https://doi.org/10.5194/hess-2021-599
Preprint. Discussion started: 13 December 2021

Hydrology and Earth System Sciences
Discussions

Model Comparisons Between Canonical Vine Copulas and Meta-Gaussian for Agricultural Drought Forecasting over China

Authors: Haijiang Wu[1,2], Xiaoling Su[1,2*], Vijay P. Singh[3,4], Te Zhang[2], and Jixia Qi[2]

Affiliation:

[1]*Key Laboratory for Agricultural Soil and Water Engineering in Arid Area of Ministry of Education, Northwest A&F University, Yangling, Shaanxi, 712100, China*

[2]*College of Water Resources and Architectural Engineering, Northwest A&F University, Yangling, Shaanxi, 712100, China*

[3]*Department of Biological and Agricultural Engineering & Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, TX 77843-2117, USA*

[4]*National Water and Energy Center, UAE University, Al Ain, UAE*

*Corresponding Author:

Dr. Xiaoling Su, College of Water Resources and Architectural Engineering, Northwest A&F University, Weihui Road 23, Yangling, Shaanxi, China, *Email*: xiaolingsu@nwafu.edu.cn (X. Su).

https://doi.org/10.5194/hess-2021-599
Preprint. Discussion started: 13 December 2021

Hydrology and Earth System Sciences
Open Access
Discussions

Abstract

Agricultural drought is caused by reduced soil moisture and precipitation and affects the growth of crops and vegetation, and in turn agricultural production and food security. For developing measures for drought mitigation, reliable agricultural drought forecasting is essential. In this study, we developed an agricultural drought forecasting model based on canonical vine copulas under three-dimensions (3C-vine model), in which the antecedent meteorological drought and agricultural drought persistence were utilized as predictors. Besides, the meta-Gaussian (MG) model was selected as a reference model to evaluate the forecast skill. The agricultural drought in August of 2018 was selected as a case study, and the spatial patterns of 1–3-month lead forecasts of agricultural drought utilizing the 3C-vine model resembled the corresponding observations, indicating the predictive ability of the model. The performance metrics (NSE, R^2, and RMSE) showed that the 3C-vine model outperformed the MG model for August under diverse lead times. Also, the 3C-vine model exhibited excellent forecast skills in capturing the extreme agricultural drought over different selected typical regions. This study may help with drought early warning, drought mitigation, and water resources scheduling.

Keywords: agricultural drought forecasting, model comparison, vine copulas, meta-Gaussian

1. Introduction

Agriculture is the source of livelihoods of over 2.5 billion people worldwide, and the agricultural sector also sustains 82% of all drought impacts (FAO, 2021). A cascade of impacts of droughts, such as crop reduction and failure, increased human and tree mortality, and ecological disturbance, have attracted considerable attention (FAO, 2021; Lu et al., 2012; Modanesi et al., 2020; Su et al., 2018; Zhang et al., 2018; Zhang et al., 2019; Zscheischler et al., 2020). Droughts have

https://doi.org/10.5194/hess-2021-599
Preprint. Discussion started: 13 December 2021

reduced global crop production by about 9–10% for the period 1964–2007 (Lesk et al., 2016). Additionally, droughts have caused overall crop and livestock production loss of $37 billion over the least developed and lower-middle-income countries (FAO, 2021). Agricultural drought forecasting, therefore, lies at the core of overall drought risk management and is critical for food security, early warning, and drought mitigation.

Agricultural drought is generally referred to as soil moisture shortage, which affects crop yield and vegetation health (Modanesi et al., 2020; Zhang et al., 2016; Zhang et al., 2021). Under natural conditions, atmospheric precipitation is a paramount source for replenishment of soil moisture (Wu et al., 2021). Therefore, reduced soil moisture (agricultural drought) is mainly due to precipitation deficit (meteorological drought) (Modanesi et al., 2020; Orth & Destouni, 2018). Moreover, soil moisture has a good memory to drought because of the time-integration effects (Long et al., 2019), i.e., agricultural drought persistence. The previous meteorological drought and antecedent agricultural drought can be taken into consideration as predictors of subsequent agricultural drought.

In hydrology, the traditional methods have been extensively employed to forecast drought, such as regression models, machine learning models, and hybrid models (by considering both statistical and dynamical predictions) (Hao et al., 2016). Yet, these models tend to be limited in considering the complex nonlinear (e.g., regression models), explicit physical mechanisms and over-fitting (e.g., machine learning models), as well as the demand of massive hydroclimatic data input (e.g., hybrid models). The copula functions overcome the limitations of the conventional statistical methods. Since copulas can flexible joining arbitrary marginal distributions of variables, they have been widely employed in risk assessment (Hao et al., 2017; Liu et al., 2021; Sarhadi et al., 2016; Xu et al., 2021; Zhang et al., 2021; Zhou et al., 2019), flood and runoff forecasting (Bevacqua et al., 2017;

https://doi.org/10.5194/hess-2021-599
Preprint. Discussion started: 13 December 2021

Hemri et al., 2015; Liu et al., 2018; Zhang & Singh, 2019), and drought forecasting (Ganguli & Reddy, 2014; Wu et al., 2021). However, when bivariate copulas are extended to higher-dimensional (≥ three-dimensions) cases, they are restricted due to nonexistence of analytical expressions (Liu et al., 2021). Symmetric Archimedean copulas and nested Archimedean copulas partially have addressed the issues of dimensionality, but single parameter and Archimedean class are difficult to characterize the various dependence structures (Aas & Berg, 2009; Hao et al., 2016; Wu et al., 2021). Fortunately, the vine copulas addressed these limitations (Aas et al., 2009; Bedford & Cooke, 2002; Joe, 1996).

Vine copulas are flexible in decomposing any multi-dimensional joint distribution into a hierarchy of bivariate copulas or pair copula constructions (Aas et al., 2009; Bedford & Cooke, 2002; Liu et al., 2021; Vernieuwe et al., 2015; Xiong et al., 2014). These copulas have been extensively applied in the hydrological field (Bevacqua et al., 2017; Liu et al., 2021; Vernieuwe et al., 2015; Wu et al., 2021). For instance, Xiong et al. (2014) derived the annual runoff distributions using canonical vine copulas. Liu et al. (2018) developed a framework to investigate compound floods based on canonical vine copulas. Wang et al. (2019) utilized regular vine copulas with historical streamflow and climate drivers to simulate monthly streamflow for the headwater catchment of the Yellow River basin. Liu et al. (2021) developed a hybrid ensemble forecast model, using the Bayesian model averaging combined canonical vine copulas, to forecast water level. Wu et al. (2021) proposed an agricultural drought forecast model based on vine copulas under four-dimensional scenarios.

The meta-Gaussian (MG) model, a popular statistical model in the hydrometeorological community, is capable of joining multiple variables and have explicit conditional distributions, which is apt for forecasting and risk assessment purposes (Hao et al., 2016; Hao et al., 2019; Wu et

https://doi.org/10.5194/hess-2021-599
Preprint. Discussion started: 13 December 2021

al., 2021; Zhang et al., 2021). For example, the forecasting of compound dry-hot events in summer over Southern Africa was investigated, based on the MG model under 1-month and 3-month lead times (Hao et al., 2019). The propagation between meteorological drought and agricultural drought was characterized via the MG model (Xu et al., 2021). However, there has been a rather limited investigation, to our knowledge, that carrying out model comparisons between vine copulas and MG for agricultural drought forecasting under the same conditions. Therefore, the MG model was selected as a competition (or reference) model.

The objective of this study therefore was to compare the forecast ability of agricultural drought in August of every year in the period 1961–2018 between canonical vine copulas (i.e., 3C-vine model) and MG model under three-dimensional scenario. In the following, we briefly describe the study area and data used in Section 2. The MG and 3C-vine models and performance metrics utilized are presented in Section 3. Results of the 3C-vine model application and assessment are given in Section 4. Finally, the discussion and conclusions are presented in Section 5.

2. Study area and data used

China stretches across a vast area covering diverse climate regimes and is a major agricultural-producing country (Wu et al., 2021; Zhang et al., 2015). For the convenience of analyzing spatial patterns of agricultural drought, the climate of China was divided into seven sub-climate regions on the basis of Yao et al. (2018) and Zhao (1983), as shown in Figure 1. For each sub-climate region, the temperature and moisture conditions when combined are roughly similar, and the type of soil and vegetation have a certain common characteristic (Zhao, 1983).

---Figure 1. ---

https://doi.org/10.5194/hess-2021-599
Preprint. Discussion started: 13 December 2021

In this study, the gridded monthly precipitation with a 0.25°×0.25° spatial resolution was obtained from the CN05.1 dataset for the 1961–2018 period over the mainland of China (excluding the Taiwan province), which was provided by the China National Climate Center. The Copernicus Climate Change Service (C3S) at European Center for Medium-Range Weather Forecast (ECMWF) has begun the release of the ERA5 back extension data covering the period 1950–1978 on the Climate Data Store (CDS). Therefore, the gridded monthly soil moisture with a 0.25°×0.25° spatial resolution corresponding to three soil depths (0–7 cm, 7–28 cm, and 28–100 cm) are available from the ECMWF ERA5 reanalysis datasets for 1961–1978: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels-monthly-means-preliminary-back-extension?tab=overview and 1979–2018: https://cds.climate.copernicus.eu./cdsapp#!/dataset/reanalysis-era5-single-levels-monthly-means?tab=overview. The CN05.1 and ERA5 reanalysis datasets have been extensively utilized numerous studies, e.g., drought monitoring and forecasting (Wu et al., 2021; Zhang et al., 2021), long-term climatic analysis (He et al., 2021; Wu et al., 2017), and flash drought attribution analysis (Wang & Yuan, 2021).

3. Methodology

We employed the Standardized Precipitation Index (SPI, based on monthly precipitation) and Standardized Soil moisture Index (SSI, based on monthly cumulative soil moisture at three soil depths), respectively, to characterize meteorological drought and agricultural drought at a 6-month timescale. The empirical Gringorten plotting position formula (Gringorten, 1963) was used to obtain the empirical cumulative probabilities of these two indexes, which were then transformed into standardized variables via the normal quantile transformation. Since meteorological drought is a source of other drought types (e.g., agricultural drought), the antecedent precipitation deficiency (i.e., meteorological drought) has a stronger effect on the subsequent soil moisture deficiency (i.e., agricultural drought). Moreover, soil moisture has a good memory for prior drought, i.e., agricultural drought persistence, which is attributed to the soil porosity characteristics and time-integration effects (Long et al., 2019; Wu et al., 2021). We attempted to use the prior meteorological drought (SPI_{t-i}; t denotes the target month (e.g., August), and i indicates lead time (month)) and agricultural drought persistence (SSI_{t-i}) to forecast the subsequent agricultural drought (SSI_t) based on the canonical vine copulas under three-dimensional scenarios (3C-vine model). We selected the meta-Gaussian (MG) model as a reference model to assess the agricultural drought forecast performance of the 3C-vine model. More detailed information is given below.

3.1. Meta-Gaussian model under three-dimensional scenarios

The meta-Gaussian (MG) model can effectively combine multiple hydrometeorological variables, which have gained attention for drought forecasting and risk assessment (Hao et al., 2019; Hao et al., 2019; Wu et al., 2021; Zhang et al., 2021). Suppose the series of SPI_{t-i}, SSI_{t-i}, and SSI_t correspond to random variables Y_1, Y_2, and Y_3, respectively, the predictand y_3 under the given conditions of y_1 and y_2 based on the MG model can be expressed as (Wilks, 2014):

$$y_3 \mid (y_1, y_2) \sim N(\mu_{y_3|(y_1, y_2)}, \Sigma_{y_3|(y_1, y_2)}) \tag{1}$$

where N signifies the Gaussian distribution function, $\mu_{y_3|(y_1, y_2)}$ denotes the conditional mean, and $\Sigma_{y_3|(y_1, y_2)}$ represents the conditional covariate matrix.

Furthermore, we removed the forecast values in a specific year of y_1, y_2, and y_3, which denote y_1^{-yr}, y_2^{-yr}, and y_3^{-yr}, respectively. Under this circumstance, the covariate matrix $\mathbf{\Sigma}$ regarding y_1^{-yr},

y_2^{-yr}, and y_3^{-yr} can be written as:

$$\mathbf{\Sigma} = Cov\left[\begin{array}{cc|c} (y_1^{-yr}, y_1^{-yr}) & (y_1^{-yr}, y_2^{-yr}) & (y_1^{-yr}, y_3^{-yr}) \\ (y_2^{-yr}, y_1^{-yr}) & (y_2^{-yr}, y_2^{-yr}) & (y_2^{-yr}, y_3^{-yr}) \\ \hline (y_3^{-yr}, y_1^{-yr}) & (y_3^{-yr}, y_2^{-yr}) & (y_3^{-yr}, y_3^{-yr}) \end{array}\right] = \left[\begin{array}{cc|c} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ \hline C_{31} & C_{32} & C_{33} \end{array}\right] = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix} \quad (2)$$

The forecast of specific years, i.e., y_3^{yr}, can be derived as (Wilks, 2014):

$$y_3^{yr} = \mu_{y_3|(y_1, y_2)} = \mu_{y_3^{-yr}} + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\begin{bmatrix} y_1^{yr} - \mu_{y_1^{-yr}} \\ y_2^{yr} - \mu_{y_2^{-yr}} \end{bmatrix} \quad (3)$$

where $\mu_{y_1^{-yr}}$, $\mu_{y_2^{-yr}}$, and $\mu_{y_3^{-yr}}$ represent the mean of y_1^{-yr}, y_2^{-yr}, and y_3^{-yr}, respectively. y_1^{yr} and y_2^{yr} denote that y_1 and y_2 provided the forecast information at time t–i in a specific year.

3.2. Canonical vine copulas model under three-dimensional scenarios

Copulas can effectively combine multiple variables without the restriction of marginal distributions (Nelsen, 2013; Sarhadi et al., 2016; Wang et al., 2019; Xiong et al., 2014). They were initially utilized for deriving joint distributions of two-dimensional variables, since parameters are easy to assess and the analytical solution is apt to obtain (Liu et al., 2021; Sadegh et al., 2017). However, under higher-dimensional (e.g., $d \geqslant 3$) scenarios, owing to the limitations of a great deal of parameters and complexity, the copulas (mainly referred to bivariate copulas) are difficult to promote and apply (Joe, 2014; Liu et al., 2018; Liu et al., 2021; Sadegh et al., 2017). To overcome these limitations, Joe (1996) and Aas et al. (2009) developed vine copulas, a hierarchy of pair copula constructions, for multi-dimensional cases. Vine copulas possess two sub-classes: canonical vine copulas (C-vine copulas) and drawable vine copulas (D-vine copulas). Here, we mainly employed the C-vine copulas to establish the forecast model of agricultural drought under three-dimensional conditions. Undoubtedly, a similar scheme is capable of applying to D-vine copulas.

C-vine copulas may have numerous tree structures, especially for the case of higher dimensions, which are associated with the quantity and ordering of variables (Aas et al., 2009; Liu et al., 2018; Liu et al., 2021; Wu et al., 2021). Also, different ordering of variables affects the estimation of the parameters of C-vine copulas (Liu et al., 2021; Wang et al., 2019). Given the ordering of variables Y_1, Y_2, and Y_3 for three-dimensional C-vine copula model (termed as 3C-vine model hereinafter; Figure 2a), the joint probability density function (PDF), g_{123}, can be expressed as (Aas et al., 2009):

$$g_{123} = g_1 \bullet g_2 \bullet g_3 \bullet c_{12} \bullet c_{13} \bullet c_{23|1} \tag{4}$$

where g_1, g_2, and g_3 correspond to the margin density functions of $g_1(y_1)$, $g_2(y_2)$, and $g_3(y_3)$, respectively; c is the bivariate copula density; c_{12}, c_{13}, and $c_{23|1}$ signify the abbreviation of $c_{1,2}[G_1(y_1), G_2(y_2)]$, $c_{1,3}[G_1(y_1), G_3(y_3)]$, and $c_{2,3|1}[G(y_2|y_1), G(y_3|y_1)]$, respectively. The selected bivariate copulas utilized in this study comprised Gaussian (or Normal), Student-t, Clayton, and Frank, as well as the corresponding survival functions. We used the R function *CDVineCondFit* in the "*CDVineCopulaConditional*" R package (Bevacqua, 2017), based on the Akaike information criterion (AIC), to select the suitable bivariate copula for each pair of variables.

--Figure 2. --

A conditional copula density needs to be addressed in Equation 4, i.e., $G(y|\mathbf{w})$, where $\mathbf{w}$ is a d-dimensional vector $\mathbf{w} = (w_1, \ldots, w_d)$. Here, regarding the conditional distribution of z given the conditions $\mathbf{w}$, we introduced the h-function, $h(y, \mathbf{w}; \theta)$, to indicate the $G(y|\mathbf{w})$ as follows (Aas et al., 2009; Joe, 1996):

$$h(y,\mathbf{w};\theta) := G(y\,|\,\mathbf{w}) = \frac{\partial C_{y,w_j|\mathbf{w}_{-j}}\left[G(y\,|\,\mathbf{w}_{-j}), G(w_j\,|\,\mathbf{w}_{-j})\right]}{\partial G(w_j\,|\,\mathbf{w}_{-j})} \tag{5}$$

where θ denotes the parameter(s) of bivariate copula function $C_{yw_j|\mathbf{w}_{-j}}$; w_j represents an arbitrary

https://doi.org/10.5194/hess-2021-599
Preprint. Discussion started: 13 December 2021

Hydrology and Earth System Sciences
Discussions
Open Access

component of **w**; and w_{-j} indicates the excluding element w_j from the vector **w**.

Let the ordering variables be y_1, y_2, and y_3, the conditional variables be y_1 and y_2, and the predictand be y_3. Accordingly, the expression of $G(y_3|y_1, y_2)$, based on Equation 5, can be written as:

$$G(y_3 \mid y_1, y_2) = \frac{\partial C_{z_3, z_1|z_2}\left[G(y_3 \mid y_1), G(y_2 \mid y_1)\right]}{\partial G(y_2 \mid y_1)} = h\left\{h(u_3 \left| u_1; \theta_{12}) \right| h(u_2 \left| u_1; \theta_{11}); \theta_{21}\right.\right\} \tag{6}$$

where θ_{ij} (i denotes a tree and j is an edge) represents the parameters of different conditional copulas in the 3C-vine model (Figure 2a); and u_k(k = 1, …, 3) is the marginal cumulative distribution function (CDF) of y_k. The CDF for each variable is substituted by the empirical Gringorten cumulative probability (Bevacqua et al., 2017; Genest et al., 2009; Wu et al., 2021).

Here, we introduced the τ-th copula–quantile curve (Chen et al., 2009; Liu et al., 2018) to simulate u_3 based on Equation 6 and derived its inverse distribution function as follows:

$$y_3 = N^{-1}\left\{G(\tau \mid z_1, z_2)\right\} = N^{-1}(u_3) = N^{-1}\left[h^{-1}\left\{h^{-1}(\tau \left| h(u_2 \left| u_1; \theta_{11}); \theta_{21}) \right| u_1; \theta_{12}\right.\right\}\right] \tag{7}$$

where N^{-1} and h^{-1} signify the inverse form of Gaussian distribution and h-function, respectively; y_3 is the agricultural drought forecast at time t (i.e., SSI_t); y_1 and y_2 are the predictors corresponding to the antecedent meteorological drought and agricultural drought persistence at time $t–i$ (i.e., $SPI_{t–i}$ and $SSI_{t–i}$). The R functions of *BiCopHfunc* and *BiCopHinv* in the R package "*VineCopula*" (Nagler et al., 2021) were utilized to model the h-function and its inverse form for Equation 7, respectively.

The tree structure is related to the ordering variables, so when the ordering variables are y_2, y_1, and y_3 (conditional variables are y_1 and y_2; Figure 2b), Equations 6 and 7 can be changed analogously as:

$$G(y_3 \left| y_2, y_1\right.) = h\left\{h(u_3 \left| u_2; \theta_{12}) \right| h(u_1 \left| u_2; \theta_{11}); \theta_{21}\right.\right\} \tag{8}$$

$$y_3 = N^{-1}(u_3) = N^{-1}\left[h^{-1}\left\{h^{-1}(\tau \left| h(u_1 \left| u_2;\theta_{11}\right.);\theta_{21}\right.) \left| u_2;\theta_{12}\right.\right\}\right] \tag{9}$$

We first generated a sample size of 1,000 uniformly distributed random values over the interval [0, 1] by Monte Carlo simulation. Then, the best 3C-vine model (i.e., selected the best model from Equation 7 and Equation 9 according to AIC) was utilized to obtain 1,000 simulations (or estimations) for y_3. The best forecast of y_3 was finally calculated by the mean value of these simulations. Note that we applied the leave-one-out cross validation (LOOCV) (Wilks, 2014) to forecast agricultural drought in August of every year during 1961–2018 for the 3C-vine model or MG model, namely, the validation sample was left one in each time, and the rest were used to establish the 3C-vine model or MG model and obtain the corresponding parameters.

3.3. Performance metrics

The Nash-Sutcliffe efficiency (NSE), coefficient of determination (R^2), and root mean square error (RMSE) were utilized to assess the forecast performance of 3C-vine model or MG model. These metrics can be expressed as:

$$NSE = 1 - \frac{\sum_{i=1}^{n}(AP_i - AO_i)^2}{\sum_{i=1}^{n}(AO_i - \overline{AO})^2} \qquad NSE \in (-\infty, 1] \tag{10}$$

$$R^2 = \frac{\left[\sum_{i=1}^{n}(AO_i - \overline{AO})^2 (AP_i - \overline{AP})\right]^2}{\sum_{i=1}^{n}(AO_i - \overline{AO})^2 \bullet \sum_{i=1}^{n}(AP_i - \overline{AP})^2} \qquad R^2 \in [0,1] \tag{11}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(AP_i - AO_i)^2} \qquad RMSE \in [0, +\infty) \tag{12}$$

where n is the number of forecast periods; AO_i and AP_i are the i-th observed and forecasted

https://doi.org/10.5194/hess-2021-599
Preprint. Discussion started: 13 December 2021

agricultural droughts (i.e., SSI), respectively; $\overline{AO}$ and $\overline{AP}$ denote the mean of the SSI observations and forecasts in the target month (e.g., August), respectively. Moreover, a most positive NSE and R^2 value and a lower RMSE value expressed a good forecast performance for the 3C-vine model or MG model.

4. Results

4.1. Correlation patterns of agricultural drought with potential predictors

The dependence between variables can be measured by the correlation coefficient, indirectly characterizing the quantity of common information between the two variables. In this study, we employed Kendall's correlation coefficient (τ_k) to measure the dependence of agricultural drought at current time t (SSI_t, herein t is August) with the previous meteorological drought (SPI_{t-i}, i indicates the lag or lead time with 1–3-month herein) and agricultural drought persistence (SSI_{t-i}). It should be mentioned that the significant correlation prevalent used may overestimate or overinterpret the dependence between variables (Wilks, 2016). Therefore, we adopted the maximum false discovery rate (FDR) of 0.1 to correct τ_k at the 0.05 significance level (Benjamini & Hochberg, 1995; Röthlisberger & Martius, 2019; Wilks, 2016).

--Figure 3. ---

Figure 3 summarizes 1–3-month lag τ_k between antecedent SPI (SSI) and succedent SSI for August during 1961–2018 over China. For most regions of China under 1–3-month lag time, the previous meteorological drought or agricultural drought persistence (memory) showed significant positive correlations with the target agricultural drought (i.e., the stippling in Figure 3). Also, we found perfect agricultural drought memory over many regions of China (excluding D4, a humid

https://doi.org/10.5194/hess-2021-599
Preprint. Discussion started: 13 December 2021

climate region) (Figures 3e and 3f), as the overlapping information existed in SSI_t and SSI_{t-i}. Additionally, the dependency pattern varied temporally and spatially, and this phenomenon evidently occurred with the lag (or lead) time extended, especially between SPI_{t-i} and SSI_t (Figure 3a–3c). Overall, the prior meteorological drought and agricultural drought memory provided reliable and useful forecast information for the subsequent agricultural drought for most areas of China.

4.2. Forecast performance comparison between 3C-vine model and MG Model

We leveraged the MG model as a reference model to measure the performance of 3C-vine model in forecasting the agricultural drought for the period 1961–2018 over China. Figures 4a–4i show the difference between the 3C-vine model and MG model with respect to NSE_{3C-MG}, R^2_{3C-MG}, and $RMSE_{3C-MG}$ under 1–3-month leads for August, respectively. In terms of the spatial extent of $NSE_{3C-MG} > 0$, $R^2_{3C-MG} > 0$, and $RMSE_{3C-MG} < 0$, the agricultural drought forecast ability of 3C-vine model superior MG model was occupied 65%, 68%, and 58% of land areas in China, respectively, under the 1-month lead SSI forecast (Figures 4a, 4d, and 4g), except for western China (D3 and D7) and central China (D4). The relationship between predictors and the forecasted variable was simple under 1–month lead time, so the MG model better showed their connection. However, with the lead time prolonged, the forecast skills of 3C-vine model outperformed the MG model for most regions of China (e.g., Figures 4e and 4f, accounting 72% and 74% of land areas in China for $R^2_{3C-MG} > 0$ under 2–3-month lead times, respectively). This indicates the 3C-vine model sufficiently utilized the forecasted information contained by previous meteorological drought and agricultural drought persistence in comparison with the MG model under the same conditions.

It can be seen that the forecast ability of 3C-vine model, compared with the MG model, is limited over climate region D5 (e.g., Figures 4b and 4c). This may be related to the fact that D5 is a

https://doi.org/10.5194/hess-2021-599
Preprint. Discussion started: 13 December 2021

crucial grain-producing region in China (Lu et al., 2012; Xiao et al., 2019; Zhang et al., 2016), the intensive anthropogenic activities (e.g., irrigation and urbanization) may alter the linkage between meteorological drought and agricultural drought, as well as the strength of agricultural drought memory (AghaKouchak et al., 2021). To ensure food security, if D5 experiences a drought event at the previous stage, agricultural managers and policymakers would mitigate the drought through irrigation in a variety of ways, such as groundwater exploitation and reservoir operation (Zhang et al., 2016). However, under this circumstance, the soil water obtaining the supplement from the irrigation water would affect the performance of agricultural drought forecast.

---Figure 4. --

In contrast with the MG model, the 3C-vine model yielded a better forecast performance for August under 1–3-month leads agricultural drought across most areas of China, except for the climate region D5.

4.3. Case study and sub-climate region assessment

The severe drought hit most regions of China in summer 2018, especially in southern and northern China, as the western North Pacific subtropical high abnormally impacted (Liu & Zhu, 2019; Zhang et al., 2020; Zhang et al., 2018). We chose the agricultural drought that occurred in August of 2018 as a case study to investigate the forecast ability of 3C-vine model. Similarly, the MG model was selected as a benchmark model. Figure 5 presents the SSI observations and 1–3-month lead SSI forecasts for this agricultural drought using the 3C-vine model and MG model. Obviously, the 1–3-month lead SSI forecasts via 3C-vine model resembled the observations (Figures 5a–5d), which captured the droughts that emerged in southern China, northern China, and northeastern China, i.e., climate regions D1–D2 and D4–D6. Comparing the 3C-vine model with

https://doi.org/10.5194/hess-2021-599
Preprint. Discussion started: 13 December 2021

the MG model under 2–3-month leads (Figures 5c–5d versus Figures 5f–5g), we observed the deteriorating forecast skill of MG model in climate region D5, which tended to non-drought state (i.e., SSI > 0), but the 3C-vine model better forecasted the agricultural drought for these regions under the same conditions, although the severity of agricultural drought had some decrement. The above analyses indicated that the 3C-vine model, using previous meteorological drought and agricultural drought persistence as two predictors, had the ability for reliable drought forecast over many regions of China.

---Figure 5. ---

---Figure 6. ---

Furthermore, to explore the skill of 3C-vine model in capturing the extremum of agricultural drought (i.e., minimum and maximum SSI), we randomly selected a typical region (black rectangle boxes in Figure 5b) in each climate region. Note that these extreme SSI values were calculated using the spatial average in each typical region. Figure 6 shows the probability density function (PDF) curve of minimum and maximum SSIs for these selected typical regions (D1S–D7S) via the 3C-vine model for 1–3-month leads of August. Here, the vertical black dash line denotes the SSI observation in each subplot. The *x*-axis value of peak point (i.e., high probability) for each PDF curve is regarded as the best estimation of SSI under diverse lead times. For minimum SSI with 1–2-month lead times, the difference between forecasted SSI and observed SSI was slight (except for D3S), which all reflected the drought state for these typical regions (Figure 6a). The deteriorated skills of 3C-vine model in a typical region D3S may be attributed to the lengthy response time existing between precipitation deficiency and soil moisture shortage, which is caused by the limited precipitation that cannot effectively replenish the soil moisture depletion due to the incrassation of vadose zone. For the 3-month lead time, the poor forecasts were produced in a typical region D5S for the minimum SSI. This phenomenon may result in the agricultural manager utilizing irrigation to mitigate the effect of drought on crop growth, thus, the response relationship between meteorological drought and agricultural drought accordingly would change (Xu et al., 2021).

For the forecasted maximum SSI utilizing 3C-vine model over diverse regions, the excellence forecast ability is displayed for the 1–3-month leads (Figure 6b), excluding the typical regions D5S and D6S (PDF curve shifted left). For the abundant precipitation and higher soil moisture content in D6S, the shortened response time between precipitation and soil moisture (Xu et al., 2021) may cause inferior forecasts of 3C-vine model for the target month.

Moreover, to assess the forecast performance (according to NSE, R^2, and RMSE) of the 3C-vine model over each climate region, we counted the pixel contained in each climate region and constructed the boxplots for these performance metrics (Figures 4j–4l). We still selected the MG model as the reference model, and obtained the difference between these two models, i.e., $NSE_{3C–MG}$, $R^2_{3C–MG}$, and $RMSE_{3C–MG}$. The forecast performances of 3C-vine model and MG model were generally consistent for 1-month lead of August over climate regions D1–D7 (Figures 4j–4l, the median percentile of $NSE_{3C–MG}$, $R^2_{3C–MG}$, and $RMSE_{3C–MG}$ were all around the 0 line), indicating the improved skills of 3C-vine model was limited under the same condition. Obviously, the median percentile of $NSE_{3C–MG}$ and $R^2_{3C–MG}$ were greater than 0 as well as $RMSE_{3C–MG}$ was lower than 0, respectively, for 2–3-month leads SSI forecast of August in different climate regions D1–D7 (except for D5), indicating that the 3C-vine model more accurately forecasted agricultural drought than did the MG model in diverse climate regions.

In conclusion, based the ability of typical agricultural drought forecasted (Figure 5) and

https://doi.org/10.5194/hess-2021-599
Preprint. Discussion started: 13 December 2021

Hydrology and Earth System Sciences
Discussions

extremum agricultural drought captured in selected typical regions (Figure 6) and the comprehensive forecast performance showed in diverse climate regions (Figures 4j–4l), the 3C-vine model had a good forecast skill for 1–3-month leads agricultural drought of August over most areas of China.

5. Discussion and Conclusions

This study developed a C-vine copula model for forecasting agricultural drought over China under three dimensions, in which antecedent meteorological drought and agricultural drought persistence at time t–1 (t denotes target month) was primarily employed as two predictors. We selected the MG model as a competition model, in terms of the difference in NSE, R^2, and RMSE between 3C-vine and MG models, to evaluate the forecast performance of 3C-vine model. These performance metrics all displayed that the 3C-vine model, especially for 2–3-month lead times, outperformed the MG model in many climate regions over China (except for D5, which lies in humid and subhumid regions of northern China) (Figure 4). Compared with the MG model, the 3C-vine model yielded a good forecast skill for the selected typical agricultural droughts (Figure 5). Besides, the nearly perfect forecast of extremum agricultural drought in typical regions (Figure 6) further certified the excellent ability of 3C-vine model.

Heterogeneous topography and anthropogenic activities (e.g., irrigation and urbanization) have certainly impacted precipitation interpolation and soil moisture simulation, which may depart from the actual precipitation or soil moisture conditions, notwithstanding the precipitation of CN05.1 and soil moisture of ERA5 that show good performances with respect to drought monitoring and forecasting over China (Wang & Yuan, 2021; Wu et al., 2021; Xu et al., 2009; Zhang et al., 2021; Zhang et al., 2019). It can also influence the response (propagation) time between meteorological drought and agricultural drought as well as agricultural drought memory and can thus lead to the

3C-vine model falling short in some climate regions. To address this issue, we can comprehensively utilize multiple reanalysis data sets, e.g., the precipitation and soil moisture data in Global Land Data Assimilation System (GLDAS) and ERA5, to reduce the uncertainty resulting from a single data source (Wang & Yuan, 2021; Wu et al., 2021). Currently, it is a challenge to consider irrigation activities into agricultural drought forecasting, especially at large spatial scales. In addition to antecedent precipitation deficit, air temperature, relative humidity, and evapotranspiration may influence soil moisture budget. Moreover, from the perspective of driving mechanisms, the effect of certain atmospheric circulation anomalies (e.g., El Niño-Southern Oscillation (ENSO), Pacific Decadal Oscillation (PDO), and North Arctic Oscillation (NAO)) on agricultural drought at regional and global scales can also be considered as predictors (Zhang et al., 2021). Therefore, a more efficient space can be established by leveraging these predictors for agricultural drought forecasting.

In recent years, a myriad of extreme events, such as heatwaves and flash droughts, have swept many regions around the globe. These extreme events have a rapid onset with a few days or weeks and lead to devastating impacts on agricultural production, water resource security, and human well-being (Wang & Yuan, 2021; Yuan et al., 2019; Zscheischler et al., 2020). Therefore, agricultural drought forecasting at finer temporal scales (e.g., weekly) is essential for agricultural managers and policymakers to manage and plan water use. Yet, with limited spatiotemporal resolution and the length of model sample, we temporally have not carried out agricultural drought forecasting at sub-monthly or pentad temporal scales.

The limitation of this study is that we choose a single "best" model from two C-vine copula candidate models (i.e., Figure 2) as the ideal forecast. However, as the inherent structural differences (i.e., ordering variables are different), the utilized best model may underestimate the forecast uncertainty (Liu et al., 2021). Therefore, to reduce the predictive uncertainty and improve the forecast performance, a multi-model combination technique (e.g., Bayesian model averaging (Liu et al., 2021; Long et al., 2017)) can be considered to merge different C-vine copula candidate models. Moreover, as we only pay attention to the C-vine copulas and several bivariate copula functions, the other D-vine copulas or regular vine copulas, as well as a multitude of bivariate copula families (Sadegh et al., 2017) can be investigated to establish the forecast model for agricultural drought in the next work.

Data availability

The grided monthly precipitation data with a 0.25° spatial resolution was provided by the CN05.1 (http://data.cma.cn) for the period of 1961–2018. The gridded monthly soil moisture data with three soil depths (0–7 cm, 7–28 cm, and 28–100 cm) from the European Center for Medium-Range Weather Forecast (ECMWF) ERA5 reanalysis datasets are available at 1961–1978: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels-monthly-means-preliminary-back-extension?tab=overview and 1979–2018: https://cds.climate.copernicus.eu./cdsapp#!/dataset/reanalysis-era5-single-levels-monthly-means?tab=overview.

Author contribution

Haijiang Wu: Conceptualization, Methodology, Software, Visualization, Writing - original draft. Xiaoling Su: Data curation, Validation, Investigation, Funding acquisition, Supervision, Formal analysis. Vijay P. Singh: Writing - review & editing, Supervision. Te Zhang: Formal analysis, Investigation. Jixia Qi: Data curation, Investigation.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

This study was financially supported by the National Natural Science Foundation of China (Grants No. 51879222 and 52079111).

References

Aas, K. & Berg, D. (2009). Models for construction of multivariate dependence – a comparison study. *The European Journal of Finance*, 15(7-8), 639-659. https://doi.org/10.1080/13518470802588767

Aas, K., Czado, C., Frigessi, A. & Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2), 182-198. https://doi.org/10.1016/j.insmatheco.2007.02.001

AghaKouchak, A., Mirchi, A., Madani, K., Di Baldassarre, G., Nazemi, A., Alborzi, A., et al. (2021). Anthropogenic Drought: Definition, Challenges, and Opportunities. *Reviews of Geophysics*, 59(2), e2019RG000683. https://doi.org/10.1029/2019rg000683

Bedford, T. & Cooke, R. M. (2002). Vines–A new graphical model for dependent random variables. *Annals of Statistics* 30(4), 1031-1068.

Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, 57(1), 289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bevacqua, E. (2017). CDVineCopulaConditional: Sampling from conditional C- and D-vine copulas,

https://doi.org/10.5194/hess-2021-599
Preprint. Discussion started: 13 December 2021

R package. version 0.1.1. https://CRAN.R-project.org/package=CDVineCopulaConditional

Bevacqua, E., Maraun, D., Hobæk Haff, I., Widmann, M. & Vrac, M. (2017). Multivariate statistical modelling of compound events via pair-copula constructions: analysis of floods in Ravenna (Italy). *Hydrology and Earth System Sciences*, 21(6), 2701-2723. https://doi.org/10.5194/hess-21-2701-2017

Chen, X., Koenker, R. & Xiao, Z. (2009). Copula-based nonlinear quantile autoregression. *Econometrics Journal*, 12, S50-S67. https://doi.org/10.1111/j.1368-423X.2008.00274.x

FAO. (2021). The impact of disasters and crises on agriculture and food security. *Food and Agriculture Organization of the United Nations, Rome*. https://doi.org/10.4060/cb3673en

Ganguli, P. & Reddy, M. J. (2014). Ensemble prediction of regional droughts using climate inputs and the SVM-copula approach. *Hydrological Processes*, 28(19), 4989-5009. https://doi.org/10.1002/hyp.9966

Genest, C., Rémillard, B. & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44(2), 199-213. https://doi.org/10.1016/j.insmatheco.2007.10.005

Gringorten, I. I. (1963). A plotting rule for extreme probability paper. *Journal of Geophysical Research*, 68(3), 813-814. https://doi.org/10.1029/JZ068i003p00813

Hao, Z., Hao, F., Singh, V. P., Sun, A. Y. & Xia, Y. (2016). Probabilistic prediction of hydrologic drought using a conditional probability approach based on the meta-Gaussian model. *Journal of Hydrology*, 542, 772-780. https://doi.org/10.1016/j.jhydrol.2016.09.048

Hao, Z., Hao, F., Singh, V. P. & Ouyang W. (2017). Quantitative risk assessment of the effects of drought on extreme temperature in eastern China. *Journal of Geophysical Research: Atmosphere*, 122, 9050-9059, https://doi.org/10.1002/2017JD027030

https://doi.org/10.5194/hess-2021-599
Preprint. Discussion started: 13 December 2021

Hao, Z., Hao, F., Singh, V. P. & Zhang, X. (2019). Statistical prediction of the severity of compound dry-hot events based on El Niño-Southern Oscillation. *Journal of Hydrology*, 572, 243-250. https://doi.org/10.1016/j.jhydrol.2019.03.001

Hao, Z., Hao, F., Xia, Y., Singh, V. P. & Zhang, X. (2019). A monitoring and prediction system for compound dry and hot events. *Environmental Research Letters*, 14(11), 114034. https://doi.org/10.1088/1748-9326/ab4df5

He, L., Hao, X., Li, H. & Han, T. (2021). How Do Extreme Summer Precipitation Events Over Eastern China Subregions Change? *Geophysical Research Letters*, 48, e2020GL091849. https://doi.org/10.1029/2020GL091849

Hemri, S., Lisniak, D. & Klein, B. (2015). Multivariate postprocessing techniques for probabilistic hydrological forecasting. *Water Resources Research*, 51(9), 7436-7451. https://doi.org/10.1002/2014wr016473

Joe, H. (1996). Families of m-variate distributions with given margins and m(m−1)/2 bivariate dependence parameters. Distributions with fixed marginals and related topics. *Institute of Mathematical Statistics Lecture Notes—Monograph Series*, 28, 120-141. https://doi.org/10.1214/lnms/1215452614

Joe, H. (2014). Dependence modeling with copulas. Chapman and Hall/CRC.

Lesk, C., Rowhani, P. & Ramankutty, N. (2016). Influence of extreme weather disasters on global crop production. *Nature*, 529(7584), 84-87. https://doi.org/10.1038/nature16467

Liu, B. & Zhu, C. (2019). Extremely Late Onset of the 2018 South China Sea Summer Monsoon Following a La Niña Event: Effects of Triple SST Anomaly Mode in the North Atlantic and a Weaker Mongolian Cyclone. *Geophysical Research Letters*, 46(5), 2956-2963. https://doi.org/10.1029/2018gl081718

Liu, Z., Cheng, L., Hao, Z., Li, J., Thorstensen, A. & Gao, H. (2018). A Framework for Exploring Joint Effects of Conditional Factors on Compound Floods. *Water Resources Research*, 54(4), 2681-2696. https://doi.org/10.1002/2017wr021662

Liu, Z., Cheng, L., Lin, K. & Cai, H. (2021). A hybrid bayesian vine model for water level prediction. *Environmental Modelling & Software*, 142, 105075. https://doi.org/10.1016/j.envsoft.2021.105075

Liu, Z., Xie, Y., Cheng, L., Lin, K., Tu, X. & Chen, X. (2021). Stability of spatial dependence structure of extreme precipitation and the concurrent risk over a nested basin. *Journal of Hydrology*, 602, 126766. https://doi.org/10.1016/j.jhydrol.2021.126766

Long, D., Bai, L., Yan, L., Zhang, C., Yang, W., Lei, H., et al. (2019). Generation of spatially complete and daily continuous surface soil moisture of high spatial resolution. *Remote Sensing of Environment*, 233, 111364. https://doi.org/10.1016/j.rse.2019.111364

Long, D., Pan, Y., Zhou, J., Chen, Y., Hou, X., Hong, Y., et al. (2017). Global analysis of spatiotemporal variability in merged total water storage changes using multiple GRACE products and global hydrological models. *Remote Sensing of Environment*, 192, 198-216. https://doi.org/10.1016/j.rse.2017.02.011

Lu, Y., Wu, K., Jiang, Y., Guo, Y. & Desneux, N. (2012). Widespread adoption of Bt cotton and insecticide decrease promotes biocontrol services. *Nature*, 487(7407), 362-365. https://doi.org/10.1038/nature11153

Modanesi, S., Massari, C., Camici, S., Brocca, L. & Amarnath, G. (2020). Do Satellite Surface Soil Moisture Observations Better Retain Information About Crop-Yield Variability in Drought Conditions? *Water Resources Research*, 56(2), e2019WR025855. https://doi.org/10.1029/2019wr025855

Nagler, T., Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., Erhardt, T., et al. (2021). VineCopula: Statistical Inference of Vine Copulas, R Package Version 2.4.2. https://CRAN.R-project.org/package=VineCopula

Nelsen, R. B. (2013). An Introduction to Copulas. 2nd ed., Springer, N. Y.

Orth, R. & Destouni, G. (2018). Drought reduces blue-water fluxes more strongly than green-water fluxes in Europe. *Nature Communications*, 9(1), 3602. https://doi.org/10.1038/s41467-018-06013-7

Röthlisberger, M. & Martius, O. (2019). Quantifying the Local Effect of Northern Hemisphere Atmospheric Blocks on the Persistence of Summer Hot and Dry Spells. *Geophysical Research Letters*, 46(16), 10101-10111. https://doi.org/10.1029/2019gl083745

Sadegh, M., Ragno, E. & AghaKouchak, A. (2017). Multivariate Copula Analysis Toolbox (MvCAT): Describing dependence and underlying uncertainty using a Bayesian framework. *Water Resources Research*, 53(6), 5166-5183. https://doi.org/10.1002/2016wr020242

Sarhadi, A., Burn, D. H., Concepción Ausín, M. & Wiper, M. P. (2016). Time-varying nonstationary multivariate risk analysis using a dynamic Bayesian copula. *Water Resources Research*, 52(3), 2327-2349. https://doi.org/10.1002/2015wr018525

Su, B., Huang, J., Fischer, T., Wang, Y., Kundzewicz, Z. W., Zhai, J., et al. (2018). Drought losses in China might double between the 1.5 degrees C and 2.0 degrees C warming. *Proceedings of the National Academy of Sciences of the United States of America*, 115(42), 10600-10605. https://doi.org/10.1073/pnas.1802129115

Vernieuwe, H., Vandenberghe, S., De Baets, B. & Verhoest, N. E. C. (2015). A continuous rainfall model based on vine copulas. *Hydrology and Earth System Sciences*, 19(6), 2685-2699. https://doi.org/10.5194/hess-19-2685-2015

https://doi.org/10.5194/hess-2021-599
Preprint. Discussion started: 13 December 2021

Wang, W., Dong, Z., Lall, U., Dong, N. & Yang, M. (2019). Monthly Streamflow Simulation for the Headwater Catchment of the Yellow River Basin With a Hybrid Statistical-Dynamical Model. *Water Resources Research*, 55(9), 7606-7621. https://doi.org/10.1029/2019wr025103

Wang, Y. & Yuan, X. (2021). Anthropogenic Speeding Up of South China Flash Droughts as Exemplified by the 2019 Summer-Autumn Transition Season. *Geophysical Research Letters*, 48(9), e2020GL091901. https://doi.org/10.1029/2020gl091901

Wilks, D. S. (2014). Statistical methods in the atmospheric sciences. Academic Press.

Wilks, D. S. (2016). "The Stippling Shows Statistically Significant Grid Points": How Research Results are Routinely Overstated and Overinterpreted, and What to Do about It. *Bulletin of the American Meteorological Society*, 97(12), 2263-2273. https://doi.org/10.1175/bams-d-15-00267.1

Wu, H., Su, X., Singh, V. P., Feng, K. & Niu, J. (2021). Agricultural Drought Prediction Based on Conditional Distributions of Vine Copulas. *Water Resources Research*, 57(8), e2021WR029562. https://doi.org/10.1029/2021wr029562

Wu, H., Su, X. & Zhang, G. (2021). Prediction of agricultural drought in China based on Meta-Gaussian model. *Aata Geographica Sinaca*, 76(3), 525-538. https://doi.org/10.11821/dlxb202103003

Wu, J., Gao, X., Giorgi, F. & Chen, D. (2017). Changes of effective temperature and cold/hot days in late decades over China based on a high resolution gridded observation dataset. *International Journal of Climatology*, 37, 788-800. https://doi.org/10.1002/joc.5038

Xiao, G., Zhao, Z., Liang, L., Meng, F., Wu, W. & Guo, Y. (2019). Improving nitrogen and water use efficiency in a wheat-maize rotation system in the North China Plain using optimized farming practices. *Agricultural Water Management*, 212, 172-180.

https://doi.org/10.1016/j.agwat.2018.09.011

Xiong, L., Yu, K.-x. & Gottschalk, L. (2014). Estimation of the distribution of annual runoff from climatic variables using copulas. *Water Resources Research*, 50(9), 7134-7152. https://doi.org/10.1002/2013wr015159

Xu, Y., Gao, X., Shen, Y., Xu, C., Shi, Y. & Giorgi, F. (2009). A daily temperature dataset over China and its application in validating a RCM simulation. *Advances in Atmospheric Sciences*, 26(4), 763-772. https://doi.org/10.1007/s00376-009-9029-z

Xu, Y., Zhang, X., Hao, Z., Singh, V. P. & Hao, F. (2021). Characterization of agricultural drought propagation over China based on bivariate probabilistic quantification. *Journal of Hydrology*, 598, 126194. https://doi.org/10.1016/j.jhydrol.2021.126194

Yao, N., Li, Y., Lei, T. & Peng, L. (2018). Drought evolution, severity and trends in mainland China over 1961-2013. *Science of the Total Environment*, 616-617, 73-89. https://doi.org/10.1016/j.scitotenv.2017.10.327

Yuan, X., Wang, L., Wu, P., Ji, P., Sheffield, J. & Zhang, M. (2019). Anthropogenic shift towards higher risk of flash drought over China. *Nature Communications*, 10(1), 4661. https://doi.org/10.1038/s41467-019-12692-7

Zhang, J., Mu, Q. & Huang, J. (2016). Assessing the remotely sensed Drought Severity Index for agricultural drought monitoring and impact analysis in North China. *Ecological Indicators*, 63, 296-309. https://doi.org/10.1016/j.ecolind.2015.11.062

Zhang, L. & Singh, V. P. (2019). Copulas and their applications in water resources engineering. Cambridge University Press.

Zhang, L., Zhou, T., Chen, X., Wu, P., Christidis, N. & Lott, F. C. (2020). The late spring drought of 2018 in South China. *Bulletin of the American Meteorological Society*, 101(1), S59-S64.

https://doi.org/10.1175/BAMS-D-19-0202.1

Zhang, Q., Qi, T., Singh, V. P., Chen, Y. D. & Xiao, M. (2015). Regional Frequency Analysis of Droughts in China: A Multivariate Perspective. *Water Resources Management*, 29(6), 1767-1787. https://doi.org/10.1007/s11269-014-0910-x

Zhang, Q., Li, Q., Singh, V. P., Shi, P., Huang, Q. & Sun, P. (2018). Nonparametric integrated agrometeorological drought monitoring: Model development and application. *Journal of Geophysical Research: Atmospheres*, 123, 73-88. https://doi.org/10.1002/2017JD027448

Zhang, Q., Yu, H., Sun, P., Singh, V. P. & Shi, P. (2019). Multisource data based agricultural drought monitoring and agricultural loss in China. *Global and Planetary Change*, 172, 298-306. https://doi.org/10.1016/j.gloplacha.2018.10.017

Zhang, T., Su, X. & Feng, K. (2021). The development of a novel nonstationary meteorological and hydrological drought index using the climatic and anthropogenic indices as covariates. *Science of the Total Environment*, 786, 147385. https://doi.org/10.1016/j.scitotenv.2021.147385

Zhang, X., Su, Z., Lv, J., Liu, W., Ma, M., Peng, J., et al. (2019). A Set of Satellite-Based Near Real-Time Meteorological Drought Monitoring Data over China. *Remote Sensing*, 11(4), 453. https://doi.org/10.3390/rs11040453

Zhang, Y., Hao, Z., Feng, S., Zhang, X., Xu, Y. & Hao, F. (2021). Agricultural drought prediction in China based on drought propagation and large-scale drivers. *Agricultural Water Management*, 255, 107028. https://doi.org/10.1016/j.agwat.2021.107028

Zhang, Y., Wang, Z., Sha, S. & Feng, J. (2018). Drought Events and Its Causes in Summer of 2018 in China. *Journal of Arid Meteorology*, 36(5), 884-892. https://doi.org/10.11755/j.issn.1006-7639(2018)-05-0884

Zhao, S. (1983). A new scheme for comprehensive physical regionalization in China. *Acta*

https://doi.org/10.5194/hess-2021-599
Preprint. Discussion started: 13 December 2021

Geographica Sinica, 38(1), 1-10.

Zhou, S., Williams, A. P., Berg, A. M., Cook, B. I., Zhang, Y., Hagemann, S., et al. (2019). Land-atmosphere feedbacks exacerbate concurrent soil drought and atmospheric aridity. *Proceedings of the National Academy of Sciences of the United States of America*, 116(38), 18848-18853. https://doi.org/10.1073/pnas.1904955116

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R. M., et al. (2020). A typology of compound weather and climate events. *Nature Reviews Earth & Environment*, 1(7), 333-347. https://doi.org/10.1038/s43017-020-0060-z

https://doi.org/10.5194/hess-2021-599
Preprint. Discussion started: 13 December 2021

Hydrology and Earth System Sciences Discussions
Open Access
EGU

Figure Captions

Figure 1. Seven sub-climate regions division over China. The specific information of climate regions D1–D7 is listed at the left-bottom in the panel.

Figure 2. Different schematic (two types) of C-vine copulas under three-dimensional scenarios. For the first type (a), the ordering variables are y_1, y_2, and y_3, while for the second type (b) that are y_2, y_1, and y_3. $C_{12}(C_{21})$, $C_{13}(C_{23})$, and $C_{23|1}(C_{13|2})$ denotes bivariate copulas with parameters θ_{11}, θ_{12}, and θ_{21}, respectively. Here, θ_{ij} signifies the parameters of the j-th edge with respect to the i-th tree. $G(\bullet|\bullet)$ denote conditional distribution functions.

Figure 3. Spatial patterns of 1–3-months lag Kendall's correlation coefficient (τ_k) between SPI_{t-i} and SSI_t (t denotes August, and i is 1–3-month lag time) (top row), as well as SSI_{t-i} and SSI_t (bottom row) for August during 1961–2018 over China. Note the stippling indicates where τ_k is at a 0.05 significance level, which is corrected via the false discovery rate (FDR) of 0.1.

Figure 4. Forecast performance of the 3C-vine model based on (a–c) NSE_{3C-MG} (difference of NSE between 3C-vine model and MG model), (d–f) R^2_{3C-MG} (difference of R^2 between 3C-vine and MG models), and (g–i) RMSE_{3C-MG} (difference in RMSE between 3C-vine and MG models) for the 1–3-month leads of August during 1961–2018 over China. The corresponding boxplots of (j) NSE_{3C-MG}, (k) R^2_{3C-MG}, and (l) RMSE_{3C-MG} relative to a threshold of 0 (horizontal black dash line) for agricultural drought forecast in August under 1–3-month leads in climate regions D1–D7 over China. The percentage of $\mathrm{NSE}_{3C-MG} > 0$, $R^2_{3C-MG} > 0$, and $\mathrm{RMSE}_{3C-MG} < 0$ is listed in the left-bottom of corresponding sub-figure, respectively.

https://doi.org/10.5194/hess-2021-599
Preprint. Discussion started: 13 December 2021

Figure 5. SSI observations in August of 2018 (a) as well as the corresponding SSI forecasts under 1–3-month lead times utilizing 3C-vine model (b–d) and MG model (e–g) over China. The black rectangle boxes (as shown in b) denote the typical regions (corresponding to signify D1S–D7S) selected in climate regions D1–D7.

Figure 6. Probability density function (PDF) curve of (a) minimum and (b) maximum SSI under 1–3-month lead times for August during the 1961–2018 period over seven selected typical regions in climate regions D1–D7 (i.e., these black rectangle boxes in Figure 5b correspond to signify D1S–D7S, respectively). Black dash line and text indicate the (a) minimum and (b) maximum observations of SSI in D1S–D7S. These texts with red, blue, and cyan colors of top-right in each sub-figure are SSI forecasts under 1–3-month lead times of August, which correspond to the abscissa projected by the peak point of each PDF.

https://doi.org/10.5194/hess-2021-599
Preprint. Discussion started: 13 December 2021

Hydrology and Earth System Sciences
Discussions

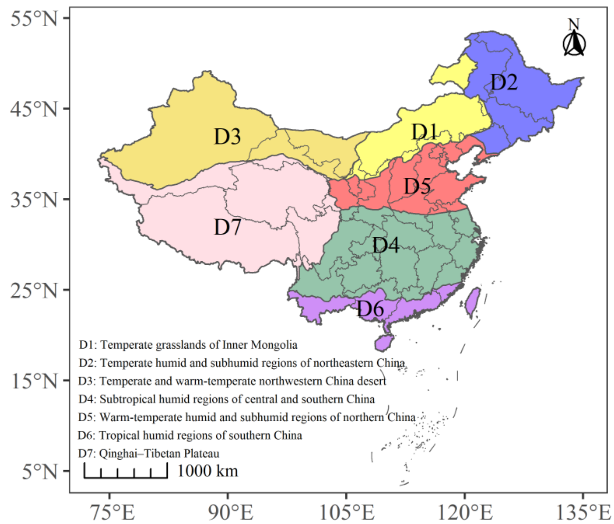

Figure 1. Seven sub-climate regions division over China. The specific information of climate regions D1–D7 is listed at the left-bottom in the panel.

https://doi.org/10.5194/hess-2021-599
Preprint. Discussion started: 13 December 2021

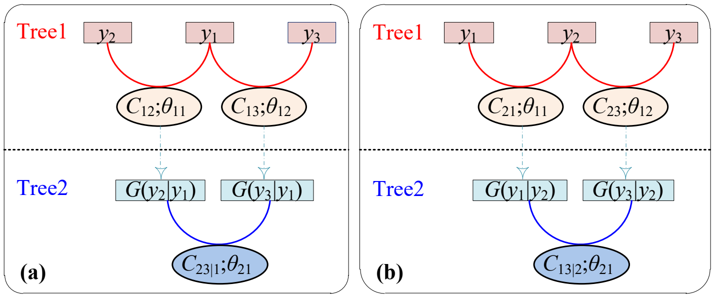

Figure 2. Different schematic (two types) of C-vine copulas under three-dimensional scenarios. For the first type (a), the ordering variables are y_1, y_2, and y_3, while for the second type (b) that are y_2, y_1, and y_3. C_{12}(C_{21}), C_{13}(C_{23}), and $C_{23|1}$($C_{13|2}$) denotes bivariate copulas with parameters θ_{11}, θ_{12}, and θ_{21}, respectively. Here, θ_{ij} signifies the parameters of the j-th edge with respect to the i-th tree. $G(\bullet|\bullet)$ denote conditional distribution functions.

https://doi.org/10.5194/hess-2021-599
Preprint. Discussion started: 13 December 2021

Hydrology and Earth System Sciences
Open Access
Discussions

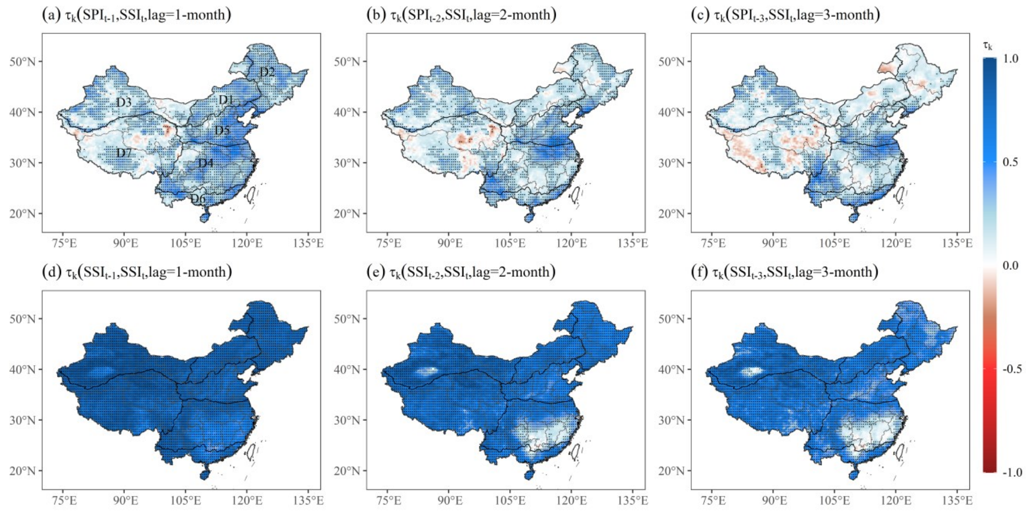

Figure 3. Spatial patterns of 1–3-months lag Kendall's correlation coefficient (τ_k) between SPI_{t-i} and SSI_t (t denotes August, and i is 1–3-month lag time) (top row), as well as SSI_{t-i} and SSI_t (bottom row) for August during 1961–2018 over China. Note the stippling indicates where τ_k is at a 0.05 significance level, which is corrected via the false discovery rate (FDR) of 0.1.

https://doi.org/10.5194/hess-2021-599
Preprint. Discussion started: 13 December 2021

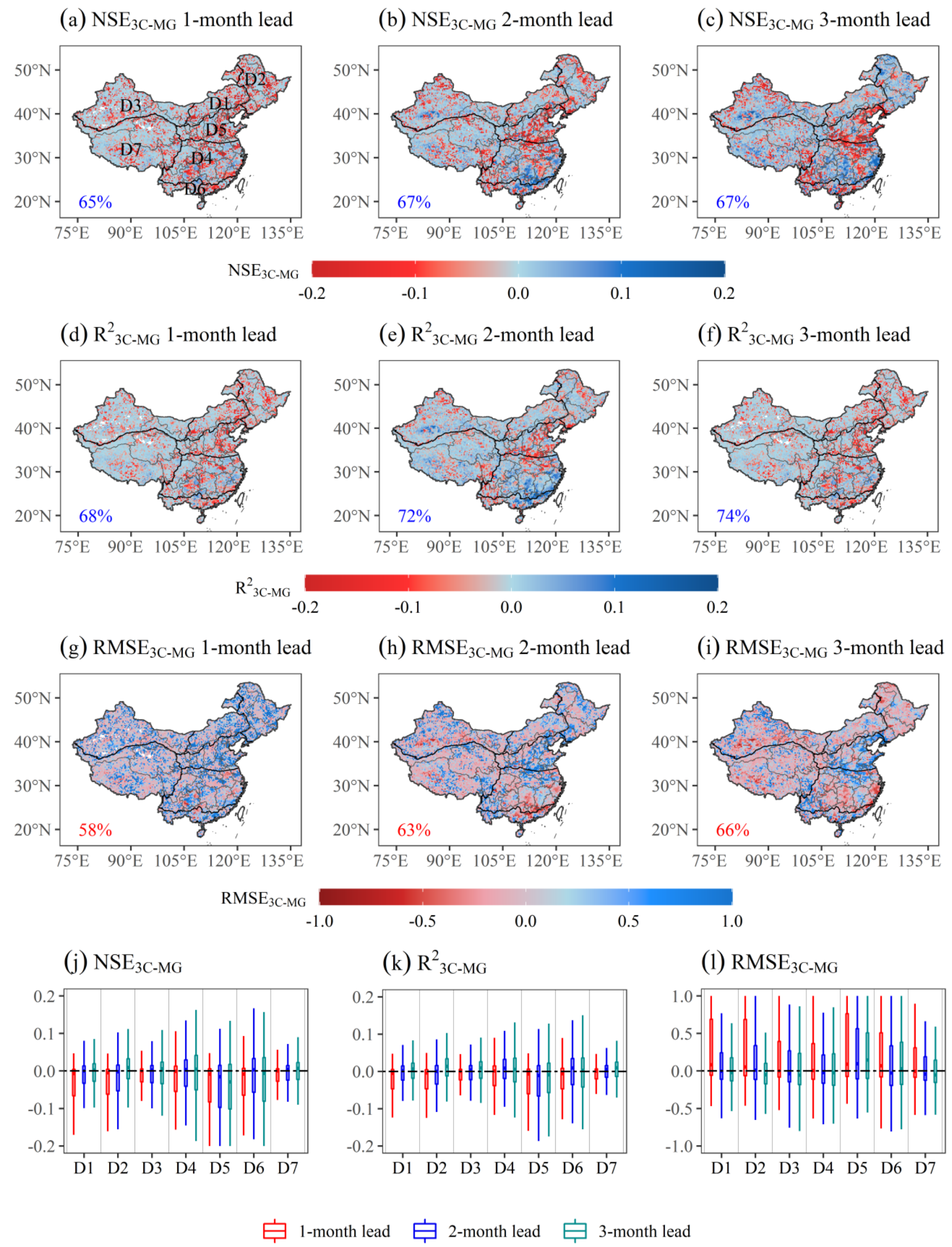

Figure 4. Forecast performance of the 3C-vine model based on (a–c) $NSE_{3C\text{–}MG}$ (difference of NSE between 3C-vine model and MG model), (d–f) $R^2_{3C\text{–}MG}$ (difference of R^2 between 3C-vine and MG

https://doi.org/10.5194/hess-2021-599
Preprint. Discussion started: 13 December 2021

models), and (g–i) $RMSE_{3C–MG}$ (difference in RMSE between 3C-vine and MG models) for the 1–3-month leads of August during 1961–2018 over China. The corresponding boxplots of (j) $NSE_{3C–MG}$, (k) $R^2_{3C–MG}$, and (l) $RMSE_{3C–MG}$ relative to a threshold of 0 (horizontal black dash line) for agricultural drought forecast in August under 1–3-month leads in climate regions D1–D7 over China. The percentage of $NSE_{3C–MG} > 0$, $R^2_{3C–MG} > 0$, and $RMSE_{3C–MG} < 0$ is listed in the left-bottom of corresponding sub-figure, respectively.

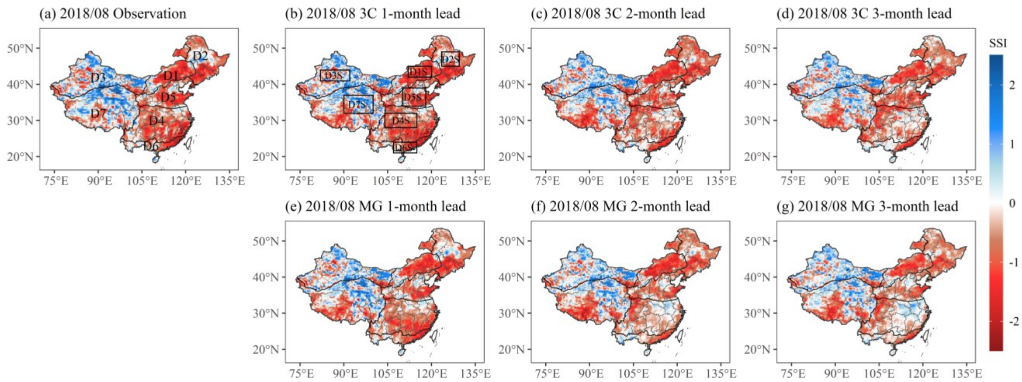

Figure 5. SSI observations in August of 2018 (a) as well as the corresponding SSI forecasts under 1–3-month lead times utilizing 3C-vine model (b–d) and MG model (e–g) over China. The black rectangle boxes (as shown in b) denote the typical regions (corresponding to signify D1S–D7S) selected in climate regions D1–D7.

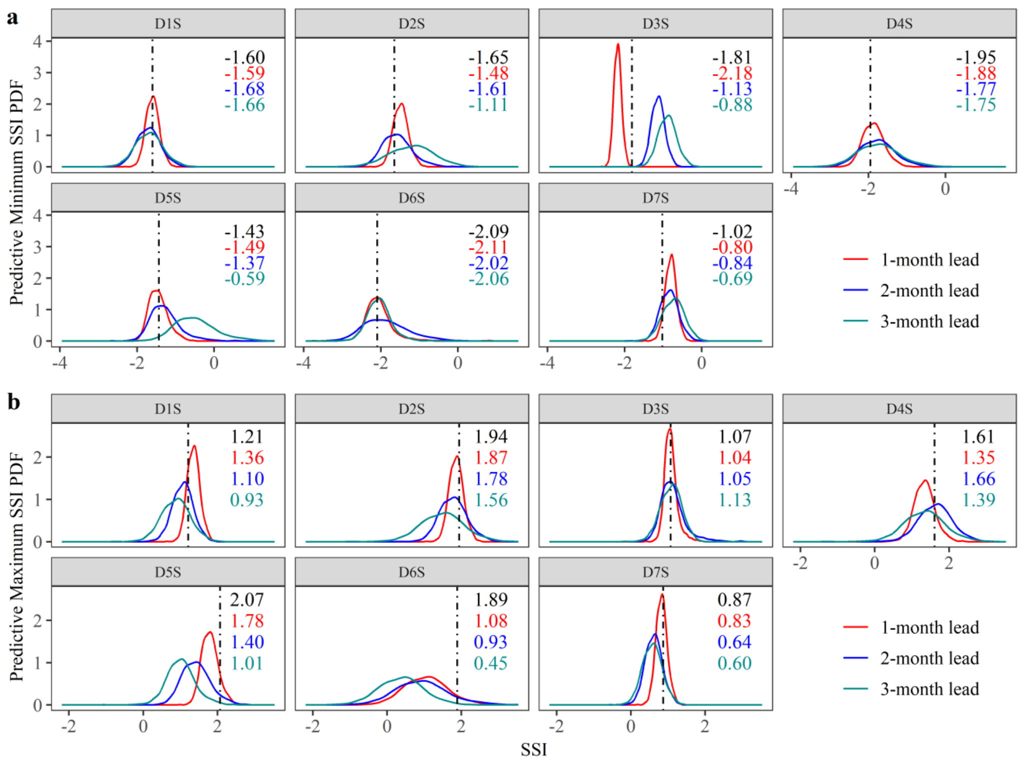

Figure 6. Probability density function (PDF) curve of (a) minimum and (b) maximum SSI under 1–3-month lead times for August during the 1961–2018 period over seven selected typical regions in climate regions D1–D7 (i.e., these black rectangle boxes in Figure 5b correspond to signify D1S–D7S, respectively). Black dash line and text indicate the (a) minimum and (b) maximum observations of SSI in D1S–D7S. These texts with red, blue, and cyan colors of top-right in each sub-figure are SSI forecasts under 1–3-month lead times of August, which correspond to the abscissa projected by the peak point of each PDF.