Making the Output of Seasonal Climate Models More Palatable to Agriculture: A Copula-Based Postprocessing Method^①

MING LI

CSIRO Data61, Kensington, Western Australia, Australia

HUIDONG JIN

CSIRO Data61, Canberra, Australian Capital Territory, Australia

JACLYN N. BROWN

CSIRO Agriculture and Food, Sandy Bay, Tasmania, Australia

(Manuscript received 22 April 2019, in final form 5 December 2019)

ABSTRACT

Seasonal climate forecasts from raw climate models at coarse grids are often biased and statistically unreliable for credible crop prediction at the farm scale. We develop a copula-based postprocessing (CPP) method to overcome this mismatch problem. The CPP forecasts are ensemble based and are generated from the predictive distribution conditioned on raw climate forecasts. CPP performs univariate postprocessing procedures at each station, lead time, and variable separately and then applies the Schaake shuffle to reorder ensemble sequence for a more realistic spatial, temporal, and cross-variable dependence structure. The use of copulas makes CPP free of strong distributional assumptions and flexible enough to describe complex dependence structures. In a case study, we apply CPP to postprocess rainfall, minimum temperature, maximum temperature, and radiation forecasts at a monthly level from the Australian Community Climate and Earth-System Simulator Seasonal model (ACCESS-S) to three representative stations in Australia. We evaluate forecast skill at lead times of 0–5 months on a cross-validation theme in the context of both univariate and multivariate forecast verification. When compared with forecasts that use climatological values as the predictor, the CPP forecast has positive skills, although the skills diminish with increasing lead times and finally become comparable at long lead times. When compared with the bias-corrected forecasts and the quantile-mapped forecasts, the CPP forecast is the overall best, with the smallest bias and greatest univariate forecast skill. As a result of the skill gain from univariate forecasts and the effect of the Schaake shuffle, CPP leads to the most skillful multivariate forecast as well. Further results investigate whether using ensemble mean or additional predictors can enhance forecast skill for CPP.

1. Introduction

Year-to-year variations in agricultural productivity are largely underpinned by variations in the seasonal climate. Knowledge of the likely season ahead allows stakeholders along the agricultural value chain—farmers, agronomists, suppliers, distributors, and insurers—to better manage risk in bad years and enhance production

in good years (Hammer et al. 2000; Ines and Hansen 2006; McIntosh et al. 2005, 2007). Seasonal climate forecasts (SCFs) linked with crop simulation tools [such as potentially the national-scale C-Crop model (Donohue et al. 2018) and Agricultural Production System Simulator (APSIM) model (Keating et al. 2003)] inform the optimum combination of crop genetics and agronomic managements, and the quantification of associated possible profits and risks for famers or farming practice (Brown et al. 2018; McCrear et al. 2005; Meinke and Stone 2005; Rodriguez et al. 2018). For Australia alone, the Centre for International Economics (2014) found that the potential value of skillful SCFs for the agricultural sector is significant,

^① Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JAMC-D-19-0093.s1>.

Corresponding author: Ming Li, ming.li@csiro.au

about AUD \$1.5 billion per year at 2012 prices, or 7.31% of industry value added.

Nowadays climate forecasts at seasonal time scales have been increasingly available from coupled ocean–atmosphere general circulation models (GCMs) (Graham et al. 2005; Hudson et al. 2013; Paolino et al. 2012; Saha et al. 2006). GCMs are designed primarily for understanding climate change and for predicting climate variations. Growing understanding of ocean–atmosphere interactions, advanced modeling of global climate systems, better observational data, and improved computer technology have made SCFs from GCMs more accurate. To represent the chaotic nature of the climate system, GCMs generate a set of probable outcomes as ensemble members for SCFs. Doblas-Reyes et al. (2013) provided a state-of-the-art review of global seasonal climate predictability and argued that SCFs obtained from GCMs are potentially skillful at large temporal and spatial scales for many parts of the world.

Nevertheless, the coarse grid structures, imperfect physical parameterization, and sensitivity to perturbed initial conditions make raw GCM forecasts (typically available from 100 to 200 km) inadequate for credible crop simulations at the farm scale. The local climate features simulated from GCMs are potentially biased and/or statistically unreliable (i.e., the spread of ensemble forecast is too wide or narrow to accurately characterize the underlying uncertainty) (Graham et al. 2005). As a result, raw GCM forecasts are not good enough directly for crop simulation studies (Challinor et al. 2005; Mavromatis and Jones 1999). A standard practice to solve this “connectivity problem” (Stone and Meinke 2005) is statistical postprocessing.

Two most commonly used postprocessing methods are bias correction and quantile mapping. Bias correction adjusts the GCM outputs by using the mean bias between local observations and the GCM forecasts in a reference period (Hawkins et al. 2013; Huntingford et al. 2005; Ines and Hansen 2006). Ho et al. (2012) extended the simple bias correction to a more general case where both the mean bias and the temporal variability of the GCM forecasts are corrected in according to the corresponding observations. Quantile mapping uses a transfer function to align the cumulative distribution of GCM forecasts with that of the corresponding observations (Michelangeli et al. 2009; Wood et al. 2002, 2004). After quantile mapping, climate simulations from GCMs are statistically closer to observations in the reference period. In the context of univariate case, it can be shown that quantile mapping is the optimal transport map [see Santambrogio (2015, theorem 2.5) or Farchi et al. (2016, their appendix A) for a detailed proof and Robin et al. (2019) for an explanation on bias correction]. Currently,

multivariate quantile-mapping-based methods has been developed in literature, including a multivariate bias correction algorithm (MBCn; Cannon 2018), rank resampling for distributions and dependences (R^2D^2 ; Vrac 2018), and dynamical optimal transport correction (dOTC; Robin et al. 2019). Both bias correction and quantile mapping have been extensively used in long-range climate change projection (Grillakis et al. 2013; Li et al. 2010; Olsson et al. 2015). Some applications have used both bias correction and quantile mapping (Kokic et al. 2013).

Postprocessing climate model outputs for SCFs is not as simple as systematic bias removal. Both bias correction and quantile mapping are capable of removing systematic bias but not ensuring forecasts reliable in ensemble spread (Zhao et al. 2017). To support the use for crop simulation and other impact model applications, ensemble SCFs are actually desired to be skillful to the extent of outperforming the “climatology” forecast (Hersbach 2000; Murphy 1993). The forecast skill includes forecast accuracy, reliability, and sharpness (Gneiting et al. 2007) and retrospective forecasts need to be carefully verified against observed data in a reference period under cross validation. The task of postprocessing methods beyond bias correction is to establish the dependence structure between local observed data, raw GCM forecasts, and other possible predictors. This can be done by a conditional approach to establish a joint probability distribution or equivalently a conditional forecast distribution in a probabilistic framework (Glahn and Lowry 1972; Schepen and Wang 2014). Bias correction and quantile mapping are typically “deterministic” mapping (Maraun 2013) and the conditional approach is “probabilistic” mapping where the uncertainty of the raw forecasts is taken into account.

Because of the complex nature of climate variables, it is challenging to model such a joint probability distribution by a simple parametric multivariate distribution. A pragmatic approach is to apply a data transformation (e.g., the Box–Cox transformation and log transformation) on each marginal variable and describe the forecast distribution at the transformed space by a multivariate Gaussian distribution (Wang and Robertson 2011). The goal of data transformation is to transform data into a Gaussian distribution. Nevertheless, this goal cannot always be achieved, even approximately. For example, a multimodal distribution (e.g., probability density function with multiple peaks) cannot be transformed into any single-modal distributions (including Gaussian distribution) by any Box–Cox transformation. Alternatively, a multivariate probability distribution can be modeled directly by a copula without data transformation (Nelsen 2006). Copulas

conveniently separate the dependence structure of random variables from their marginal distributions. Since introduced in climate science by Schölzel and Friederichs (2008), copulas have been applied to solve many problems in climate study (Madadgar and Moradkhani 2013; Piani and Haerter 2012; Wilks 2015).

In this paper, we develop a new statistical postprocessing method based on copulas to improve SCFs at the farm scale for use in yield prediction. We call this method the copula-based postprocess method (CPP). The method uses copulas to model the joint probability distribution of GCM forecasts and the corresponding local observations and generates the CPP forecast from the conditional forecast distribution accordingly. The use of copulas makes CPP free of strong distributional assumptions (such as multivariate Gaussian after data transformation) and flexible to describe complex dependence structures. CPP starts from generating univariate CPP forecasts at each station, lead time, and variable separately and then apply the Schaake shuffle (Clark et al. 2004) to reorder ensembles for a more realistic spatial, temporal and cross-variable dependence structure. By using pair-copula constructions (PCC) (Aas et al. 2009; Bedford and Cooke 2001, 2002; Bevacqua et al. 2017), CPP can be extended to include the GCM forecast of large-scale climate indices as additional predictors in postprocessing. The idea has been called “bridging” in some previous studies (Schepen and Wang 2014; Schepen et al. 2014, 2016).

Climate forecasts at a daily scale or a seasonal scale (e.g., 3-month totals) are the most sought-after for crop model simulations or simple correlation analysis, though monthly climate forecasts also have been directly used for crop yield prediction (Peng et al. 2018). Following a brief review from Hawkins et al. (2013), we find that quality forecasts of monthly average are needed for some existing methods of transforming climate model outputs for use with crop models, including weather generators and the delta method. For example, some previous research has been done to use weather generators to disaggregate monthly rainfall forecasts to daily for crop model simulation (Hansen and Ines 2005; Ines et al. 2011; Osborne et al. 2013). In a case study at three representative stations in Australia, we apply CPP to postprocess the seasonal forecasts of four climate variables (including rainfall, minimum temperature, maximum temperature, and radiation) at a monthly resolution from an Australian state-of-the-art seasonal forecast GCM known as the Australian Community Climate and Earth-System Simulator Seasonal model (ACCESS-S). The forecast skill extending out to 0–5-month lead times is verified in comparison with bias correction and quantile mapping. Some further results show whether the use of

ensemble mean or additional predictors can bring skill improvement for CPP.

We organize this paper as follows. The CPP method is introduced in section 2. We describe the three stations and ACCESS-S forecasts used to test the method in section 3. We present the skill assessment of CPP forecasts in comparison with two other postprocessing methods in section 4. Discussion and conclusions are provided in sections 5 and 6, respectively.

2. Method

a. Bias correction and quantile mapping

For a univariate climate variable X , we obtain a raw forecast of this variable X_f from a GCM. As the GCM is run at a large spatial scale, the accuracy of X_f is generally not satisfactory at a local scale. A postprocessing procedure is often applied to raw forecasts before we use them as input variables of an impact model, such as crop models. Two commonly used postprocessing methods are bias correction (BC) and quantile mapping (QM).

The BC method simply corrects the forecast mean by extracting a constant from the raw forecast. In mathematical terms, the BC forecast X^{BC} can be expressed as

$$X^{\text{BC}} = X_f - \bar{X}_f + \bar{X}_o, \quad (1)$$

where $\bar{X}_f$ and $\bar{X}_o$ are the mean of raw forecast and observations over a reference period, respectively. The value of rainfall forecasts should be at least zero, but BC sometimes can lead to negative rainfall forecasts when a positive bias occurs. In this study, we force negative BC rainfall forecasts back to zero.

The QM method maps the raw forecast to the corresponding quantiles of historical observations. We denote the cumulative probability distributions (CDFs) of raw forecasts and observations by F_f and F_o , respectively. The QM forecast X^{QM} can be formulated as

$$X^{\text{QM}} = F_o^{-1}[F_f(X_f)], \quad (2)$$

where F_o^{-1} is the inverse function of F_o (also called the quantile function of X_o). In this study, we use the empirical distribution of raw forecasts and observations over a reference period as the estimates of F_f and F_o . Maraun (2018) called this approach the empirical quantile mapping. The main drawback of empirical quantile mapping is that an additional extrapolation is needed when the forecast value is beyond the range of raw forecasts in the historical period. Alternatively, the estimation of F_f and F_o can be done parametrically (i.e., fitted by a parametric distribution) (Piani et al. 2010; Wood et al. 2004). The parametric quantile mapping

requires a lot of efforts to choose an appropriate parametric distribution for each climate variable; otherwise, we may be exposed to the risk of distribution misspecification.

In this study, BC and QM are carried out at the level of individual ensemble members. As a result, the BC and QM forecasts keep the same ensemble size as the raw forecasts.

b. A review of copula and pair-copula construction

A copula is a statistical model to represent a multivariate uniform distribution (i.e., the marginal distribution of each variable is uniform). Sklar's theorem (Sklar 1959) stated that every multivariate CDF $F(y_1, \dots, y_n)$ can be represented by a copula C and its marginal CDFs $F(y_1), \dots, F(y_n)$:

$$F(y_1, \dots, y_n) = C[F_1(y_1), \dots, F_n(y_n)]. \quad (3)$$

Sklar's theorem provides a way to separate the modeling of the dependence structure of multivariate distributions from the modeling of the marginal distributions. For this reason, copulas can be used to construct a wide range of multivariate parametric distributions.

It is not always obvious to identify a copula (Frees and Valdez 1998). For the bivariate case, an extensive range of copula families has been well investigated in the literature—for example, Nelsen (2006, chapter 3). However, for the multivariate case, the choices of copula families are often limited to the multivariate Gaussian, Student's t , and exchangeable Archimedean copulas. These multivariate parametric copulas usually presume that the variables in all the pairs have homogeneous dependence structures (e.g., they are of the same type). In the case in which a complex dependence

structure presents, multivariate parametric copulas can be inadequate and not produce satisfying results.

PCCs provide a tool to flexibly model the dependence structure of multivariate distributions (Aas et al. 2009; Bevacqua et al. 2017). PCCs decompose the multivariate dependence structure into a cascade of bivariate copulas (conditionally and unconditionally on the original variables). In particular, PCCs mathematically decompose an n -dimensional copula density into a product of multiple bivariate copulas. Each pair copula is identified from a large number of bivariate parametric copula families, and such an identification is independent from the identification of other pair copulas.

There are many possible and equally valid PCCs for a particular multivariate distribution, and Bedford and Cooke (2001, 2002) introduced the regular vine, a graphical model to organize PCCs. We use one of the simplest regular vines, the canonical (C) vine, in this study for ease of computation. The decomposition of the n -dimensional density $f(y_1, \dots, y_n)$ on the basis of a C vine is generally expressed as

$$f(y_1, \dots, y_n) = \prod_{k=1}^n f(y_k) \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{j+i|1, \dots, j-1} \times [F(y_j|y_1, \dots, y_{j-1}), F(y_{j+i}|y_1, \dots, y_{j-1})], \quad (4)$$

where $c_{j+i|1, \dots, j-1}$ is the conditional copula density of y_j and y_{j+i} given values for $y_1, \dots, y_{j-1}$. The marginal conditional distributions $F(y_j|y_1, \dots, y_{j-1})$ and $F(y_{j+i}|y_1, \dots, y_{j-1})$ can be calculated from

$$F(y_j|y_1, \dots, y_{j-1}) = \frac{\partial C_{j-1,j|1, \dots, j-2}[F(y_j|y_1, \dots, y_{j-2}), F(y_{j-1}|y_1, \dots, y_{j-2})]}{\partial F(y_{j-1}|y_1, \dots, y_{j-2})} \quad \text{and} \quad (5)$$

$$F(y_{j+i}|y_1, \dots, y_{j-1}) = \frac{\partial C_{j-1,j+i|1, \dots, j-2}[F(y_{j+i}|y_1, \dots, y_{j-2}), F(y_{j-1}|y_1, \dots, y_{j-2})]}{\partial F(y_{j-1}|y_1, \dots, y_{j-2})}. \quad (6)$$

For example, the decomposition of the four-dimensional density on the basis of a C vine can be written as

$$f(y_1, y_2, y_3, y_4) = f(y_1)f(y_2)f(y_3)f(y_4)c_{12}[F(y_1), F(y_2)]c_{13}[F(y_1), F(y_3)]c_{14}[F(y_1), F(y_4)]c_{23|1} \times [F(y_2|y_1), F(y_3|y_1)]c_{24|1}[F(y_2|y_1), F(y_4|y_1)]c_{34|12}[F(y_3|y_1, y_2), F(y_4|y_1, y_2)]. \quad (7)$$

Figure 1 presents a graphical representation of the four-dimensional C vine in Eq. (7), which includes three trees

and six edges. Each edge corresponds to a pair-copula density and the edge label specifies the definition of pair

FIG. 1. Representation of the four-dimensional C vine in Eq. (7).

variables. For example, the edge labeled 34|12 represents the copula density of y_3 and y_4 conditioned on y_1 and y_2 . More detail about the C-vine representation can be found at Aas et al. (2009).

Each pair copula in the decomposition Eq. (4) involves at least one parameter, and we apply the algorithm based on maximum pseudolikelihood described by Aas et al. (2009) to estimate these parameters of C-vine density. This algorithm is applied to the normalized ranks of the data instead of the original data. By doing this, we can assume that the margin at each dimension is uniform on $[0, 1]$ and greatly simplify the evaluation of pseudolikelihood. This algorithm consists of only multiple bivariate copula parameter estimations and thus is very computationally efficient. The detail of this parameter estimation algorithm can be found at Aas et al. (2009).

c. Copula-based postprocessing method

In the framework of copulas, we postprocess a univariate climate variable X and formulate the joint distribution of raw forecasts and observations $F(X_f, X_o)$ by a bivariate copula as

$$F(X_f, X_o) = C[F_f(X_f), F_o(X_o)], \quad (8)$$

where C is a bivariate copula function and F_f and F_o are the marginal distributions of X_f and X_o . We estimate the marginal distributions F_f and F_o by the empirical distribution of X_f and X_o from the historical data. We use the R package Vinecopula (<https://CRAN.R-project.org/package=VineCopula>) to select an appropriate bivariate copula family by using the Akaike information criterion and estimate the corresponding parameters of C by the maximum likelihood estimation.

For a given C , F_f , and F_o , the conditional forecast distribution $F(X_o|X_f)$ can be calculated by

$$F(X_o|X_f) = \frac{\partial C[F_f(X_f), F_o(X_o)]}{\partial F_f(X_f)}. \quad (9)$$

We thereafter generate random samples from $F(X_o|X_f)$ given by Eq. (9) and use them to form the base of the CPP forecasts. We derive the conditional forecast distribution at the level of each individual ensemble member. In particular, we draw p random samples (hereinafter referred to as subensembles) from $F(X_o|X_f)$ for a given ensemble member of the raw forecast and combine all subensembles to create a superensemble forecast, which is referred to as the “unordered” CPP forecast. By this way, we augment the ensemble size by p times.

The p subensembles for a given ensemble member of the raw forecast are generated independently for different lead times, stations, and variables. As a result, spatial, temporal, and cross-variable correlations are not retained in the subensembles of the unordered CPP forecasts. Prior to forcing a crop model for yield prediction, we apply a pragmatic approach known as the Schaake shuffle (SS) (Clark et al. 2004) to reorder ensemble sequence for a more realistic representation of the spatial variability, temporal persistence, and variable association: 1) select p random months from similar months in the historical record, 2) generate a climatology forecast based on the historical observations on the selected months, and 3) reassign the historical observations in the climatology forecast with new values obtained from the unordered CPP subensembles by matching with the rank in p ensembles. Detailed tutorial-style examples are provided by Clark et al. (2004). We call the forecasts after the SS as the CPP forecasts.

In summary, the CPP forecasts are generated in three steps:

- 1) the estimation step, in which we estimate a copula and the marginal distribution of raw forecasts and observations,
- 2) the forecasting step, in which we generate unordered ensemble forecasts from the conditional forecast distribution, and

- 3) the reordering step, in which we apply the SS to unordered ensemble forecasts for more realistic temporal, spatial, and cross-variable dependence.

d. Two variants of the CPP forecast

Many researchers have shown that some large-scale climate indices, denoted by Y , have strong dependence with local climate variables (Crimp et al. 2019; Kokic et al. 2013). Therefore, in addition to the raw forecast X_f , the GCM forecasts of large-scale climate indices Y_f may provide useful information for postprocessing the raw forecast. In this case, we consider the conditional distribution $F(X_o|X_f, Y_f)$ as the forecast distribution and generate the CPP forecasts with external variables (referred to as the CPP-E forecasts). For the CPP-E forecasts, the C-vine copula is used to estimate the joint distribution $F(X_o, X_f, Y_f)$ and the corresponding conditional distribution $F(X_o|X_f, Y_f)$ by Eq. (4).

The CPP forecasts derive the forecast distribution based on the exact value of each ensemble member in the raw forecast. Alternatively, we can formulate the distribution of X_o conditioned on the ensemble mean of the raw forecast, denoted by X_f^M . Analogous to the CPP forecasts, we generate the random samples from $F(X_o|X_f^M)$ and apply the SS to reorder the ensembles. We call this variant the CPP forecast with ensemble mean (referred to as the CPP-M forecast).

e. Forecast verification

The continuous ranked probability score (CRPS) (Grimm et al. 2006), a surrogate measure of forecast reliability, sharpness, and efficiency (Hersbach 2000), is used to evaluate the overall forecast skill for each univariate forecast. To conveniently compare a forecast among different target months, we normalize CRPS with respect to a reference forecast and thereafter define the “CRPS skill score” as

$$\text{CRPS skill score} = \frac{\text{CRPS}_{\text{REF}} - \text{CRPS}}{\text{CRPS}_{\text{REF}}}, \quad (10)$$

where CRPS_{REF} is the CRPS calculated from a reference forecast. The reference forecast used in this study is a cross-validation version of the climatology forecast. For example, in the case of generating the reference forecasts of January in 2000, we use the historical observations from January other than year 2000 (i.e., from 1992 to 1999 and from 2001 to 2012 for our study) to generate a reference forecast with 22 ensemble members corresponding to 22 cross-validated years (see section 3 for a detailed data description). A positive CRPS skill score suggests a skillful forecast (with respect to the reference forecast), and a perfectly skillful

forecast is indicated by the maximum of CRPS skill score being 1.

To complement overall forecast skill assessment to evaluate univariate ensemble forecasts, we consider three additional performance measures suggested by Engeland et al. (2010) to test for important attributes of ensemble (or probabilistic) forecasts including efficiency, reliability, and sharpness. Efficiency is simply measured by the relative forecast bias, defined by the difference between the ensemble mean and observations normalized by climatology. Relative bias checks whether a forecast tends to make an over or underestimation (indicated by positive or negative biases) and how much forecast mean is deviated from observations. Reliability (also called calibration) examines how well a forecast can reproduce the frequency of an event and is measured by the uniformity of forecast integral transform (PIT), defined by

$$\text{PIT} = \text{CDF}_f(X_o), \quad (11)$$

where CDF_f is the cumulative distribution function of an ensemble forecast of a univariate variable X and X_o is the corresponding observed value of X . If a forecast is reliable, PIT will follow a uniform distribution over $[0, 1]$. We evaluate the forecast reliability by visually checking the uniformity of the PIT histogram, which is also quantified by the α index (Renard et al. 2010), formulated as

$$\alpha = 1 - \frac{2}{n} \sum_{i=1}^n \left| \text{PIT}_i^* - \frac{t}{n+1} \right|, \quad (12)$$

where PIT_i^* is the sorted PIT_i in ascending order. The range of the α index is from 0 (worst reliability) to 1 (perfect reliability). Sharpness is a measure of ensemble dispersion, and a sharp forecast is more decisive and more certain. Note that sharpness is desirable only if a forecast is reliable. Sharpness is indicated by the average width of the 95% confidence interval (95% AWCI). To facilitate the comparison over different climate variables, we standardize the 95% AWCI by the standard deviation of observations to make it unitless.

In addition to univariate ensemble forecast skill assessments, we also consider multivariate ensemble forecast skill assessments because all forecasts of four climate variables need to be used jointly to drive crop models. Generalized from CRPS, the energy score (ES) (Gneiting and Raftery 2007; Gneiting et al. 2008) is a measure of statistical discrepancy between two multivariate distribution and is used to evaluate the skill of multivariate ensemble forecasts. That is, the CRPS assesses the individual elements of the forecast

FIG. 2. The locations of the three stations in Australia that are used in this study.

variables in the univariate sense, whereas the ES assesses all forecast variables jointly (including interdependence structure between different variables). We define the ES for a multivariate ensemble forecast P of an observed vector $\mathbf{x}$ by

$$ES(P, \mathbf{x}) = \frac{1}{2} E \|\mathbf{X} - \mathbf{X}'\|^\beta - E \|\mathbf{X} - \mathbf{x}\|^\beta, \quad (13)$$

where $\mathbf{X}$ and $\mathbf{X}'$ are two independent realizations of the forecast P and $\beta \in (0, 2)$ is a shape parameter. We choose $\beta = 1$ in this study, and the ES becomes a direct multivariate generalization of the CRPS. The ES is a strictly proper score and is a surrogated metric of forecast calibration, sharpness, and efficiency in the context of multivariate forecasting (Gneiting and Raftery 2007). Analogous to the CRPS, we normalize the ES with respect to the reference forecast described earlier and define the “ES skill score” as

$$ES \text{ skill score} = \frac{ES_{\text{REF}} - ES}{ES_{\text{REF}}}, \quad (14)$$

where ES_{REF} is the ES calculated from the reference forecast mentioned earlier.

3. Study stations and data

We test our approach at three stations from distinct climate zones in Australia (Fig. 2). The three study stations are Birchip from a grassland climate zone in Victoria, Parkes from a temperate climate zone in New South Wales, and Mingenew from a subtropical climate zone in Western Australia. All three of the stations are surrounded by rich wheat farming lands and seasonal climate forecasts at these stations are of great interest to

agricultural planning. In the future research, we will examine more climate regions in Australia, such as the central-northern Queensland region where highly variable climate and climate change beyond natural variability have great impacts on agricultural sectors.

The daily average of four climate variables are forecast at lead times of 0–5 months in this study, including rainfall, minimum temperature T_{min} , maximum temperature T_{max} , and shortwave radiation (referred to as radiation). These four variables at a daily scale are commonly used as core input variables of a crop model (such as the APSIM model; Keating et al. 2003) for crop yield simulation. In the future work, we will disaggregate monthly forecasts to daily steps prior to using them as inputs for a crop model. We use the “SILO” patched point data as observations of these variables (obtained free of charge from Queensland Government, licensed under Creative Commons Attribution 4.0, at <https://legacy.longpaddock.qld.gov.au/silo>). The predictor variables are based on model outputs from the seasonal prediction version of ACCESS-S (Hudson et al. 2017). The ACCESS-S model is a coupled general circulation model developed and tested by the Bureau of Meteorology Australia with the key collaboration of the Met Office, based on its Global Seasonal Forecast System, version 5 (GloSea5). ACCESS-S produces daily forecasts of many atmospheric climate variables (including all the four variables used in this study) on grid points at the resolution of 60 km. The version of ACCESS-S used in this study outputs 11 ensemble members per start date with lead times of 0–5 months. The hindcasts of ACCESS-S are available at the period of 1990–2012 (i.e., 23 yr).

We choose to harness the raw forecasts and two ACCESS-S-derived sea surface temperature (SST) climate indices to form the CPP-E forecast for each climate variable as described in section 2c. These two SST climate indices are the El Niño–Southern Oscillation 3.4 (Niño-3.4) and the Indian Ocean dipole (IOD) (Saji et al. 1999), which have been linked to Australian climate in previous studies (Cai et al. 2011; Risbey et al. 2009).

The daily averages of each climate variable are post-processed with lead times of 0–5 months by the BC, QM, and CPP methods described in section 2. The lead time is defined by the time between the initialization date of the ACCESS forecast (e.g., 1 January 1990) and the start date of the forecast period (e.g., 1 February 1990). For example, the 0-month-lead-time rainfall forecast issued at 1 January 1990 is the forecast of monthly average rainfall from 1 January 1990 to 31 January 1990. We evaluate the skills of downscaled forecasts at the three stations mentioned in section 3. We use the entire ACCESS-S hindcasts from 1990 to 2012 in skill verification

and perform leave-one-year-out cross validation to avoid potential overfitting.

4. Results

a. Overall forecast skill

At each site we evaluate the univariate forecast for each of the three methods (BC, QM, and CPP) using four performance measures (Fig. 3). For all variables except rainfall, QM and CPP produce forecasts with a bias of near to zero and BC leads to exact zero bias because of the fact that BC corrects the quantity that is evaluated (i.e., the mean bias). For rainfall, the median bias of BC is still close to zero but some positive bias is observed because negative BC rainfall forecasts are forced back to zero. The median bias of CPP (2%–4%) is substantially smaller than that of QM (8%–13%).

CPP leads to the most reliable forecasts with the greatest α index on average (close to 95%), and the difference of the α index among various climate variables and sites is generally small for CPP. The BC method has marginally better forecast reliability on average than QM but leads to a greater dispersion in the α index. We further check the PIT histogram for different postprocessed forecasts in Fig. 4. To obtain a sufficiently large sample size, we pool the forecasts from all sites, lead times, and target months together to make the PIT histogram. Figure 4 confirms that CPP is almost perfectly reliable for all four variables, suggested by almost perfectly uniform PIT histogram. The only exception is that CPP is slightly underdispersive (also called overconfident) for small rainfall. The strongly U-shaped PIT histogram in the first and second rows of Fig. 4 suggests that BC and QM are far from reliable. A highly underdispersive character is found for BC and QM as verification observations often fall either below or above all ensembles. In conclusion, the PIT histogram as well as the α index indicate that CPP improves reliability compared to BC and QM.

The standardized 95% AWCI in the third row of Fig. 3 indicates that CPP is less sharp than BC or QM. This is consistent with our findings of forecast reliability. CPP corrects the problem of underdispersion in raw forecasts by enhancing the ensemble dispersion to include more verification observations in ensembles. Instead, QM does not identify the underdispersion but further reduces ensemble dispersion.

As seen in the fourth row of Fig. 3, CPP leads to consistently greater overall CRPS skill score than BC and QM, indicating that CPP produces more skillful univariate forecasts than BC and QM. Furthermore, CRPS skill score of CPP has a smaller dispersion than

those of BC and QM. The reduced dispersion of CRPS skill score of the CPP forecasts is to be expected, because CPP is the only method that accounts—in a convoluted way—for limited predictability. As CPP has the tendency to only retain predictable variability and will thereby produce forecasts that are less variable on average and hence reduce variability in scores as well. From the median CRPS skill score, we can see that CPP leads to more skillful overall univariate forecasts than the climatology forecasts by 3%–5% for $T_{\max}$, $T_{\min}$ and radiation. Rainfall is the most difficult variable to forecast and CPP does not show better overall forecast skill than the climatology forecast for rainfall.

Extending to multivariate forecast evaluation (Fig. 5), we still find that CPP is the most skillful among these three methods. The median ES skill score calculated from the BC and QM forecasts is negative and it implies that BC and QM multivariate forecasts are not generally as skillful as the climatology counterpart. Instead, the CPP multivariate forecast is at least as skillful as the climatology forecast. We can observe that the SS indeed improves the multivariate forecast skill though in a small magnitude.

b. Impact of target month and lead time on forecast skill

Forecast skill can vary with different target months and lead times, which we now explore. In the previous subsection, we compared the overall forecast skill of different postprocessing methods for all target months and lead times. In this section, we investigate the impact of lead time and target month on forecast skill indicated by CRPS skill score. Figure 6 presents CRPS skill score as a function of lead time while averaging over all target months. As expected, the univariate forecast skill tends to decrease with increasing lead times. At 0-month lead time, all the three postprocessing methods lead to positive CRPS skill score for all the variables except for rainfall where BC and QM lead to zero to negative skill at all lead times. At lead times of 1–5 months, the univariate forecasts from BC and QM lose skill quickly and are often not as skillful as the reference forecast. Instead, CPP maintains positive skill at increasing lead times for all variables except for rainfall. Figure 7 presents CRPS skill score as a function of target month while averaging over all lead times. The forecast skill varies greatly across 12 months and three sites. CPP overperforms BC and QM at most target months and exhibits a smaller variability in CRPS skill score over different target months than the two counterparts. Detailed results on the impact of lead time and target

FIG. 3. A comparison of relative bias, alpha index, standardized average width of the 95% confidence interval (standardized 95% AWCI), and CRPS skill score calculated from bias-corrected (BC), quantile-mapped (QM), and copula-based postprocessed (CPP) forecasts. Each boxplot is drawn on the basis of the performance measures pooled for all test years, target months, and lead times. The performance measures can be interpreted as follows: 1) an unbiased forecast is indicated by relative bias close to 0; 2) a reliable forecast is indicated by large alpha index close to 100%; 3) a sharp forecast is indicated by small standardized 95% AWCI; 4) a skillful forecast is indicated by large CRPS skill score close to 100% (0 implies a forecast with no overall skill improvement, and a negative value implies a degraded forecast).

month on forecast skill for each site can be found in the online supplemental materials.

c. Comparing CPP with CPP-E and CPP-M

Figure 8 presents the CRPS skill score calculated from the CPP-E and CPP-M forecasts as a function of lead

month. Except for rainfall we observe positive forecast skill for CPP-E and CPP-M for all variables. However, we do not find an evidence of consistent skill improvement of CPP-E and CPP-M compared to CPP. The impact of two external predictors (Niño-3.4 and IOD) on forecast skill is site dependent. For example, the

FIG. 4. The PIT histogram calculated from bias-corrected, quantile-mapped, and copula-based postprocessed forecasts. All sites, lead times, and target months are pooled for each PIT histogram. A uniform PIT histogram indicates a reliable forecast.

inclusion of Niño-3.4 and IOD leads an improvement for $T_{\min}$ at the Parkes site. A further investigation is required to explore the effect of including other large-scale climate indices. The difference between CPP and CPP-M indicates the value of utilizing the exact value of individual ensemble member instead of ensemble mean, which is often more significant at long lead times. CPP-M works like or better than CPP at 0-month lead time for which forecast uncertainty is often small and a less complex model may lead to better performance. Detailed comparisons of CPP, CPP-E, and CPP-M at

each lead time and each target month can be found in the online supplemental materials.

5. Discussion

The skill improvement of CPP compared to BC and QM in the univariate context is contributed by the following three factors: 1) CPP is effective for bias removal. Copulas implicitly account for the potential nonlinear relationship between GCM forecasts and local observations, which cannot be addressed by a linear-scale

FIG. 5. A comparison of ES skill score calculated from bias-corrected, quantile-mapped, and copula-based postprocessed forecasts. Each boxplot is drawn on the basis of the performance measures for all years, target months, and lead times. A greater ES skill score indicates a more skillful forecast (100% implies a perfectly skillful forecast, 0% implies a forecast with no skill, and a negative value implies a degraded forecast).

factor in BC. QM is based on the oversimplified assumption that raw forecasts are strongly rank correlated with observations. When raw forecasts have weak or even negative correlation with observations, the QM assumption is invalid and CPP leads to better bias removal than QM (as shown in Fig. 3). 2) CPP is designed to generate fully reliable forecasts. CPP aims to predict the forecast distribution conditioning on raw forecasts and the use of copulas makes CPP flexible to model multivariate dependence structure beyond multivariate Gaussian. BC does not change the ensemble spread and cannot correct the overdispersion or underdispersion of the raw forecasts at all. QM also cannot ensure forecast reliability as seen from previous studies (Zhao et al. 2017). 3) CPP is flexible to increase or decrease the ensemble size. CPP essentially predicts a smooth forecast distribution and discretises it to a set of ensemble members. Verification scores (such as the Brier score and CRPS) can be dependent on the ensemble size and a larger ensemble size can lead to a greater verification score. We follow a suggestion from an anonymous reviewer and calculate the ensemble-size-adjusted CRPS (Ferro et al. 2008) by using the R package SpecsVerification, as shown in Fig. 9. Note that we keep the CRPS skill score for BC and QM the same for all ensemble sizes in Fig. 9 for two reasons: 1) Once an ensemble forecast system (e.g., BC and QM) is made, verification score will correctly reflect the accuracy of the forecast system (Casati et al. 2008). 2) Although we can apply random permutations to increase ensemble size for BC and QM artificially, this will remain the

forecast distribution as well as the verification score (e.g., CRPS). The skill gain for CPP is not observed when we keep the ensemble size of BC and QM forecasts (i.e., ensemble size = 11) or even double it (i.e., ensemble size = 22). Because CPP predicts the conditional forecast distribution at each of 11 ensemble members, an ensemble size of 11 or 22 implies that we only generate one or two random samples from each conditional forecast distribution. Given that forecast uncertainty is often large, one or two random samples are far from enough to represent the conditional forecast distribution. Nevertheless, practitioners usually prefer a small ensemble size to reduce the effect on storage and computation for crop simulation. We feel that an ensemble size of 110 configured for CPP is a good practical choice.

In the context of multivariate forecasts, the small skill improvement of CPP is mainly from the transfer of skill gain in the univariate forecasts. The effectiveness of the SS to transfer independent univariate forecasts to multivariate forecasts is limited. Other methods can be considered to establish the multivariate dependence in the future research. For example, the SS does not capture the multivariate dependence conditional on the state of atmosphere but on the forecast issue date alone (Verkade et al. 2013). To address this issue, we could apply the ensemble copula coupling developed by Schefzik et al. (2013) to essentially sort the ensemble members according to the order in the raw forecasts. Several state-of-the-art multivariate post-processing methods originally designed for deterministic

FIG. 6. CRPS skill score as a function of lead time for three postprocessing methods. The CRPS skill score presented here is averaged over all target months. A larger CRPS skill score indicates a more skillful forecast (100% implies a perfectly skillful forecast, 0% implies a forecast with no skill, and a negative value implies a degraded forecast).

FIG. 7. CRPS skill score as a function of target month for three postprocessing methods. The CRPS skill score presented here is averaged over all lead times. A larger CRPS skill score indicates a more skillful forecast (100% implies a perfectly skillful forecast, 0% implies a forecast with no skill, and a negative value implies a degraded forecast).

FIG. 8. CRPS skill score for CPP-E and CPP-M as a function of lead time. A larger CRPS skill score indicates a more skillful forecast (100% implies a perfectly skillful forecast, 0% implies a forecast with no skill, and a negative value implies a degraded forecast).

forecasts, including MBCn (Cannon 2018), R^2D^2 (Vrac 2018), and dOTC (Robin et al. 2019), are good inspiration for the further development of CPP for formally establishing the dependence structure in multivariate

ensemble forecasts. Furthermore, the forecast verification of multivariate forecasts is preliminary and more comprehensive assessments should be considered in the future work. For example, the variogram score

FIG. 9. Ensemble-size-adjusted CRPS skill score as a function of ensemble size. CRPS skill scores are averaged over all sites, lead times, and target months. A larger CRPS skill score indicates a more skillful forecast (100% implies a perfectly skillful forecast, 0% implies a forecast with no skill, and a negative value implies a degraded forecast).

proposed by [Scheuerer and Hamill \(2015\)](#) considers pairwise differences of the components at each dimension and is known to be more sensitive to misspecification of a dependence structure than the energy score.

Another possibility is that we could use PCCs to directly model the space, time, and variable dependence in CPP instead of the SS. In this case, we could derive the joint distribution of four climate variables conditioned on the corresponding raw forecasts and produce a multivariate ensemble forecast directly. However, this alternative can substantially increase computing time on estimating the parameters of PCCs, which would not be practical for a real-time forecasting exercise. In the abovementioned example, we have to fit an eight-dimensional C-vine copula conditioned on four variables. Following the discussion by [Bevacqua et al. \(2017\)](#), the number of pair copula decompositions is $4! \times 4! = 576$. We keep a parsimonious model for computational convenience. In the future research, we would link a few

strongly associated variables and derive the joint forecast distribution from a low-dimensional copula.

In this study, we perform CPP at the level of each ensemble member and treat each ensemble member as a deterministic forecast. This relies on the assumption that ensemble members are traceable over time (i.e., member 1 at a given initialization date from different years are statistically consistent). This assumption is vaguely valid for the ACCESS-S model and current operational postprocessing (e.g., BC and QM) is carried out in the same fashion as CPP. When CPP is extended to other climate models or in a different context, we would suggest that the user exercises caution to check whether it is wise to treat ensemble members individually instead of as a whole. [Möller et al. \(2018\)](#) provided a valuable alternative to the setup of CPP that uses ensemble dispersion more comprehensively.

We have seen that it is a challenging task to use additional GCM forecasts as “informative” predictors in

CPP. The added value of the simple inclusion of Niño-3.4 and IOD is limited and largely varies across locations and target months. Many previous research (Cai et al. 2011; Risbey et al. 2009) found the correlation between large-scale climate indices and regional rainfall and temperature in Australia, but these correlation studies were based on climate indices from reanalysis datasets instead from GCM model outputs. It is still elusive to use additional GCM forecasts in CPP. Some variable selection analysis, such as Schepen et al. (2012), would be carried out in the future study to investigate when and how to use additional GCM forecasts in CPP.

The CPP proposed in this paper is generally applicable to various SCF models. It could preprocess the climate models with daily data available such as ACCESS-S (Hudson et al. 2017), but it also could postprocess climate models with monthly level or seasonal-level forecasts only, such as North American Multimodel Ensemble Forecasts (Kirtman et al. 2014). In essence, what we have achieved is to take an existing seasonal climate forecast dataset that performs only marginally better than climatology and offer improvements to it that will potentially allow agricultural scientists to make better yield prediction. They can have confidence that they are retrieving maximized skill in climate forecasts in formats to grow crops effectively and make crucial evidence-based management decisions.

6. Conclusions

We have developed a new copula-based postprocessing method to generate fully calibrated multivariate seasonal climate forecasts, which can be potentially used for yield prediction in conjunction with appropriate temporal disaggregation. Our method uses copulas to characterize the joint probability distribution of seasonal climate model forecasts and the corresponding local observations. We perform CPP independently for each climate variable at the level of each ensemble member and use the Schaake shuffle to reorder ensemble sequence for more realistic cross-variable dependence structures. CPP can be extended with the inclusion of the forecasts of some large-scale climate indices as additional predictors, and in this case pair-copula constructions are used to construct multivariate probability distributions. We apply the method to postprocess monthly rainfall, minimum temperature, maximum temperature, and radiation forecasts from the Australian Community Climate and Earth-System Simulator Seasonal model. We evaluate forecast skill at three representative Australian sites and compare CPP with bias correction and quantile mapping on a cross-validation theme. We make the following conclusions:

- 1) CPP leads to better overall forecast skill than BC and QM in the context of both univariate and multivariate skill assessment.
- 2) The CPP univariate forecasts are more skillful than the climatology forecast at all lead times for all variables except for rainfall.
- 3) CPP is effective to generate reliable forecasts, while the BC and QM forecasts are underdispersive.
- 4) The simple inclusion of Niño-3.4 and IOD provides limited added value to forecast skill.
- 5) The exact value of ensemble members indeed provides more information than the ensemble mean in postprocessing ACCESS-S forecasts.

Acknowledgments. This research has been supported by the CSIRO Digiscape future science platform. We thank Dr. Andrew Schepen for valuable suggestions to improve the quality of the paper. We are grateful to three anonymous reviewers and the editor for providing constructive comments and suggestions.

REFERENCES

- Aas, K., C. Czado, A. Frigessi, and H. Bakken, 2009: Pair-copula constructions of multiple dependence. *Insur. Math. Econ.*, **44**, 182–198, <https://doi.org/10.1016/j.insmatheco.2007.02.001>.
- Bedford, T., and R. M. Cooke, 2001: Probability density decomposition for conditionally dependent random variables modeled by vines. *Ann. Math. Artif. Intell.*, **32**, 245–268, <https://doi.org/10.1023/A:1016725902970>.
- , and —, 2002: Vines—A new graphical model for dependent random variables. *Ann. Stat.*, **30**, 1031–1068, <https://doi.org/10.1214/aos/1031689016>.
- Bevacqua, E., D. Maraun, I. H. Haff, M. Widmann, and M. Vrac, 2017: Multivariate statistical modelling of compound events via pair-copula constructions: Analysis of floods in Ravenna (Italy). *Hydrol. Earth Syst. Sci.*, **21**, 2701–2723, <https://doi.org/10.5194/hess-21-2701-2017>.
- Brown, J. N., Z. Hochman, D. Holzworth, and H. Horan, 2018: Seasonal climate forecasts provide more definitive and accurate crop yield predictions. *Agric. For. Meteorol.*, **260–261**, 247–254, <https://doi.org/10.1016/j.agrformet.2018.06.001>.
- Cai, W. J., P. van Rensch, T. Cowan, and H. H. Hendon, 2011: Teleconnection pathways of ENSO and the IOD and the mechanisms for impacts on Australian rainfall. *J. Climate*, **24**, 3910–3923, <https://doi.org/10.1175/2011JCLI4129.1>.
- Cannon, A. J., 2018: Multivariate quantile mapping bias correction: An N-dimensional probability density function transform for climate model simulations of multiple variables. *Climate Dyn.*, **50**, 31–49, <https://doi.org/10.1007/s00382-017-3580-6>.
- Casati, B., and Coauthors, 2008: Forecast verification: Current status and future directions. *Meteor. Appl.*, **15**, 3–18, <https://doi.org/10.1002/met.52>.
- Centre for International Economics, 2014: Analysis of the benefits of improved seasonal climate forecasting for agriculture. Managing Climate Variability Program Rep., 50 pp., <http://www.climatekelpie.com.au/Files/MCV-CIE-report-Value-of-improved-forecasts-agriculture-2014.pdf>.

- Challinor, A. J., J. M. Slingo, T. R. Wheeler, and F. J. Doblas-Reyes, 2005: Probabilistic simulations of crop yield over western India using the DEMETER seasonal hindcast ensembles. *Tellus*, **57A**, 498–512, <https://doi.org/10.1111/j.1600-0870.2005.00126.x>.
- Clark, M., S. Gangopadhyay, L. Hay, B. Rajagopalan, and R. Wilby, 2004: The Schaake shuffle: A method for reconstructing space–time variability in forecasted precipitation and temperature fields. *J. Hydrometeorol.*, **5**, 243–262, [https://doi.org/10.1175/1525-7541\(2004\)005<0243:TSSAMF>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0243:TSSAMF>2.0.CO;2).
- Crimp, S., H. Jin, P. Kokic, S. Bakar, and N. Nicholls, 2019: Possible future changes in south east Australian frost frequency: An inter-comparison of statistical downscaling approaches. *Climate Dyn.*, **52**, 1247–1262, <https://doi.org/10.1007/s00382-018-4188-1>.
- Doblas-Reyes, F. J., J. Garcia-Serrano, F. Lienert, A. P. Biescas, and L. R. L. Rodrigues, 2013: Seasonal climate predictability and forecasting: Status and prospects. *Wiley Interdiscip. Rev. Climate Change*, **4**, 245–268, <https://doi.org/10.1002/wcc.217>.
- Donohue, R. J., R. A. Lawes, G. Mata, D. Gobbett, and J. Ouzman, 2018: Towards a national, remote-sensing-based model for predicting field-scale crop yield. *Field Crops Res.*, **227**, 79–90, <https://doi.org/10.1016/j.fcr.2018.08.005>.
- Engeland, K., B. Renard, I. Steinsland, and S. Kolberg, 2010: Evaluation of statistical models for forecast errors from the HBV model. *J. Hydrol.*, **384**, 142–155, <https://doi.org/10.1016/j.jhydrol.2010.01.018>.
- Farchi, A., M. Bocquet, Y. Roustan, A. Mathieu, and A. Querel, 2016: Using the Wasserstein distance to compare fields of pollutants: Application to the radionuclide atmospheric dispersion of the Fukushima-Daiichi accident. *Tellus*, **68B**, 31682, <https://doi.org/10.3402/tellusb.v68.31682>.
- Ferro, C. A. T., D. S. Richardson, and A. P. Weigel, 2008: On the effect of ensemble size on the discrete and continuous ranked probability scores. *Meteor. Appl.*, **15**, 19–24, <https://doi.org/10.1002/met.45>.
- Frees, E. W., and E. A. Valdez, 1998: Understanding relationships using copulas. *North Amer. Actuar. J.*, **2** (1), 1–25, <https://doi.org/10.1080/10920277.1998.10595667>.
- Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203–1211, [https://doi.org/10.1175/1520-0450\(1972\)011<1203:TUOMOS>2.0.CO;2](https://doi.org/10.1175/1520-0450(1972)011<1203:TUOMOS>2.0.CO;2).
- Gneiting, T., and A. E. Raftery, 2007: Strictly proper scoring rules, prediction, and estimation. *J. Amer. Stat. Assoc.*, **102**, 359–378, <https://doi.org/10.1198/016214506000001437>.
- , F. Balabdaoui, and A. E. Raftery, 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>.
- , L. I. Stanberry, E. P. Grit, L. Held, and N. A. Johnson, 2008: Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *Test*, **17**, 211–235, <https://doi.org/10.1007/s11749-008-0114-x>.
- Graham, R. J., M. Gordon, P. J. Mclean, S. Ineson, M. R. Huddleston, M. K. Davey, A. Brookshaw, and R. T. H. Barnes, 2005: A performance comparison of coupled and uncoupled versions of the Met Office seasonal prediction general circulation model. *Tellus*, **57A**, 320–339, <https://doi.org/10.1111/j.1600-0870.2005.00116.x>.
- Grillakis, M. G., A. G. Koutroulis, and I. K. Tsanis, 2013: Multisegment statistical bias correction of daily GCM precipitation output. *J. Geophys. Res. Atmos.*, **118**, 3150–3162, <https://doi.org/10.1002/JGRD.50323>.
- Grit, E. P., T. Gneiting, V. J. Berrocal, and N. A. Johnson, 2006: The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification. *Quart. J. Roy. Meteor. Soc.*, **132**, 2925–2942, <https://doi.org/10.1256/qj.05.235>.
- Hammer, G. L., N. Nicholls, and C. D. Mitchell, 2000: *Applications of Seasonal Climate Forecasting in Agricultural and Natural Ecosystems: The Australian Experience*. Springer Science and Business Media, 472 pp.
- Hansen, J. W., and A. V. M. Ines, 2005: Stochastic disaggregation of monthly rainfall data for crop simulation studies. *Agric. For. Meteorol.*, **131**, 233–246, <https://doi.org/10.1016/j.agrformet.2005.06.006>.
- Hawkins, E., T. M. Osborne, C. K. Ho, and A. J. Challinor, 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. *Wea. Forecasting*, **15**, 559–570, [https://doi.org/10.1175/1520-0434\(2000\)015<0559:DOTCRP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2).
- Ho, C. K., D. B. Stephenson, M. Collins, C. A. T. Ferro, and S. J. Brown, 2012: Calibration strategies a source of additional uncertainty in climate change projections. *Bull. Amer. Meteor. Soc.*, **93**, 21–26, <https://doi.org/10.1175/2011BAMS3110.1>.
- Hudson, D., A. G. Marshall, Y. H. Yin, O. Alves, and H. H. Hendon, 2013: Improving intraseasonal prediction with a new ensemble generation strategy. *Mon. Wea. Rev.*, **141**, 4429–4449, <https://doi.org/10.1175/MWR-D-13-00059.1>.
- , and Coauthors, 2017: ACCESS-S1: The new Bureau of Meteorology multi-week to seasonal prediction system. *J. South. Hemisphere Earth Syst. Sci.*, **67**, 132–159, <https://doi.org/10.22499/3.6703.001>.
- Huntingford, C., F. H. Lambert, J. H. C. Gash, C. M. Taylor, and A. J. Challinor, 2005: Aspects of climate change prediction relevant to crop productivity. *Philos. Trans. Roy. Soc.*, **B360**, 1999–2009, <https://doi.org/10.1098/rstb.2005.1748>.
- Ines, A. V. M., and J. W. Hansen, 2006: Bias correction of daily GCM rainfall for crop simulation studies. *Agric. For. Meteorol.*, **138**, 44–53, <https://doi.org/10.1016/j.agrformet.2006.03.009>.
- , —, and A. W. Robertson, 2011: Enhancing the utility of daily GCM rainfall for crop yield prediction. *Int. J. Climatol.*, **31**, 2168–2182, <https://doi.org/10.1002/joc.2223>.
- Keating, B. A., and Coauthors, 2003: An overview of APSIM, a model designed for farming systems simulation. *Eur. J. Agron.*, **18**, 267–288, [https://doi.org/10.1016/S1161-0301\(02\)00108-9](https://doi.org/10.1016/S1161-0301(02)00108-9).
- Kirtman, B. P., and Coauthors, 2014: The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, **95**, 585–601, <https://doi.org/10.1175/BAMS-D-12-00050.1>.
- Kokic, P., H. D. Jin, and S. Crimp, 2013: Improved point scale climate projections using a block bootstrap simulation and quantile matching method. *Climate Dyn.*, **41**, 853–866, <https://doi.org/10.1007/s00382-013-1791-z>.
- Li, H. B., J. Sheffield, and E. F. Wood, 2010: Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *J. Geophys. Res.*, **115**, D10101, <https://doi.org/10.1029/2009JD012882>.

- Madadgar, S., and H. Moradkhani, 2013: Drought analysis under climate change using copula. *J. Hydrol. Eng.*, **18**, 746–759, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000532](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000532).
- Maraun, D., 2013: Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. *J. Climate*, **26**, 2137–2143, <https://doi.org/10.1175/JCLI-D-12-00821.1>.
- , 2018: *Statistical Downscaling and Bias Correction for Climate Research*. Cambridge University Press, 347 pp.
- Mavromatis, T., and P. D. Jones, 1999: Evaluation of HadCM2 and direct use of daily GCM data in impact assessment studies. *Climatic Change*, **41**, 583–614, <https://doi.org/10.1023/A:1005336608651>.
- McCrea, R., L. Dalglish, and W. Coventry, 2005: Encouraging use of seasonal climate forecasts by farmers. *Int. J. Climatol.*, **25**, 1127–1137, <https://doi.org/10.1002/joc.1164>.
- McIntosh, P. C., A. J. Ash, and D. M. Stafford Smith, 2005: From oceans to farms: The value of a novel statistical climate forecast for agricultural management. *J. Climate*, **18**, 4287–4302, <https://doi.org/10.1175/JCLI3515.1>.
- , M. J. Pook, J. S. Risbey, S. N. Lisson, and M. Rebbeck, 2007: Seasonal climate forecasts for agriculture: Towards better understanding and value. *Field Crops Res.*, **104**, 130–138, <https://doi.org/10.1016/j.fcr.2007.03.019>.
- Meinke, H., and R. Stone, 2005: Seasonal and inter-annual climate forecasting: The new tool for increasing preparedness to climate variability and change in agricultural planning and operations. *Climatic Change*, **70**, 221–253, <https://doi.org/10.1007/s10584-005-5948-6>.
- Michelangeli, P. A., M. Vrac, and H. Loukos, 2009: Probabilistic downscaling approaches: Application to wind cumulative distribution functions. *Geophys. Res. Lett.*, **36**, L11708, <https://doi.org/10.1029/2009GL038401>.
- Möller, A., L. Spazzini, D. Kraus, T. Nagler, and C. Czado, 2018: Vine copula based post-processing of ensemble forecasts for temperature. arXiv, <https://arxiv.org/abs/1811.02255>.
- Murphy, A. H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281–293, [https://doi.org/10.1175/1520-0434\(1993\)008<0281:WIAGFA>2.0.CO;2](https://doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2).
- Nelsen, R. B., 2006: *An Introduction to Copulas*. 2nd ed. Springer, 269 pp.
- Olsson, T., J. Jakkila, N. Veijalainen, L. Backman, J. Kaurola, and B. Vehviläinen, 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**, 3217–3238, <https://doi.org/10.5194/hess-19-3217-2015>.
- Osborne, T., G. Rose, and T. Wheeler, 2013: Variation in the global-scale impacts of climate change on crop productivity due to climate model uncertainty and adaptation. *Agric. For. Meteorol.*, **170**, 183–194, <https://doi.org/10.1016/j.agrformet.2012.07.006>.
- Paolino, D. A., J. L. Kinter, B. P. Kirtman, D. H. Min, and D. M. Straus, 2012: The impact of land surface and atmospheric initialization on seasonal forecasts with CCSM. *J. Climate*, **25**, 1007–1021, <https://doi.org/10.1175/2011JCLI3934.1>.
- Peng, B., K. Y. Guan, M. Pan, and Y. Li, 2018: Benefits of seasonal climate prediction and satellite data for forecasting US maize yield. *Geophys. Res. Lett.*, **45**, 9662–9671, <https://doi.org/10.1029/2018GL079291>.
- Piani, C., and J. O. Haerter, 2012: Two dimensional bias correction of temperature and precipitation copulas in climate models. *Geophys. Res. Lett.*, **39**, L20401, <https://doi.org/10.1029/2012GL053839>.
- , —, and E. Coppola, 2010: Statistical bias correction for daily precipitation in regional climate models over Europe. *Theor. Appl. Climatol.*, **99**, 187–192, <https://doi.org/10.1007/s00704-009-0134-9>.
- Renard, B., D. Kavetski, G. Kuczera, M. Thyer, and S. W. Franks, 2010: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resour. Res.*, **46**, W05521, <https://doi.org/10.1029/2009WR008328>.
- Risbey, J. S., M. J. Pook, P. C. McIntosh, M. C. Wheeler, and H. H. Hendon, 2009: On the remote drivers of rainfall variability in Australia. *Mon. Wea. Rev.*, **137**, 3233–3253, <https://doi.org/10.1175/2009MWR2861.1>.
- Robin, Y., M. Vrac, P. Naveau, and P. Yiou, 2019: Multivariate stochastic bias corrections with optimal transport. *Hydrol. Earth Syst. Sci.*, **23**, 773–786, <https://doi.org/10.5194/hess-23-773-2019>.
- Rodriguez, D., P. de Voil, D. Hudson, J. N. Brown, P. Hayman, H. Marrou, and H. Meinke, 2018: Predicting optimum crop designs using crop models and seasonal climate forecasts. *Sci. Rep.*, **8**, 2231, <https://doi.org/10.1038/s41598-018-20628-2>.
- Saha, S., and Coauthors, 2006: The NCEP Climate Forecast System. *J. Climate*, **19**, 3483–3517, <https://doi.org/10.1175/JCLI3812.1>.
- Saji, N. H., B. N. Goswami, P. N. Vinayachandran, and T. Yamagata, 1999: A dipole mode in the tropical Indian Ocean. *Nature*, **401**, 360–363, <https://doi.org/10.1038/43854>.
- Santambrogio, F., 2015: *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*. Birkhäuser, 353 pp.
- Schefzik, R., T. L. Thorarindottir, and T. Gneiting, 2013: Uncertainty quantification in complex simulation models using ensemble copula coupling. *Stat. Sci.*, **28**, 616–640, <https://doi.org/10.1214/13-STS443>.
- Schepen, A., and Q. J. Wang, 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>.
- , —, and D. Robertson, 2012: Evidence for using lagged climate indices to forecast Australian seasonal rainfall. *J. Climate*, **25**, 1230–1246, <https://doi.org/10.1175/JCLI-D-11-00156.1>.
- , —, and D. E. Robertson, 2014: Seasonal forecasts of Australian rainfall through calibration and bridging of coupled GCM outputs. *Mon. Wea. Rev.*, **142**, 1758–1770, <https://doi.org/10.1175/MWR-D-13-00248.1>.
- , —, and Y. Everingham, 2016: Calibration, bridging, and merging to improve GCM seasonal temperature forecasts in Australia. *Mon. Wea. Rev.*, **144**, 2421–2441, <https://doi.org/10.1175/MWR-D-15-0384.1>.
- Scheuerer, M., and T. M. Hamill, 2015: Statistical postprocessing of ensemble precipitation forecasts by fitting censored, shifted gamma distributions. *Mon. Wea. Rev.*, **143**, 4578–4596, <https://doi.org/10.1175/MWR-D-15-0061.1>.
- Schölzel, C., and P. Friederichs, 2008: Multivariate non-normally distributed random variables in climate research—Introduction to the copula approach. *Nonlinear Processes Geophys.*, **15**, 761–772, <https://doi.org/10.5194/npg-15-761-2008>.
- Sklar, A., 1959: Fonctions de répartition à n dimensions et leurs marges (n dimensional distribution functions and their margins). Vol. 8, Publications de l'Institut de Statistique de l'Université de Paris, 229–231.
- Stone, R. C., and H. Meinke, 2005: Operational seasonal forecasting of crop performance. *Philos. Trans. Roy. Soc.*, **B360**, 2109–2124, <https://doi.org/10.1098/rstb.2005.1753>.
- Verkade, J. S., J. D. Brown, P. Reggiani, and A. H. Weerts, 2013: Post-processing ECMWF precipitation and temperature

- ensemble reforecasts for operational hydrologic forecasting at various spatial scales. *J. Hydrol.*, **501**, 73–91, <https://doi.org/10.1016/j.jhydrol.2013.07.039>.
- Vrac, M., 2018: Multivariate bias adjustment of high-dimensional climate simulations: The rank resampling for distributions and dependences (R^2D^2) bias correction. *Hydrol. Earth Syst. Sci.*, **22**, 3175–3196, <https://doi.org/10.5194/hess-22-3175-2018>.
- Wang, Q. J., and D. E. Robertson, 2011: Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences. *Water Resour. Res.*, **47**, W02546, <https://doi.org/10.1029/2010WR009333>.
- Wilks, D. S., 2015: Multivariate ensemble model output statistics using empirical copulas. *Quart. J. Roy. Meteor. Soc.*, **141**, 945–952, <https://doi.org/10.1002/qj.2414>.
- Wood, A. W., E. P. Maurer, A. Kumar, and D. P. Lettenmaier, 2002: Long-range experimental hydrologic forecasting for the eastern United States. *J. Geophys. Res.*, **107**, 4429, <https://doi.org/10.1029/2001JD000659>.
- , L. R. Leung, V. Sridhar, and D. P. Lettenmaier, 2004: Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change*, **62**, 189–216, <https://doi.org/10.1023/B:CLIM.0000013685.99609.9e>.
- Zhao, T. T. G., J. C. Bennett, Q. J. Wang, A. Schepen, A. W. Wood, D. E. Robertson, and M. H. Ramos, 2017: How suitable is quantile mapping for postprocessing GCM precipitation forecasts? *J. Climate*, **30**, 3185–3196, <https://doi.org/10.1175/JCLI-D-16-0652.1>.