The Interdependence between Rainfall and Temperature: Copula Analyses

Cong, Ronggang; Brady, Mark

Published in:
Scientific World Journal

DOI:
10.1100/2012/405675

Published: 2012-01-01

Citation for published version (APA):
Research Article
The Interdependence between Rainfall and Temperature: Copula Analyses

Rong-Gang Cong¹ and Mark Brady¹,²

¹ Centre for Environmental and Climate Research (CEC), Lund University, Lund S-22362, Sweden
² AgriFood Economics Centre, Department of Economics, Swedish University of Agricultural Sciences, Lund S-22007, Sweden

Correspondence should be addressed to Rong-Gang Cong, ronggang.cong@cec.lu.se

Received 17 August 2012; Accepted 9 October 2012

Academic Editors: G. O. Thomas and C. Varotsos

Copyright © 2012 R.-G. Cong and M. Brady. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Rainfall and temperature are important climatic inputs for agricultural production, especially in the context of climate change. However, accurate analysis and simulation of the joint distribution of rainfall and temperature are difficult due to possible interdependence between them. As one possible approach to this problem, five families of copula models are employed to model the interdependence between rainfall and temperature. Scania is a leading agricultural province in Sweden and is affected by a maritime climate. Historical climatic data for Scania is used to demonstrate the modeling process. Heteroscedasticity and autocorrelation of sample data are also considered to eliminate the possibility of observation error. The results indicate that for Scania there are negative correlations between rainfall and temperature for the months from April to July and September. The student copula is found to be most suitable to model the bivariate distribution of rainfall and temperature based on the Akaike information criterion (AIC) and Bayesian information criterion (BIC). Using the student copula, we simulate temperature and rainfall simultaneously. The resulting models can be integrated with research on agricultural production and planning to study the effects of changing climate on crop yields.

1. Introduction

Weather is the key source of uncertainty affecting crop yield especially in the context of climate change [1–3]. For example, Vergara et al. studied the potential impact of catastrophic weather on the crop insurance industry and found that 93% of crop loss was directly related to unfavorable weather [4]. Accurate modeling of multivariate weather distributions would allow farmers to make better decisions for reducing their exposure to weather risk or take advantage of favorable climatic relationships [5]. Among variables relevant to weather, rainfall and temperature are two important factors which have a large effect on crop yield [6–9]. Typically, temperature affects the length of the growing season and rainfall affects plant production (leaf area and the photosynthetic efficiency) [10, 11].

There is a lot of literature studying the effects of temperature and rainfall on crop yield. Erskine and El Ashkar quantified the effect of rainfall on lentil seed yield and found that rainfall accounted for 79.8% of the variance of seed yield [12]. Lobell et al. studied 12 major Californian crops and found rainfall was able to explain more than 60% of the observed variability in yields for most crops [13]. Cooper et al. found that not only the seasonal rainfall totals and their season-to-season variability were important, but also the “within season” variability had a major effect on crop productivity [14], which implies that monthly data is needed in crop production analysis.

Muchow et al. found that lower temperature increased the length of time that the maize could intercept radiation and hence grow [15]. Lobell and Asner found a roughly 17% relative decrease in both corn and soybean yield in the USA for each degree of increase in growing season temperature [16]. In summary, it is well established that rainfall and temperature are two important climatic factors affecting agricultural production [17–19].

Since temperature and rainfall are critical determinants of crop yield, accurate simulation of temperature and rainfall
is important not only for meteorology but also for agricultural economics. However, in reality it is difficult to simulate rainfall and temperature simultaneously due to the interdependence (correlation) between them [20–22]. Spatially, it is generally believed that there exists significant correlation between rainfall and temperature over tropical oceans and land [23]. For example, Aldrian and Dwi Susanto examined the relationship between rainfall and sea surface temperature and found that Indonesian rainfall variability revealed some sensitivity to sea-surface temperature variability in adjacent parts of the Indian and Pacific Oceans [24]. Black also studied the relationship between Indian Ocean sea surface temperature and East Africa rainfall and concluded that strong East African rainfall was associated with warming in the Pacific and Western Indian Oceans and cooling in the Eastern Indian Ocean [25].

Temporally, it is generally believed that the correlation between rainfall and temperature changes between months. For example, Rajeevan et al. examined the temporal relationship between land surface temperature and rainfall [26]. They found that temperature and rainfall were positively correlated during January and May but negatively correlated during July. Using annual data Huang et al. also found a negative correlation between rainfall and temperature in Yellow River basin of China [27].

To take the interdependence between rainfall and temperature into account, multivariate probability simulation is needed. Traditionally multivariate probability density functions, however, are generally limited to the multivariate normal distribution or mixtures of it [28]. A possible method that provides an alternative is the copula method. Copulas are advantageous because they can model joint distributions of random variables with greater flexibility both in terms of marginal distributions and the dependence structure [29]. Copulas have been used in financial economics for quite some time [30–32]. However, there are relatively few applications to agricultural weather simulation.

In respect to temperature and rainfall, AghaKouchak et al. applied two different elliptical copula families, namely, Gaussian and t-copula, to simulate the spatial dependence of rainfall and found that using the t-copula might have significant advantages over the well-known Gaussian copula particularly with respect to extremes [33]. Serinaldi also studied the spatial dependence of rainfall and confirmed that only positive contemporaneous pairs of rainfall observations correctly described the intersite dependence [34]. Laux et al. highlighted the importance of pretreatment of meteorological data in the copula modeling process [35]. Laux et al. used the Clayton copula to construct the bivariate distribution of drought duration and intensity [36]. Similar applications of the Clayton copula can also be found in the studies of Favre et al. and Shiau et al. [37, 38]. Furthermore, they raised the question as to which copula model best fitted the empirical data. The only literature concerning the application of copula simulation to model the interdependence between temperature and rainfall up to now is that of Schölzel and Friederichs [39]. They used a simple statistical model based on the copula approach to describe the phenomenon that cold periods were accompanied by small precipitation amounts.

Inspired by Dupuis’s study on hydrological random variables [40], the purpose of this paper is to illustrate the pretreatment process of meteorological data, demonstrate the application of different copulas to modeling of joint distributions of rainfall and temperature, select the most suitable copula function according to information criteria, and finally simulate rainfall and temperature simultaneously.

2. Materials and Methods

2.1. Study Area. Scania is Sweden’s southernmost province and one of Northern Europe’s most fertile farming districts with the main crops being winter wheat, rapeseed, sugar beets, and barley. As Scania is surrounded by water on three sides (the Baltic Sea, the Kattegat Sea, and the Öresund Sound), it has a maritime climate, especially along the south and east coasts. The winters are mild (few days of snow), but the summers are similar to those in the rest of southern Sweden.

2.2. Data Collection and Preliminary Analysis. Monthly temperature and rainfall data for Scania from 1961 to 2010 was obtained from the Swedish Meteorological and Hydrological Institute.

2.2.1. Temperature. Monthly average temperature in Scania shows a clear seasonal cycle from 1961 to 2010 (Figure 1). The average temperature usually reaches its peak in July and its bottom in February. From April to November, the average temperature is always above 0°C. The variability of average temperature in January and February is though relatively large. Some descriptive temperature statistics are listed in Table 1.
2.2.2. Rainfall. Compared with temperature, monthly total rainfall in Scania does not show a clear seasonal cycle from 1961 to 2010. From June to November, the average monthly total rainfall is relatively high (Figure 2). Some descriptive rainfall statistics are listed in Table 2.

2.2.3. The Relationship between Rainfall and Temperature. The physical rationale behind the relationship between rainfall and temperature is that rainfall may affect soil moisture which in turn affect surface temperature by controlling the partitioning between the sensible and latent heat fluxes [41]. Because the sample data is non-Gaussian distributed and skewed, the Kendall correlation coefficient is employed to calculate the correlation between monthly rainfall and temperature. It is found that there are negative correlations between rainfall and temperature from April to July and in September (at the 10% confidence level) (Table 3).

2.3. Methods. Here we use the copula functions to model the interdependence between the probability distributions of a certain month’s temperature and rainfall. Let $X$ and $Y$ be continuous random variables representing temperature and rainfall, with cumulative distribution functions $F_X(x) = \Pr(X \leq x)$ and $G_Y(y) = \Pr(Y \leq y)$, respectively. Following Sklar [42], there is a unique function $C$ such that

$$\Pr(X \leq x, Y \leq y) = C(F(x), G(y)),$$

where $C(u, v) = \Pr(U \leq u, V \leq v)$ is the distribution of the pair $(U, V) = (F(X), G(Y))$ whose margins are uniform on $[0, 1]$. The function $C$ is called a copula. As argued by Joe [43]...
Table 3: Correlation analysis for monthly temperature and rainfall from 1961 to 2010.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Kendall correlation coefficients</td>
<td>0.12</td>
<td>0.13</td>
<td>0.07</td>
<td>-0.27</td>
<td>-0.3</td>
<td>-0.17</td>
<td>-0.3</td>
<td>-0.22</td>
<td>-0.02</td>
<td>-0.19</td>
<td>-0.13</td>
<td>-0.02</td>
</tr>
<tr>
<td>P value</td>
<td>0.22</td>
<td>0.19</td>
<td>0.49</td>
<td>0.007</td>
<td>0.002</td>
<td>0.08</td>
<td>0.002</td>
<td>0.84</td>
<td>0.06</td>
<td>0.19</td>
<td>0.85</td>
<td>0.37</td>
</tr>
</tbody>
</table>

Figure 3: Temperature and rainfall in April from 1961 to 2010.

Figure 4: Sample autocorrelation function (ACF) of temperature and rainfall in April from 1961 to 2010.

Figure 5: Sample autocorrelation function (ACF) of AR adjusted temperature and rainfall in April.
and Nelsen [44] among others, $C$ characterizes the dependence in the pair $(X, Y)$. There are many parametric copula families available, which usually have parameters that control the strength of dependence. Among these, five families of commonly used copulas are considered. They are listed in Table 4, along with their parameter ranges. The first three are Archimedean [43] and the last two are metaelliptical [45].

After calculating the parameters of each copula, it is necessary to decide which family is the best representation of the dependence structure between the variables of interest. There are a few techniques to select the best copula. One of them is based on distance measures pertaining to the distributions of the candidate models (copulas) and the empirical distribution of the data [46, 47]. Alternative methods include likelihood ratio tests and approaches related to information criteria [31], such as Akaike [48] and Schwarz’s Bayesian [49] Information Criteria. Information criteria are adopted here because they can describe the tradeoff between bias (accuracy) and variance (complexity) in model construction. The Akaike information criterion (AIC) is a measure of the relative goodness of fit of a statistical model. Its definition is

$$AIC = 2k - 2 \ln(L),$$  

where $k$ is the number of parameters in the copula and $L$ is the maximized value of the likelihood function for the copula. The Bayesian information criterion (BIC) was developed by Schwarz using Bayesian formalism. Its definition is

$$BIC = -2 \ln(L) + k \ln(N),$$

where $N$ is the sample size.
3. Results and Discussion

Temperature and rainfall data in April from 1961 to 2010 is employed as an example to demonstrate the modeling process (Figure 3). There is a significant negative relationship (Kendall correlation coefficient is $-0.27$, $P$-value = 0.007) between temperature and rainfall in April. Temperature has negative skewness ($-0.35$) and rainfall has positive skewness (1.07), which may cause a heteroscedasticity problem when fitting the model [50]. Following Kim and Ahn [51], the temperature and rainfall data are log-transformed to remove this effect. The logarithmic transformation for the data is invertible, which will not affect the fitting results.

Following Benth and Šaltyte-Benth’s instructions [52], the time series of temperature and rainfall are tested for autocorrelation using the $Q$-statistics (Figure 4). Autocorrelation describes the correlation between values of temperature (or rainfall) at different points in time, as a function of the time difference. The presence of autocorrelation increases the variances of residuals and estimated coefficients, which reduces the model’s efficiency. The Ljung-Box $Q$ test is a type of statistical test of whether autocorrelations of a time series...
The real CDFs while the solid lines are the simulated CDFs.

Figure 9: Real versus Gauss fitted CDF. Note: The dashed lines are the real CDFs while the solid lines are the simulated CDFs.

Figure 10: Real versus student fitted CDF. Note: The dashed lines are the real CDFs while the solid lines are the simulated CDFs.

Figure 11: Real versus Clayton fitted CDF. Note: The dashed lines are the real CDFs while the solid lines are the simulated CDFs.

Figure 12: Real versus Frank fitted CDF. Note: The dashed lines are the real CDFs while the solid lines are the simulated CDFs.

Table 4: Five families of copulas.

<table>
<thead>
<tr>
<th>Family</th>
<th>C(u, v)</th>
<th>Range of θ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>$N_θ(\Phi^{-1}(u), \Phi^{-1}(v))$</td>
<td>$[-1, 1]$</td>
</tr>
<tr>
<td>Student</td>
<td>$T_{θ,γ}(T_{γ}^{-1}(u), T_{γ}^{-1}(v))$</td>
<td>$[-1, 1]$</td>
</tr>
<tr>
<td>Clayton</td>
<td>$(u^θ + v^{-θ} - 1)^{-1/θ}$</td>
<td>$(0, \infty)$</td>
</tr>
<tr>
<td>Frank</td>
<td>$\theta^{-1} \ln \left[1 + (e^{-θu} - 1)(e^{-θv} - 1)/(e^{-θ} - 1)\right]$</td>
<td>$(-\infty, \infty)$</td>
</tr>
<tr>
<td>Gumbel</td>
<td>$\exp \left[-\left(\ln u\right)^θ + \left(\ln v\right)^θ\right]$</td>
<td>$[1, \infty)$</td>
</tr>
</tbody>
</table>

Φ: cumulative distribution function (CDF) of a $N(0,1)$.
$N_θ$: CDF of a standard bivariate normal distribution with Pearson correlation θ.
$T_{θ,γ}$: CDF of a student distribution with γ degrees of freedom.
$T_{γ}$: CDF of a bivariate student distribution with γ degrees of freedom.

Table 5: Results of different copula models for temperature and rainfall in April.

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Student</th>
<th>Clayton</th>
<th>Frank</th>
<th>Gumbel</th>
</tr>
</thead>
<tbody>
<tr>
<td>θ</td>
<td>-0.34</td>
<td>-0.31</td>
<td>0.001</td>
<td>0.001</td>
<td>1.1</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>3.05</td>
<td>4.11</td>
<td>-0.0007</td>
<td>-0.0002</td>
<td>-1.86</td>
</tr>
<tr>
<td>AIC</td>
<td>-6.06</td>
<td>-8.15</td>
<td>0.042</td>
<td>0.041</td>
<td>3.75</td>
</tr>
<tr>
<td>BIC</td>
<td>-6.06</td>
<td>-8.07</td>
<td>0.081</td>
<td>0.08</td>
<td>3.79</td>
</tr>
</tbody>
</table>

are different from zero [53]. The Q-statistics is defined as follows:

$$Q = N(N + 2) \sum_{a=1}^{h} \frac{\hat{ρ}_a^2}{N - a},$$  \hspace{1cm} (4)

where $\hat{ρ}_a^2$ is the sample autocorrelation at lag $a$, and $h$ is the number of lags being tested. The first-order autocorrelations are found to be strong both for temperature ($Q$-stat = 6.32, $P$ value = 0.01) and rainfall ($Q$-stat = 4.52, $P$ value = 0.03), as shown in Figure 4.

Therefore, an AR(1) model is used to eliminate the autocorrelation in the series as follows:

$$\text{tempe}_t = 0.48 + 0.35 \times \text{tempe}_{t-1} + \varepsilon_t,$$

(4.7**) (2.56**),

(5)

$$\text{rain}_t = 1.85 - 0.29 \times \text{rain}_{t-1} + \mu_t,$$

(9.06**) (-2.1**).
Note that the numbers in the bracket are t-values and ** stands for the statistical significance at the 95% confidence level.

Residuals $\varepsilon_t$ and $\mu_t$ are tested where only weak autocorrelations are found (Figure 5).

In addition to autocorrelation, time trends are also found in the series of $\varepsilon_t$ and $\mu_t$. Based on Manton et al.’s research [54], the time trends should be removed from the series to obtain a stationary process. The functions used to detrend the time series are

\begin{align*}
\varepsilon_t &= -0.08 + 0.0032 \times t + \varphi_t \\
&\quad (-2.65^{**})(3.04^{**}), \\
\mu_t &= 0.17 - 0.007 \times t + \gamma_t \\
&\quad (2.3^{**})(-2.65^{**}).
\end{align*}

We find that temperature has an increasing trend and rainfall has a decreasing trend in April from 1961 to 2010 (Figure 6). The annual rate of increase in temperature in April is 0.0032°C and decrease in rainfall is 0.007 mm per year. The trend adjusted data are shown in Figure 7 where rtempe and rrain are used to represent the corrected values of $\varepsilon_t$ and $\gamma_t$, respectively.

The residuals for the trend adjusted variables have negative skewness: temperature (−1) and rainfall (−0.7). Based on the inference for the margins (IFM) [55], the parameter estimates and model evaluation indices for each copula for rtempe and rrain are presented in Table 5.

The log-likelihood ratio is largest and the AIC and BIC are smallest for the student copula, which means that the student copula is the most suitable model. A comparison of the real and simulated residuals of temperature and rainfall is shown in Figure 8.

Since the purpose of this paper is to develop a copula model of the bivariate distribution of rainfall and temperature that can be used in simulation studies, the accuracy of the resulting model is of utmost importance. Although Table 5 has provided some statistical support for the model and Figure 8 has given some visual evidence, the contours of the cumulative distribution functions can best show the difference between the real and simulated data.

In Figures 9, 10, 11, 12, and 13, the contours of the cumulative distribution functions (CDFs) for the real and simulated data from the five copula models are plotted to visualize the difference or similarity in the distributions as the case may be. It is found that the student copula model fits the real data best according to the similarity of the contour lines. Consequently the student copula is the best choice of model according to all our criteria.

Based on the estimated parameters, 1,000 draws are made from the Student copula model. The simulated data is then transformed to the original scale and compared with the real data in Figure 14.

4. Conclusions

This paper presents a copula-based methodology for modeling the joint distribution of temperature and rainfall, which are of utmost importance for agricultural production especially in the context of climate change. Copulas have been used extensively in the financial literature, but have not been widely used in weather simulation. The copula approach provides a powerful and flexible method to model multivariate distributions and thus goes beyond joint normality, regression, and mean-variance criterion. Accurate simulation of weather events may help to improve risk management in agricultural planning.

A shortcoming of the copula method is the arbitrariness of the selection of a particular copula. The main purpose of this paper is to present a complete copula modeling framework to model the interdependence of rainfall and temperature. In contrast to Schölzel and Friederichs [39], we compare different copulas and show how to select the optimal copula based on information criteria (AIC and BIC). The advantage of this approach is that it does not require any assumptions and is primarily data driven thus minimizing the subjectivity introduced by the researcher. The model selection criteria indicate that the student copula produces the best model to simulate the dependence structure between rainfall and temperature in Scania, Sweden.

Although the month of April was chosen as our working example, we have also tested the data for other months with similar results. The study is only based on meteorological data for a single region. The most suitable copula family for rainfall and temperature might change from one region to another due to differences in geographical and geophysical conditions. Our approach however can be applied in studies of other parts of the world to select the most appropriate copula model. A potentially valuable extension of this research is to connect the analysis with crop production planning and agricultural economics. If the relationship among temperature, rainfall, and crop yield can be determined, then it could be used in developing risk reducing strategies for farmers, something which will become increasingly important in the face of climate change. This is the focus of our ongoing research.
Figure 14: Real (a) and student-based copula simulated (b) temperature and rainfall data for Scania in April.

Acknowledgments

The authors gratefully acknowledge the good research environments provided by the Centre for Environmental and Climate Research (CEC), Lund University, and AgriFood Economics Centre. We also thank the editors and anonymous referees for their helpful suggestions on the earlier draft of our paper according to which we improved the content. This research is funded by “Biodiversity and Ecosystem services in a Changing Climate (BECC)” and SAPES. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the paper. The contents are the responsibility of the authors and do not necessarily reflect the views of the Centre for Environmental and Climate Research (CEC), Lund University, or AgriFood Economics Centre.

References


