (such as R99p, SDII and R10mm). The CPP-HY approach is not yet in operational use but offers an attractive means for future operational intra-seasonal extreme rainfall forecasts. We have already trialled the CPP-HY in a forecast mode with the help of software engineers and solution architects. We are further testing out the use of the CPP based postprocessed forecast as drivers to some subsequential studies such as crop yield prediction.

6. Summary

The copula-based postprocessing (CPP) method developed by Li et al. [2020] is modified to adapt to forecast extreme rainfall. In this modification, the daily rainfall distribution is represented by a hybrid probability distribution, which keeps the empirical distribution for low to medium rainfall but adopts the extended Burr XII distribution for heavy to extreme rainfall. Unlike the existing CPP which models the rainfall distribution nonparametrically, the modified CPP with the hybrid distribution may forecast extreme rainfall that has been never happened in the range of historical rainfall observations. A case study to postprocess the 23-year ACCESS-S1 hindcasts at locations that represent 17 rainfall stations in Queensland is carried out to compare the forecast performance of the existing CPP, modified CPP and two quantile mapping based methods. The performance of seasonal rainfall forecasts in terms of mean forecast and extreme forecast is evaluated from the skill assessment of forecasting 12 rainfall indices (including one mean rainfall index and 11 extreme rainfall indices) in the framework of ensemble forecast verification. The main conclusions for this study include:

1) The raw ACCESS-S1 forecast has no skill to forecast total rainfall and extreme rainfall at a seasonal scale and especially performs badly for light rainfall related indices due to “drizzle-day syndrome”;
2) All postprocessing methods lead to overall skill improvement relative to the raw forecast to some extent and the modified CPP leads to

greater skill improvement than the other postprocessing methods, including the existing CPP, in most cases;
3) The skill improvement from the modified CPP is often greater for light to medium rainfall related indices and smaller for heavy and very heavy rainfall related indices;
4) The skill improvement has been found in all selected climate regions and initialisation dates on the 1st day of each month, though more substantial improvement is observed in the tropical region where the raw forecast is particularly unskilful;
5) The forecast skill decreases with growing lead times and the modified CPP leads to neutral forecasts for most rainfall indices beyond 0-month lead time;
6) The extended Burr XII distribution and the least square parameter estimation lead to better model fitting to far tail rainfall distributions than the generalized Pareto distribution and maximum likelihood estimation.

Author statement

Ming Li: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Visualization, Huidong Jin: Project administration, Investigation, Resources, Data Curation, Funding acquisition, Writing - Review & Editing, Quanxi Shao: Investigation, Supervision, Writing - Review & Editing.

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.

Acknowledgements

This work was partially funded by the CSIRO Digiscape Future Science Platform. We are grateful to two anonymous reviewers and the editor for providing constructive comments and suggestions. We thank Andrea Powell for insightful suggestions and comments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.wace.2021.100384.

Appendix

A.1. Quantile mapping

Quantile mapping (QM) is one of the most used statistical postprocessing methods and matches the quantiles of raw GCM forecasts with the quantile of historical observations. Denote the cumulative distribution functions (CDF) of raw forecasts $X_{raw}$and historical observations X_{obs} by $F_{raw}(\cdot)$ and $F_{obs}(\cdot)$, respectively. The QM forecast X_{qm} is derived from

$$X_{qm} = F_{obs}^{-1}\{F_{raw}(X_{raw})\} \tag{A1}$$

where $F_{obs}^{-1}(\cdot)$ is the inverse function of $F_{obs}(\cdot)$(also known as the quantile function of historical observations). It is easy to show that the cumulative distribution of X_{qm} is equal to$F_{obs}(\cdot)$. This implies that the QM forecasts and historical observations share the same distribution-related characteristics, such as quantiles (that is why this method is known as quantile mapping) and mean (that is why QM is an effective bias correction). QM can be done either parametrically or nonparametrically depending on how the cumulative distributions $F_{raw}(\cdot)$ and $F_{obs}(\cdot)$ are estimated. Parametric QM fits the cumulative distributions with parametric distributions. Nonparametric QM uses nonparametric distribution estimations, for example, empirical distributions.

A.2. A review of the existing copula-based postprocessing (CPP) method

CPP, designed as a fully calibrated method, seeks to model the dependence structure between observed and GCM raw forecast variables by copulas and derive the conditional forecast distribution accordingly. Sklar's theorem [Bardossy and Plate, 1992] showed that the joint distribution between X_{obs} and X_{raw}, denoted by $F(X_{obs}, X_{raw})$, can be formulated as a function of $F_{obs}(X_{obs})$ and $F_{raw}(X_{raw})$:

$$F(X_{obs}, X_{raw}) = C\{F_{obs}(X_{obs}), F_{raw}(X_{raw})\}, \tag{A2}$$

where C is a copula function. Copula models separate the modelling of marginal distributions from the modelling of dependence structure. Based on the representation of the joint distribution, copula models can further express the distribution of X_{obs} conditioning on X_{raw}, denoted by $h(X_{obs}; X_{raw}) = F(X_{obs}|X_{raw})$, in the following equation:

$$h(X_{obs}; X_{raw}) = \frac{\partial C\{F_{obs}(X_{obs}), F_{raw}(X_{raw})\}}{\partial F_{raw}(X_{raw})}. \tag{A3}$$

The conditional forecast distribution given by (A3) forms the foundation of CPP and requires the estimation of three "parameters": F_{obs}, F_{raw} and C. In the current implementation of CPP, the marginal distributions F_{obs} and F_{raw} are estimated nonparametrically by empirical distributions. This practical estimation is originally designed to conveniently handle numerous zero values in observed and forecast rainfall but restricts the use of CPP to non-extreme rainfall. In the next section, we will explain how the estimation of F_{obs} and F_{raw} by empirical distributions becomes a problem for forecasting extreme rainfall in detail and will provide a revised estimation to overcome this problem. The copula function C is estimated by the principle of maximum likelihood. Because the presence of numeric zero values, both X_{obs} and X_{raw} actually are a mixture of discrete and continuous random variables. Li et al. [2020] treated rainfall as a left-censored random variable and derived a likelihood estimation without explicitly specifying a discrete-continuous distribution. Under the assumption of data censoring, the likelihood can be broken down into four cases to account for whether

X_{obs} or X_{raw} is censored. For a given copula family, copula parameters are estimated by maximising the log likelihood function. The estimation is carried out for a range of copula families and the best-fit copula family is selected by the Akaike information criterion (AIC).

In the context of seasonal rainfall forecasting, these three "parameters" are estimated for each lead time and each initialisation date in order to account for the effect of lead time and seasonal dependence. The hindcasts (i.e. simulation outputs for historical events), used for model training, may have a limited number of data because CPP builds separate models for each individual lead time and calendar month. In this case, the pooled data from a sliding window on the target date are recommended to augment the size of training data. In this study, an 11-day sliding window is applied to include the data from the five days before and after the target date (except for the first and last five days, where a ramped-up sliding window is used) by following the setting of operationally postprocessing the ACCESS-S1 outputs by the BoM (See Section 3 for more details).

After the CPP model parameters are estimated, the ensemble members of CPP rainfall forecasts are firstly simulated from the conditional forecast distribution given by (A3). CPP ensembles are generated differently for zero and non-zero raw rainfall forecasts. More specifically, when $X_{raw} > 0$, an ensemble member of CPP forecasts, denoted by X_{cpp}, is generated by the following expression:

$$X_{cpp} = F_{obs}^{-1}[h_{-1}\{u; F_{raw}(X_{raw})\}], \tag{A4}$$

where u is a standard uniform random number on $[0, 1]$. When $X_{raw} = 0$, X_{cpp} is generated in a slightly different way to address X_{raw} is left censored:

$$X_{cpp} = F_{obs}^{-1}\{h^{-1}(u; u_1)\}, \tag{A5}$$

where u_1 is a uniform random number on $[0, F_{raw}(0)]$. CPP ensembles are simulated independently for each lead time and each station and this causes that the spatial and temporal correlation described by any particular ensemble member is close to zero. To recover a more realistic correlation that promises better aggregated rainfall forecasts over time and space, Schaake shuffle (SS) [Clark et al., 2004] is applied after the "unordered" CPP ensemble forecasts are generated.

A.3. A hybrid distribution based on the EBXII distribution

The cumulative distribution function and the density function of the EBXII distribution, denoted by F_{EBXII} and f_{EBXII} respectively, can be written as

$$F_{EBXII}(x; \zeta, c, \lambda) = \begin{cases} 1 - \left\{1 + \zeta\left(\lambda^{-1}x\right)^c\right\}^{-\zeta^{-1}}, \zeta \neq 0 \\ 1 - \exp\left\{-\left(\lambda^{-1}x\right)^c\right\}, \zeta = 0 \end{cases}, \tag{A6}$$

$$f_{EBXII}(x; \zeta, c, \lambda) = \begin{cases} c\lambda^{-1}\left(\lambda^{-1}x\right)^{c-1}\left\{1 + \zeta\left(\lambda^{-1}x\right)^{c-1}\right\}^{-\zeta^{-1}-1}, \zeta \neq 0 \\ c\lambda^{-1}\left(\lambda^{-1}x\right)^{c-1}\exp\left\{-\left(\lambda^{-1}x\right)^c\right\}, \zeta = 0 \end{cases}, \tag{A7}$$

where ζ and c are two shape parameters and λ is a scale parameter. Shao et al. [2004] showed that the range of x is $x > 0$ when $\zeta \geq 0$ and is $0 < x < \lambda(-\zeta)^{-c^{-1}}$ when $\zeta < 0$. When $c = 1$, the EBXII distribution decays to a GPD. Because the extreme value theory proves that excesses over high thresholds asymptotically follows a GPD, the EBXII distribution, as a generalized version of the GPD, can fit the extremes over thresholds better for finite samples.

The least square estimation is used to estimate the parameters of the EBXII distribution for each month. Denote n sorted daily rainfall greater than a prescribed threshold x_{ts} by $x_{ts} < x_{(k)} \leq x_{(k+1)} \leq \cdots \leq x_{(n)}$ and the least square estimator minimises the following objective function

$$LS(\zeta, c, \lambda | x_{(k)}, x_{(k+1)}, \cdots, x_{(n)}) = \sum_{i=k}^{n}\left[x_{(i)} - F_{EBXII}^{-1}\{F_o(x_{(i)}; \zeta, c, \lambda)\}\right]^2. \tag{A8}$$

A hybrid distribution F_H is a mixture of an empirical distribution F_0 and an EBXII distribution F_{EBXII} with a prescribed threshold x_{ts}:

$$F_H(x) = F_0(x)I(x \leq x_{ts}) + F_{EBXII}(x - x_{ts})I(x > x_{ts}). \tag{A9}$$

The quantile function of the hybrid distribution is the inversion of F_H, which can be expressed as

$$F_H^{-1}(y) = F_0^{-1}(y)I(y \leq y_{ts}) + \left\{F_{EBXII}^{-1}(y) + x_{ts}\right\}I(y > y_{ts}) \tag{A10}$$

where $y_{ts} = F_0(x_{ts})$ is the threshold corresponding to x_{ts} in the probability space.

References

Abouelmagd, T.H.M., Hamed, M.S., Afify, A.Z., 2017. The extended Burr XII distribution with variable shapes for the hazard rate. Pak. J. Statistics Oper. Res. 13 (3), 687–698. https://doi.org/10.18187/pjsor.v13i3.2083.

Allan, R.P., Soden, B.J., 2008. Atmospheric warming and the amplification of precipitation extremes. Science 321 (5895), 1481–1484. https://doi.org/10.1126/science.1160787.

Australian Business Roundtable for Disaster Resilience and Safer Communities, 2016. The economic cost of the socialimpact of natural disasters. Available online at: http://australianbusinessroundtable.com.au/assets/documents/Report%20-%20Social%20costs/Report%20-%20The%20economic%20cost%20of%20the%20social%20impact%20of%20natural%20disasters.pdf.

Bardossy, A., Plate, E.J., 1992. Space-time model for daily rainfall using atmospheric circulation patterns. Water Resour. Res. 28 (5), 1247–1259. https://doi.org/10.1029/91WR02589.

Beirlant, J., de Wet, T., Goegebeur, Y., 2006. A goodness-of-fit statistic for Pareto-type behaviour. J. Comput. Appl. Math. 186 (1), 99–116. https://doi.org/10.1016/j.cam.2005.01.036.

Bureau of Meteorology, 2019. Special Climate Statement 69—an extended period of heavy rainfall and flooding in tropical Queensland. Available online at: http://www.bom.gov.au/climate/current/statements/scs69.pdf.

Cannon, A.J., Sobie, S.R., Murdock, T.Q., 2015. Bias correction of GCM precipitation by quantile mapping: how well do methods preserve changes in quantiles and extremes? J. Clim. 28 (17), 6938–6959. https://doi.org/10.1175/JCLI-D-14-00754.1.

Cecinati, F., Wani, O., Rico-Ramirez, M.A., 2017. Comparing approaches to deal with non-gaussianity of rainfall data in kriging-based radar-gauge rainfall merging. Water Resour. Res. 53 (11), 8999–9018. https://doi.org/10.1002/2016WR020330.

Choulakian, V., Stephens, M.A., 2001. Goodness-of-fit tests for the generalized Pareto distribution. Technometrics 43 (4), 478–484. https://doi.org/10.1198/00401700152672573.

Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., Wilby, R., 2004. The Schaake shuffle: a method for reconstructing space-time variability in forecasted

precipitation and temperature fields. J. Hydrometeorol. 5 (1), 243–262. https://doi.org/10.1175/1525-7541(2004)005<0243:TSSAMF>2.0.CO;2.
Coles, S., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer, London ; New York, p. 208.
Cowan, T., Wheeler, M.C., Alves, O., Narsey, S., de Burgh-Day, C., Griffiths, M., Jarvis, C., Cobon, D.H., Hawcroft, M.K., 2019. Forecasting the extreme rainfall, low temperatures, and strong winds associated with the northern Queensland floods of February 2019. Weather and Climate Extremes 26. https://doi.org/10.1016/j.wace.2019.100232.
Cressie, N., 2015. Statistics for Spatial Data. John Wiley & Sons.
Curtis, S., Fair, A., Wistow, J., Val, D.V., Oven, K., 2017. Impact of extreme weather events and climate change for health and social care systems. Environ Health-Glob 16, 23–32. https://doi.org/10.1186/s12940-017-0324-3.
Dai, A., 2006. Precipitation characteristics in eighteen coupled climate models. J. Clim. 19 (18), 4605–4630. https://doi.org/10.1175/JCLI3884.1.
Donat, M.G., Lowry, A.L., Alexander, L.V., O'Gorman, P.A., Maher, N., 2016. More extreme precipitation in the world's dry and wet regions. Nat. Clim. Change 6 (5), 508–513. https://doi.org/10.1038/nclimate2941.
Engeland, K., Renard, B., Steinsland, I., Kolberg, S., 2010. Evaluation of statistical models for forecast errors from the HBV model. J. Hydrol. 384 (1–2), 142–155. https://doi.org/10.1016/j.jhydrol.2010.01.018.
Feng, P.Y., Wang, B., Liu, D.L., Xing, H.T., Ji, F., Macadam, I., Ruan, H.Y., Yu, Q., 2018. Impacts of rainfall extremes on wheat yield in semi-arid cropping systems in eastern Australia. Climatic Change 147 (3–4), 555–569. https://doi.org/10.1007/s10584-018-2170-x.
Ferro, C.A.T., Richardson, D.S., Weigel, A.P., 2008. On the effect of ensemble size on the discrete and continuous ranked probability scores. Meteorol. Appl. 15 (1), 19–24. https://doi.org/10.1002/met.45.
Friederichs, P., 2010. Statistical downscaling of extreme precipitation events using extreme value theory. Extremes 13 (2), 109–132. https://doi.org/10.1007/s10687-010-0107-5.
Friederichs, P., Wahl, S., Buschow, S., 2018. Chapter 5 - postprocessing for extreme events. In: Vannitsem, S., Wilks, D.S., Messner, J.W. (Eds.), Statistical Postprocessing of Ensemble Forecasts. Elsevier, pp. 127–154. https://doi.org/10.1016/B978-0-12-812372-0.00005-4.
Furrer, E.M., Katz, R.W., 2008. Improving the simulation of extreme precipitation events by stochastic weather generators. Water Resour. Res. 44 (12) https://doi.org/10.1029/2008WR007316.
Gerstengarbe, F.W., Werner, P.C., 1989. A method for the statistical definition of extreme-value regions and their application to meteorological time-series. Z. Meteorol. 39 (4), 224–226.
Glahn, H.R., Lowry, D.A., 1972. The use of model output statistics (MOS) in objective weather forecasting. J. Appl. Meteorol. 11 (8), 1203–1211. https://doi.org/10.1175/1520-0450(1972)011<1203:tuomos>2.0.co;2.
Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. J. Am. Stat. Assoc. 102 (477), 359–378. https://doi.org/10.1198/016214506000001437.
Gneiting, T., Balabdaoui, F., Raftery, A.E., 2007. Probabilistic forecasts, calibration and sharpness. J. Roy. Stat. Soc. B 69, 243–268. https://doi.org/10.1111/j.1467-9868.2007.00587.x.
Gonzalez, F.R., Raval, S., Taplin, R., Timms, W., Hitch, M., 2019. Evaluation of impact of potential extreme rainfall events on mining in Peru. Nat. Resour. Res. 28 (2), 393–408. https://doi.org/10.1007/s11053-018-9396-1.
Grimit, E.P., Gneiting, T., Berrocal, V.J., Johnson, N.A., 2006. The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification. Q. J. Roy. Meteorol. Soc. 132 (621), 2925–2942. https://doi.org/10.1256/qj.05.235.
Hawkins, E., Osborne, T.M., Ho, C.K., Challinor, A.J., 2013. Calibration and bias correction of climate projections for crop modelling: an idealised case study over Europe. Agric. For. Meteorol. 170, 19–31. https://doi.org/10.1016/j.agrformet.2012.04.007.
Hersbach, H., 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather Forecast. 15 (5), 559–570. https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.
Hudson, D., et al., 2017. ACCESS-S1: the new Bureau of Meteorology multi-week to seasonal prediction system. J So Hemisph Earth 67 (3), 132–159. https://doi.org/10.22499/3.6703.001.
Hundecha, Y., Pahlow, M., Schumann, A., 2009. Modeling of daily precipitation at multiple locations using a mixture of distributions to characterize the extremes. Water Resour. Res. 45 https://doi.org/10.1029/2008WR007453.
Ines, A.V.M., Hansen, J.W., 2006. Bias correction of daily GCM rainfall for crop simulation studies. Agric. For. Meteorol. 138 (1–4), 44–53. https://doi.org/10.1016/j.agrformet.2006.03.009.
Jia, J.-m., Yan, Z.-z., Peng, X.-y., 2018. An extended burr-XII distribution with application to lifetime data. J. Test. Eval. 48 (5), 3360–3375. https://doi.org/10.1520/JTE20180144.
Kamiguchi, K., Kitoh, A., Uchiyama, T., Mizuta, R., Noda, A., 2006. Changes in precipitation-based extremes indices due to global warming projected by a global 20-km-mesh atmospheric model. Solanus 2, 64–67. https://doi.org/10.1016/j.wace.2015.09.001.
King, A.D., Hudson, D., Lim, E.P., Marshall, A.G., Hendon, H.H., Lane, T.P., Alves, O., 2020. Subseasonal to seasonal prediction of rainfall extremes in Australia. Q. J. Roy. Meteorol. Soc. 146 (730), 2228–2249. https://doi.org/10.1002/qj.3789.
Koutsoyiannis, D., 2004a. Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation. Hydrol. Sci. J. 49 (4), 575–590. https://doi.org/10.1623/hysj.49.4.575.54430.
Koutsoyiannis, D., 2004b. Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records. Hydrol. Sci. J. 49 (4), 591–610. https://doi.org/10.1623/hysj.49.4.591.54424.
Li, C., Singh, V.P., Mishra, A.K., 2012. Simulation of the entire range of daily precipitation using a hybrid probability distribution. Water Resour. Res. 48 https://doi.org/10.1029/2011WR011446.
Li, M., Jin, H.D., 2020. Development of a postprocessing system of daily rainfall forecasts for seasonal crop prediction in Australia. Theor. Appl. Climatol. 141 (3–4), 1331–1349. https://doi.org/10.1007/s00704-020-03268-3.
Li, M., Jin, H., Brown, J.N., 2020. Making the output of seasonal climate models more palatable to agriculture: a copula-based post-processing method. J Appl Meteorol Clim 497–515. https://doi.org/10.1175/jamc-d-19-0093.1.
Lim, E.P., Hendon, H.H., Anderson, D.L.T., Charles, A., Alves, O., 2011. Dynamical, statistical-dynamical, and multimodel ensemble forecasts of Australian spring season rainfall. Mon. Weather Rev. 139 (3), 958–975. https://doi.org/10.1175/2010MWR3399.1.
Lim, Y., Son, S.W., Marshall, A.G., Hendon, H.H., Seo, K.H., 2019. Influence of the QBO on MJO prediction skill in the subseasonal-to-seasonal prediction models. Clim. Dynam. 53 (3–4), 1681–1695. https://doi.org/10.1007/s00382-019-04719-y.
Liu, Y., Li, D.R., Li, Y.Q., Zhang, Z.H., 2018. Estimation of extreme value vehicle load based on the extended Burr XII distribution. Ksce J Civ Eng 22 (9), 3401–3408. https://doi.org/10.1007/s12205-017-1727-y.
Maraun, D., 2013. Bias correction, quantile mapping, and downscaling: revisiting the inflation issue. J. Clim. 26 (6), 2137–2143. https://doi.org/10.1175/JCLI-D-12-00821.1.
Maraun, D., Widmann, M., 2018. Statistical Downscaling and Bias Correction for Climate Research. Cambridge University Press.
Marshall, A.G., Hendon, H.H., 2015. Subseasonal prediction of Australian summer monsoon anomalies. Geophys. Res. Lett. 42 (24), 10913–10919. https://doi.org/10.1002/2015GL067086.
Merryfield, W.J., et al., 2020. Current and emerging developments in subseasonal to decadal prediction. Bull. Am. Meteorol. Soc. 101 (6), E869–E896. https://doi.org/10.1175/BAMS-D-19-0037.1.
Mylne, K.R., Woolcock, C., Denholm-Price, J.C.W., Darvell, R.J., 2002. Operational calibrated probability forecasts from the ECMWF ensemble prediction system: implementation and verification. In: Symposium on Observations, Data Assimilation, and Probabilistic Prediction, pp. 113–118.
Naveau, P., Huser, R., Ribereau, P., Hannart, A., 2016. Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. Water Resour. Res. 52 (4), 2753–2769. https://doi.org/10.1002/2015WR018552.
Olsson, T., Jakkila, J., Veijalainen, N., Backman, L., Kaurola, J., Vehvilainen, B., 2015. Impacts of climate change on temperature, precipitation and hydrology in Finland - studies using bias corrected Regional Climate Model data. Hydrol. Earth Syst. Sci. 19 (7), 3217–3238. https://doi.org/10.5194/hess-19-3217-2015.
Papalexiou, S.M., Montanari, A., 2019. Global and regional increase of precipitation extremes under global warming. Water Resour. Res. 55 (6), 4901–4914. https://doi.org/10.1029/2018WR024067.
Queensland Government, 2012. Queensland rainfall—past, present and future. Available online at: https://data.longpaddock.qld.gov.au/static/about/publications/pdf/walker-report-summary-brochure.pdf.
Schepen, A., Wang, Q.J., 2014. Ensemble forecasts of monthly catchment rainfall out to long lead times by post-processing coupled general circulation model output. J. Hydrol. 519, 2920–2931. https://doi.org/10.1016/j.jhydrol.2014.03.017.
Schepen, A., Wang, Q.J., Robertson, D.E., 2014. Seasonal forecasts of Australian rainfall through calibration and bridging of coupled GCM outputs. Mon. Weather Rev. 142 (5), 1758–1770. https://doi.org/10.1175/MWR-D-13-00248.1.
Shao, Q.X., Wong, H., Xia, J., Ip, W.C., 2004. Models for extremes using the extended three-parameter Burr XII system with application to flood frequency analysis. Hydrol. Sci. J. 49 (4), 685–702. https://doi.org/10.1623/hysj.49.4.685.54425.
Sperber, K.R., Palmer, T.N., 1996. Interannual tropical rainfall variability in general circulation model simulations associated with the atmospheric model intercomparison project. J. Clim. 9 (11), 2727–2750. https://doi.org/10.1175/1520-0442(1996)009<2727:ITRVIG>2.0.CO;2.
Stephens, G.L., L'Ecuyer, T., Forbes, R., Gettelman, A., Golaz, J.C., Bodas-Salcedo, A., Suzuki, K., Gabriel, P., Haynes, J., 2010. Dreary state of precipitation in global models. J. Geophys. Res. Atmos. 115 https://doi.org/10.1029/2010JD014532.
Stern, R.D., Coe, R., 1984. A model-fitting analysis of daily rainfall data. J R Stat Soc a Stat 147, 1–34. https://doi.org/10.2307/2981736.
Stott, P., 2016. How climate change affects extreme weather events. Science 352 (6293), 1517–1518. https://doi.org/10.1126/science.aaf7271.
Todorovic, P., Woolhiser, D.A., 1975. A stochastic model of n-day precipitation. J. Appl. Meteorol. 14 (1), 17–24. https://doi.org/10.1175/1520-0450(1975)014<0017:ASMODP>2.0.CO;2.
Veenhuis, B.A., 2013. Spread calibration of ensemble MOS forecasts. Mon. Weather Rev. 141 (7), 2467–2482. https://doi.org/10.1175/MWR-D-12-00191.1.
Vrac, M., Naveau, P., 2007. Stochastic downscaling of precipitation: from dry events to heavy rainfalls. Water Resour. Res. 43 (7) https://doi.org/10.1029/2006WR005308.
Wang, B., Kang, I.S., Lee, J.Y., 2004. Ensemble simulations of Asian-Australian monsoon variability by 11 AGCMs. J. Clim. 17 (4), 803–818. https://doi.org/10.1175/1520-0442(2004)017<0803:ESOAMV>2.0.CO;2.
Wilks, D.S., 1998. Multisite generalization of a daily stochastic precipitation generation model. J. Hydrol. 210 (1–4), 178–191. https://doi.org/10.1016/S0022-1694(98)00186-3.
Wilks, D.S., 1999. Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. Agric. For. Meteorol. 93 (3), 153–169. https://doi.org/10.1016/S0168-1923(98)00125-7.

Wilks, D.S., 2015. Multivariate ensemble Model Output Statistics using empirical copulas. Q. J. Roy. Meteorol. Soc. 141 (688), 945–952. https://doi.org/10.1002/qj.2414.
Wilks, D.S., 2020. Statistical Methods in the Atmospheric Sciences. Elsevier.
Woolhiser, D.A., Roldan, J., 1982. Stochastic daily precipitation models .2. A comparison of distributions of amounts. Water Resour. Res. 18 (5), 1461–1468. https://doi.org/10.1029/WR018i005p01461.
Zhang, X.B., Alexander, L., Hegerl, G.C., Jones, P., Tank, A.K., Peterson, T.C., Trewin, B., Zwiers, F.W., 2011. Indices for monitoring changes in extremes based on daily temperature and precipitation data. Wires Clim Change 2 (6), 851–870. https://doi.org/10.1002/wcc.147.