

2024

A Copula Discretization of Time Series-Type Model for Examining Climate Data

Dimuthu Fernando
Wake Forest University

Olivia Atutey
University of South Alabama

Norou Diawara
Old Dominion University, ndiawara@odu.edu

Follow this and additional works at: https://digitalcommons.odu.edu/mathstat_fac_pubs



Part of the [Data Science Commons](#), [Mathematics Commons](#), and the [Statistics and Probability Commons](#)

Original Publication Citation

Fernando, D., Atutey, O., & Diawara, N. (2024). A copula discretization of time series-type model for examining climate data. *Mathematics*, 12(15), 1-13, Article 2419. <https://doi.org/10.3390/math12152419>

This Article is brought to you for free and open access by the Mathematics & Statistics at ODU Digital Commons. It has been accepted for inclusion in Mathematics & Statistics Faculty Publications by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

Article

A Copula Discretization of Time Series-Type Model for Examining Climate Data

Dimuthu Fernando ¹, Olivia Atutey ² and Norou Diawara ^{3,*} ¹ Department of Statistical Sciences, Wake Forest University, Winston-Salem, NC 27106, USA; fernand@wfu.edu² Department of Mathematics and Statistics, University of South Alabama, Mobile, AL 36688, USA; oatutey@southalabama.edu³ Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA 23529, USA

* Correspondence: ndiawara@odu.edu

Abstract: The study presents a comparative analysis of climate data under two scenarios: a Gaussian copula marginal regression model for count time series data and a copula-based bivariate count time series model. These models, built after comprehensive simulations, offer adaptable autocorrelation structures considering the daily average temperature and humidity data observed at a regional airport in Mobile, AL.

Keywords: count time series; copula; bivariate models

MSC: 62-08; 62H10



Citation: Fernando, D.; Atutey, O.; Diawara, N. A Copula Discretization of Time Series-Type Model for Examining Climate Data. *Mathematics* **2024**, *12*, 2419. <https://doi.org/10.3390/math12152419>

Academic Editor: Manuel Alberto M. Ferreira

Received: 14 June 2024

Revised: 29 July 2024

Accepted: 31 July 2024

Published: 3 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Multivariate count time series are prevalent in modern statistical analysis, and they rarely exhibit independence between the series. Discretization of time series data has been proposed in data mining by authors such as Chaudhari et al. (2014) [1], Marquez-Grajales et al. (2020) [2], and Mordvanyuk et al. (2022) [3]. Our objective is to propose a modeling of the discrete data with temporal dependence. Pair copula construction has gained popularity, as it offers flexibility and useful advantages in the learning of the joint distribution of time series data. As we know, Pearson's correlation is based on the normal shape of the marginal distributions. Stretching and reshaping the marginal distributions will give different answers. Categorizing the values as it is usually done in climate research (e.g., saying the temperature is in the 80s) creates challenges that can be overcome with the copula-based modeling. The advantages of using the copula also include the distribution assumptions, the time dependence, the mixture of the types of marginal distributions, and the addition of covariates. Further, there are new connections or time markers, that may have been missed in how observations are linked, that can be discovered. In this paper, we propose a comparative study of the climate data under the following two different cases:

- The Gaussian copula marginal regression model for count time series data;
- The copula-based bivariate count time series model.

These models are all part of a multivariate time series which offers flexible autocorrelation structures when the data are described as discrete counts.

Weather describes the short-term fluctuations of temperature, dew point, humidity, wind speed and direction, precipitation, atmospheric pressure, and other meteorological variables at a given location. Climate, on the other hand, is the long-term average variation of these meteorological variables at a location. By way of example, when we want to decide what clothes to put on each day, we can look at the weather, and what we stock up in our closet will probably depend on the climate of the place. The day-to-day activities of humans continue to change the make up of the earth's atmosphere. The causes of climate

and weather change are air pollution, deforestation, and increasing energy demand for heating and cooling, to name a few. Some other natural processes, such as volcanic eruptions, add to the increase in greenhouse gases in the atmosphere. When these atmospheric variables change quickly and negatively, they can impact many elements of human activity. In recent years, extreme weather conditions have displaced many people and exacerbated the factors driving people into poverty. These unfavorable weather events also increase health problems, as droughts and destructive storms cause food growth problems. It is essential to continue to explore more statistical methods, ranging from data collection to model development, to explain the variability of weather and climate data and assess future weather conditions so that people can be better prepared for extreme weather events. Advanced technology makes it possible to collect data for many weather-related factors. Thus, new approaches to visualizing and analyzing any relationship between these variables are needed as multivariate climate data sets become more widely available. A range of filter colors and patterns are available in modern data visualizations, which are intended to help users interactively visualize graphs and increase the forecasting of future trends under changing environmental conditions. Since there are also possible subjective interpretations of these graphs, statistical climate models based on the mathematical representations of the atmospheric variables remain a more dependable approach to obtaining information about current and future weather and climate states.

Quite a few data visualization methods have been developed to explore the inter-variable relations of these atmospheric properties. Teuling et al. [4] present a methodology which is an extension of that of [5] for describing the inter-variable relations of atmospheric properties. Their method is based on the properties of common color schemes to plot two variables in a single color map using a two-dimensional color legend for both sequential and diverging data. Concerning climate models, Agrawal [6] investigates the effectiveness of copula models in estimating and predicting climate extremes. Their study examines the bivariate distributions of temperature–humidity, temperature–wind speed, and wind speed–humidity in Boulder County, Colorado. The study bootstraps simulated data from a climate model and examines the accuracy of extreme event probability predictions when data are of different lengths and internal variability for the different copula functions. The analysis results reveal lower bias and variance for longer data records than for shorter data records when estimating the true probability of extreme compound events. Li et al. [7], in their paper, utilize three Archimedean copula models (the Clayton, Frank, and Gumbel copulas) to compare measured wave data to simulated wave climate data at a wave energy converter test site. In assessing the goodness of fit of the three models using R^2 , the study finds that Gumbel's copula performs better compared to the other two copulas. Lee et al. [8] apply the Clayton, Frank, Gumbel, and Gaussian copula functions to analyze the joint frequency of drought intensity and duration. They examine the performance of these copulas and find that the Frank and Gumbel copulas outperform the Clayton copula in the drought bivariate frequency analysis. The impacts of climate change and human activity have led other natural scientists to develop non-stationary multivariate analysis techniques to model these environmental changes (Li et al. [9], see also Yin et al. [10]).

The two most often discussed atmospheric properties that are connected to living situations are temperature and humidity. Temperature is the degree of warmth or coldness measured on a definite temperature scale using thermometers. Though the degree Celsius ($^{\circ}\text{C}$) scale and Kelvin (K) scale are used for temperature measuring purposes, the Fahrenheit ($^{\circ}\text{F}$) temperature scale is used by the United States and very few other countries. Relative humidity, typically expressed as a percentage, measures the amount of water vapor in the air relative to its capacity to hold it at a given temperature. We find that the temperature–relative humidity relationship is inversely proportional since relative humidity decreases as temperature rises and vice versa. Thus, temperature relates to the amount of moisture the atmosphere is able to hold. The way these variables interact affects human health and well-being as well as the weather. Barma et al. [11] utilize one-parameter bivariate

Archimedean copulas to assess the conditional probability of the number of COVID-19 cases given the mean daily temperature and relative humidity.

The literature has shown that climate data at different time points are correlated. These environmental phenomena exhibit both serial and cross-correlations that add complexities in the model specification and inference. Research in the case of negative correlation is still ongoing. In the bivariate case, the joint distribution is typically assumed to be normally distributed, but that assumption is easily violated because of changes in behavior or climate. Integrating over a discrete set requires special treatment after description. The bivariate copula offers flexibility about the underlying distributions. Moreover, when counts are used in the modeling, the theoretical and sample results lack consistency. The main reason is that the count data are modeled under the assumption of a conditional discrete distribution, instead of a marginal distribution.

Here, we use data from Mobile Regional Airport in Alabama, where the temperatures are measured over a 14-month period with thermal gradients. Data on humidity for the same period are also recorded. George et al. [12], in their research, establish that the temperature and rain amount are related. In their paper, they present linear regression models in describing bivariate relationships between the two meteorological variables. One can also regress temperature and humidity, but it is not obvious to define either variable as a response or predictor variable. Additionally, even though we know that these two variables are related, the relationship may not always be linear. To gain in efficiency, we propose to capture the dependence between these variables, in a multivariate distribution format with discrete classification. To circumvent limitations due to the normality assumption, or the fact that the data exhibit many same values, we investigate the relationship by applying copula-based time series models.

Our analysis differs twofold from the above analyses. We do not classify the temperature as high, low, or medium. We convert both humidity and temperature variables into a discrete scale by partitioning them in their associated time intervals. The copula-based approach will derive the relationship in a general framework.

The paper is organized as follows. Motivation of the bivariate time series data is given in Section 2, with a review of the correlation structure. Model construction in each of the cases and inference (under maximum likelihood estimation) are provided in Sections 3 and 4, respectively. The simulations and data application are shown in Sections 5 and 6, respectively, followed by a discussion and conclusion in Section 7.

2. Distributions

2.1. The Poisson Distribution

The Poisson distribution is one of the candidate distributions to model count data. We use the Poisson distribution as a marginal distribution to build our proposed copula-based bivariate model. Suppose y_t denotes a random observed count at time t . The probability mass function (pmf) of the well-known Poisson distribution is defined as:

$$f(y_t) = \frac{e^{-\lambda} \lambda^{y_t}}{y_t!},$$

where $\lambda > 0$ is the intensity parameter with $E(y_t) = \lambda$ and $V(y_t) = \lambda$.

2.2. Copulas

As a multivariate cumulative distribution function (cdf), the copula is a joint function that captures the dependence structure between variables. With uniform margins $U(0, 1)$ as in [13], an n -dimensional copula is a function $C : [0, 1]^n \rightarrow [0, 1]$ with the following three properties:

1. $C(1, \dots, u_t, \dots, 1) = u_t, \forall t = 1, 2, \dots, n$ and $u_t \in [0, 1]$.
2. $C(u_1, u_2, \dots, u_n) = 0$ if at least one $u_t = 0$ for $t = 1, 2, \dots, n$.

3. For any $u_{t_1}, u_{t_2} \in [0, 1]$ with $u_{t_1} \leq u_{t_2}$, for $t = 1, 2, \dots, n$,

$$\sum_{j_1=1}^2 \sum_{j_2=1}^2 \dots \sum_{j_n=1}^2 (-1)^{j_1+j_2+\dots+j_n} C(u_{1j_1}, u_{2j_2}, \dots, u_{nj_n}) \geq 0.$$

Let Y_1, \dots, Y_n be random variables with marginal cdfs F_1, \dots, F_n and joint cdf F , then we have the following:

1. There exists an n -dimensional copula C such that for all $y_1, \dots, y_n \in \mathbb{R}$

$$F(y_1, y_2, \dots, y_n) = C(F_1(y_1), F_2(y_2), \dots, F_n(y_n)).$$

2. If Y_1, \dots, Y_n are continuous, then the copula C is unique. Otherwise, C can be uniquely determined on n -dimensional rectangle $Range(F_1) \times Range(F_2) \times \dots \times Range(F_n)$.

When all the margins are integer valued, the multivariate probability mass function can be obtained as

$$\begin{aligned} f(y_1, y_2, \dots, y_n) &= P(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n) \\ &= \sum_{j_1=1}^2 \sum_{j_2=1}^2 \dots \sum_{j_n=1}^2 (-1)^{j_1+j_2+\dots+j_n} C(u_{1j_1}, u_{2j_2}, \dots, u_{nj_n}) \end{aligned} \tag{1}$$

where $u_{t1} = F_t(y_t)$ and $u_{t2} = F_t(y_t^-)$. Here, $F_t(y_t^-)$ is the left-hand limit of F_t at y_t , which is equal to $F_t(y_t - 1)$. In the bivariate case,

$$\begin{aligned} Pr(Y_1 = y_1, Y_2 = y_2) &= C(F(y_1), F(y_2); \theta) - C(F(y_1^-), F(y_2); \theta) \\ &\quad - C(F(y_1), F(y_2^-); \theta) + C(F(y_1^-), F(y_2^-); \theta). \end{aligned}$$

There are a number of copula functions, i.e., C , one can choose from. Table 1 shows some of the popular functions of copula families. For more details on these families, see [14]. Bivariate copulas like Gaussian, Frank, and T distributions can accommodate both positive and negative dependencies. Gumbel, Clayton, and Plackett copulas are restricted to model positive dependencies only.

Table 1. Bivariate copula functions.

| Copula | Copula Function |
|-------------|---|
| Gaussian | $C(u_1, u_2; \delta) = \Phi_\delta(\Phi^{-1}(u_1), \Phi^{-1}(u_2)), \delta \in [-1, 1]$ |
| Frank | $C(u_1, u_2; \delta) = -\frac{1}{\delta} \log \left[1 + \frac{(e^{-\delta u_1} - 1)(e^{-\delta u_2} - 1)}{e^{-\delta} - 1} \right], \delta \in \mathbb{R} \setminus \{0\}$ |
| Gumbel | $C(u_1, u_2; \delta) = \exp \left[- \left((-\log(u_1))^\delta + (-\log(u_2))^\delta \right)^{1/\delta} \right], \delta \geq 1$ |
| Clayton | $C(u_1, u_2; \delta) = (u_1^{-\delta} + u_2^{-\delta} - 1)^{-1/\delta}, \delta > 0$ |
| Plackett | $C(u_1, u_2; \delta) = \frac{[1+(\delta-1)(u_1+u_2)] - \sqrt{[1+(\delta-1)(u_1+u_2)]^2 - 4u_1u_2\delta(\delta-1)}}{2(\delta-1)}, \delta \geq 0$ |
| Bivariate t | $C(u_1, u_2; \delta) = \tau_\delta(\tau^{-1}(u_1), \tau^{-1}(u_2)), \delta \in [-1, 1]$ |

3. Copula-Based Model for Count Time Series Data

3.1. Gaussian Copula Marginal Regression Model

The Gaussian copula provides a mathematically convenient framework to handle various forms of dependence, for example, in time series analysis. To model the time

series data, Masarotto and Varin [15] describe a Gaussian copula model that emphasizes the regression setting when covariates are present. Let us consider a regression model with count time series Y_t as the response variable with X_t as the vector of covariates or independent variables, then the regression model can be represented as:

$$Y_t = g(X_t, \epsilon_t; \Theta), \quad t = 1, 2, \dots, n \tag{2}$$

where $g(\cdot)$ represents a function of covariates X_t and the error ϵ_t which captures the serial dependence. Further, Θ represents the vector of marginal model parameters and reduces to a scalar under the Poisson distribution, i.e., $\theta = \lambda$.

3.2. Copula-Based Bivariate Model

The bivariate integer-valued time series model was constructed via copula theory. Suppose we observe a series of 2-dimensional vector, $\{Y_t\}_{t=1}^n$, where $Y_t = (Y_{1t}, Y_{2t})'$ for $t = 1, 2, \dots, n$. Assume that each series $\{Y_{1t}\}_{t=1}^n$ and $\{Y_{2t}\}_{t=1}^n$ follows a copula-based first-order Markov process. Then, the mean vector μ_t , and the covariance matrix, say, $\Gamma(t, t - 1)$, are defined below:

$$\mu_t = E(Y_t) = \begin{bmatrix} E(Y_{1t}) \\ E(Y_{2t}) \end{bmatrix},$$

and

$$\begin{aligned} \Gamma(t, t - 1) &= \text{COV}(Y_t, Y_{t-1}) \\ &= \begin{bmatrix} \text{COV}(Y_{1t}, Y_{1,t-1}) & \text{COV}(Y_{1t}, Y_{2,t-1}) \\ \text{COV}(Y_{2t}, Y_{1,t-1}) & \text{COV}(Y_{2t}, Y_{2,t-1}) \end{bmatrix}. \end{aligned}$$

The diagonal elements in the covariance matrix represent the autocovariance within each time series, whereas the off-diagonal elements represent the cross covariance between the two time series. Hence, observing both serial dependence and cross-correlation, the joint probability distribution of Y_{1t} and Y_{2t} given $Y_{1,t-1}$ and $Y_{2,t-1}$, respectively, for $t = 1, \dots, n$ is given by:

$$f(y_{1t}, y_{2t} | y_{1,t-1}, y_{2,t-1}) = \int_{V^{-1}(F_{1,t}^-)}^{V^{-1}(F_{1,t}^+)} \int_{V^{-1}(F_{2,t}^-)}^{V^{-1}(F_{2,t}^+)} V_2(z_1, z_2, R) dz_2 dz_1, \tag{3}$$

where V^{-1} denotes the inverse cdf of the normal distribution, with $V_2(\cdot, R)$ being the pdf of the bivariate normal distribution. Here, R denotes the correlation matrix associated with the joint distribution capturing the cross-sectional dependence and is given by:

$$R = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$

where ρ is a dependence parameter Gaussian copula function that describes the cross-sectional dependence between the two count time series. Also, $F_{i,t}^+ = F(y_{it} | y_{i,t-1})$ and $F_{i,t}^- = F(y_{it} - 1 | y_{i,t-1})$, for $i = 1, 2$, where:

$$F(y_{it} | y_{i,t-1}) = \frac{F_{12}(y_{it}, y_{i,t-1}) - F_{12}(y_{it}, y_{i,t-1} - 1)}{f_{t-1}(y_{i,t-1}; \theta)},$$

is the conditional cdf of Y_{it} given $Y_{i,t-1}$, for $i = 1, 2$, and

$$F_{12}(y_{it}, y_{i,t-1}) = C(F_t(y_{it}), F_{t-1}(y_{i,t-1}); \delta),$$

where $C(\cdot, \cdot; \delta)$ is a bivariate copula function with dependence parameter δ , describing the serial dependence in a single time series, and θ denotes the vector of the marginal parameters and reduces to a scalar under the Poisson distribution, i.e., $\theta = \lambda$. This proposed model can be used to analyze bivariate count time series data with counts following any marginal distribution.

4. Inference

Parameter estimation has been conducted by maximizing the likelihood function. The log-likelihood function is constructed using copula theory. However, such a function has no closed form, so its maximization does not follow the standard theory [16]. The maximization technique used is presented next.

Using the conditional density function mentioned in Equation (3) for $t = 1$, the joint distribution of Y_{11} and Y_{21} is given by

$$f(y_{11}, y_{21}) = \int_{V^{-1}(F_{1,1}^-)}^{V^{-1}(F_{1,1}^+)} \int_{V^{-1}(F_{2,1}^-)}^{V^{-1}(F_{2,1}^+)} V_2(z_1, z_2, R) dz_2 dz_1, \tag{4}$$

and for $t = 2, \dots, n$, the conditional bivariate distribution of $Y_{1t} = y_{1t}$ and $Y_{2t} = y_{2t}$ given $Y_{1,t-1} = y_{1,t-1}$ and $Y_{2,t-1} = y_{2,t-1}$ is given by

$$f(y_{1t}, y_{2t} | y_{1,t-1}, y_{2,t-1}) = \int_{V^{-1}(F_{1,t}^-)}^{V^{-1}(F_{1,t}^+)} \int_{V^{-1}(F_{2,t}^-)}^{V^{-1}(F_{2,t}^+)} V_2(z_1, z_2, R) dz_2 dz_1. \tag{5}$$

Hence, combining Equations (4) and (5), the likelihood function is given by

$$L(\boldsymbol{\vartheta}; \mathbf{y}) = f(y_{11}, y_{21}) \cdot \prod_{t=2}^n f(y_{1t}, y_{2t} | y_{1,t-1}, y_{2,t-1}), \tag{6}$$

where $\boldsymbol{\vartheta} = (\boldsymbol{\theta}', \delta_1, \delta_2, \rho)'$; here, $\boldsymbol{\theta}$ is the vector of marginal parameters, and δ_1 and δ_2 are the serial dependence parameters to deal with the first and second time series, respectively. The bivariate dependence between the two time series is captured by ρ . Therefore, taking the log of the function in (6), we can construct the log-likelihood function as follows:

$$\log L(\boldsymbol{\vartheta}; \mathbf{y}) = l(\boldsymbol{\vartheta}; \mathbf{y}) = \log f(y_{11}, y_{21}) + \sum_{t=2}^n \log f(y_{1t}, y_{2t} | y_{1,t-1}, y_{2,t-1}). \tag{7}$$

Maximizing the log-likelihood function in (7) provides ML estimates for the proposed class of model. However, within the log-likelihood function, there exists a bivariate normal integral function that does not have a closed form as shown in (3). Hence, we evaluate the bivariate integral function using the standard randomized importance sampling method presented by Genz and Bretz [17]. This method has been proven to be effective with dimensions less than ten. Hothorn et al. [18] implement this procedure in a package `mvtnorm`, available at CRAN. The package consists of a function, `pmvnorm`, for the computation of multivariate normal probabilities. Then, the parameter estimates, i.e., $\hat{\boldsymbol{\vartheta}}$, can be obtained as

$$\hat{\boldsymbol{\vartheta}} = \arg \max_{\boldsymbol{\vartheta}} l(\boldsymbol{\vartheta}; \mathbf{y}).$$

This maximization technique produces a numerically calculated Hessian matrix that provides the Fisher's information matrix (FIM) as shown by Silva and Diniz [19]. Using the inverse of the FIM yields standard errors of the ML estimates of $\boldsymbol{\vartheta}$. In the next section, we evaluate the effectiveness of the proposed class of models through a comprehensive simulation study.

5. Simulation Studies

A comprehensive simulation study is conducted to evaluate the proposed estimation method and validate the asymptotic properties of the parameter estimates. We first consider the bivariate Poisson count time series data. For each univariate time series, we consider a first-order stationary copula-based Markov model, where a copula family is used for the joint distribution of consecutive observations, and then we couple these two time series using a bivariate copula function at each time point. Here, λ_1 and λ_2 denote the means of two marginal distributions; δ_1 and δ_2 measure the serial dependence within each time series; and ρ measures the cross-correlation between the two time series. The Gaussian copula is selected as the candidate copula family with true parameters ($\lambda_1 = 4, \lambda_2 = 6, \delta_1 = 0.5, \delta_2 = 0.4, \rho = 0.5$). Assuming the process is stationary, the marginal distributions' parameters θ is set to be constant across time. Simulations are performed using sample sizes of 100, 500, and 1500 while replicating them 1000 times. For each of the above five parameter estimates, the standard error (**SE**), mean square error (**MSE**), and mean absolute error (**MAE**) are calculated, and the results are displayed in Table 2. The **SE** is the standard deviation of the estimates over 1000 replications. The **MSE** measures the average squared difference between the estimated values and the actual value, while **MAE** measures the average absolute difference between the estimated values and the actual value. Mathematically, we can define the **SE**, **MSE** and **MAE** as given below:

$$SE = \sqrt{\frac{\sum_{i=1}^m (\theta_i - \hat{\theta}_i)^2}{m - 1}}, \quad MSE = \frac{1}{m} \sum_{i=1}^m (\theta_i - \hat{\theta}_i)^2, \quad MAE = \frac{1}{m} \sum_{i=1}^m |\theta_i - \hat{\theta}_i|,$$

where $\hat{\theta}_i$ is the estimated value of the parameter and m is the number of replications. We conduct another simulation setting using Gaussian copula as the candidate copula family with true parameters ($\lambda_1 = 3, \lambda_2 = 5, \delta_1 = 0.6, \delta_2 = 0.4, \rho = -0.5$). In this simulation setting, we consider the negative cross-correlation between two time series. The corresponding simulation results are displayed in Table 3.

Table 2. Parameter estimates using Gaussian copula for univariate and joint distribution with Poisson marginals for positive cross-correlation.

| Sample Size | Parameter | Estimate | SE | MSE | MAE |
|-------------|-----------------|----------|-------|-------|-------|
| 100 | $\lambda_1(4)$ | 4.092 | 0.376 | 0.149 | 0.309 |
| | $\lambda_2(6)$ | 6.077 | 0.406 | 0.171 | 0.327 |
| | $\delta_1(0.5)$ | 0.428 | 0.067 | 0.009 | 0.08 |
| | $\delta_2(0.4)$ | 0.347 | 0.074 | 0.008 | 0.072 |
| | $\rho(0.5)$ | 0.459 | 0.071 | 0.007 | 0.066 |
| 500 | $\lambda_1(4)$ | 4.098 | 0.175 | 0.04 | 0.162 |
| | $\lambda_2(6)$ | 6.074 | 0.182 | 0.039 | 0.157 |
| | $\delta_1(0.5)$ | 0.437 | 0.031 | 0.005 | 0.063 |
| | $\delta_2(0.4)$ | 0.355 | 0.033 | 0.003 | 0.047 |
| | $\rho(0.5)$ | 0.458 | 0.032 | 0.003 | 0.048 |
| 1500 | $\lambda_1(4)$ | 4.104 | 0.105 | 0.022 | 0.124 |
| | $\lambda_2(6)$ | 6.084 | 0.105 | 0.018 | 0.106 |
| | $\delta_1(0.5)$ | 0.438 | 0.016 | 0.004 | 0.061 |
| | $\delta_2(0.4)$ | 0.358 | 0.018 | 0.002 | 0.042 |
| | $\rho(0.5)$ | 0.453 | 0.019 | 0.002 | 0.046 |

Table 3. Parameter estimates using Gaussian copula for univariate and joint distribution with Poisson marginals for negative cross-correlation.

| Sample Size | Parameter | Estimate | SE | MSE | MAE |
|-------------|-----------------|----------|-------|-------|-------|
| 100 | $\lambda_1(3)$ | 3.132 | 0.398 | 0.176 | 0.335 |
| | $\lambda_2(5)$ | 4.988 | 0.365 | 0.133 | 0.295 |
| | $\delta_1(0.6)$ | 0.501 | 0.062 | 0.014 | 0.101 |
| | $\delta_2(0.4)$ | 0.338 | 0.079 | 0.010 | 0.081 |
| | $\rho(-0.5)$ | -0.448 | 0.072 | 0.008 | 0.072 |
| 500 | $\lambda_1(3)$ | 3.136 | 0.186 | 0.053 | 0.185 |
| | $\lambda_2(5)$ | 4.978 | 0.161 | 0.026 | 0.129 |
| | $\delta_1(0.6)$ | 0.510 | 0.028 | 0.009 | 0.089 |
| | $\delta_2(0.4)$ | 0.351 | 0.033 | 0.003 | 0.051 |
| | $\rho(-0.5)$ | -0.448 | 0.033 | 0.004 | 0.059 |
| 1500 | $\lambda_1(3)$ | 3.141 | 0.111 | 0.032 | 0.152 |
| | $\lambda_2(5)$ | 4.972 | 0.094 | 0.009 | 0.077 |
| | $\delta_1(0.6)$ | 0.513 | 0.015 | 0.007 | 0.086 |
| | $\delta_2(0.4)$ | 0.353 | 0.019 | 0.002 | 0.047 |
| | $\rho(-0.5)$ | -0.441 | 0.018 | 0.004 | 0.055 |

Tables 2 and 3 illustrate that the parameter estimates are converging to true values, and the standard error decreases as the sample size increases. The results show that the estimates become more and more robust as the sample size increase. Figures 1 and 2 show the quantile plots of the estimated parameters. They are approximately normally distributed.

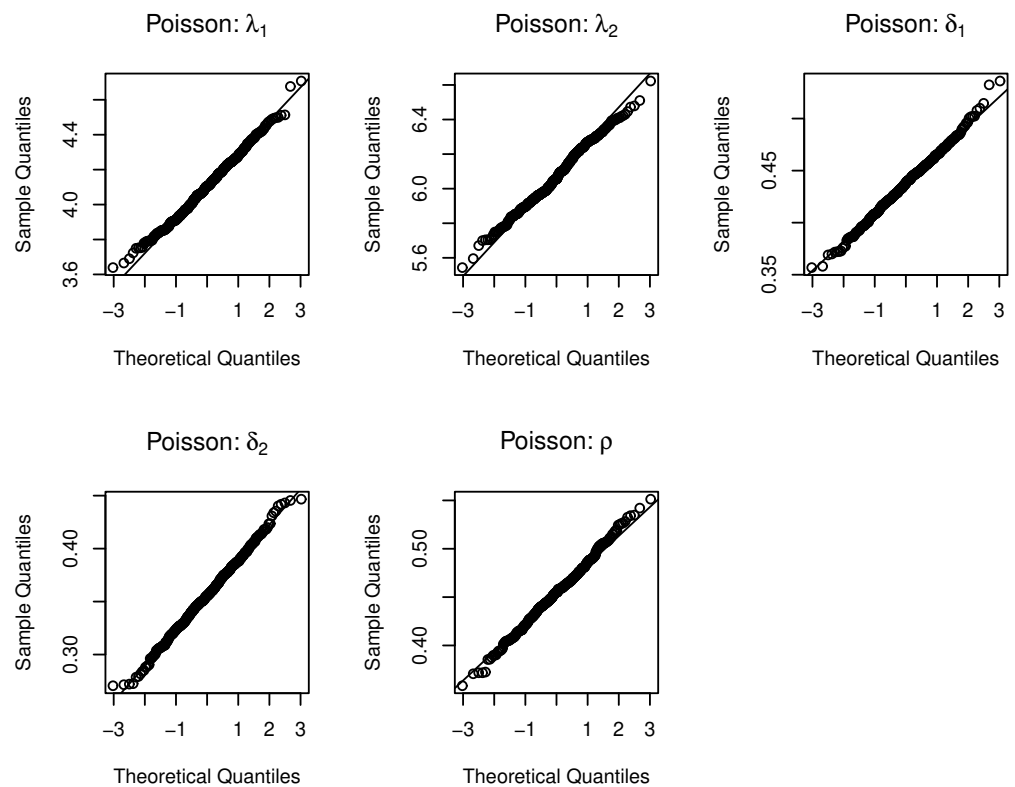


Figure 1. Q-Q plots of the ML estimates for $n = 500$ with positive cross-correlation.

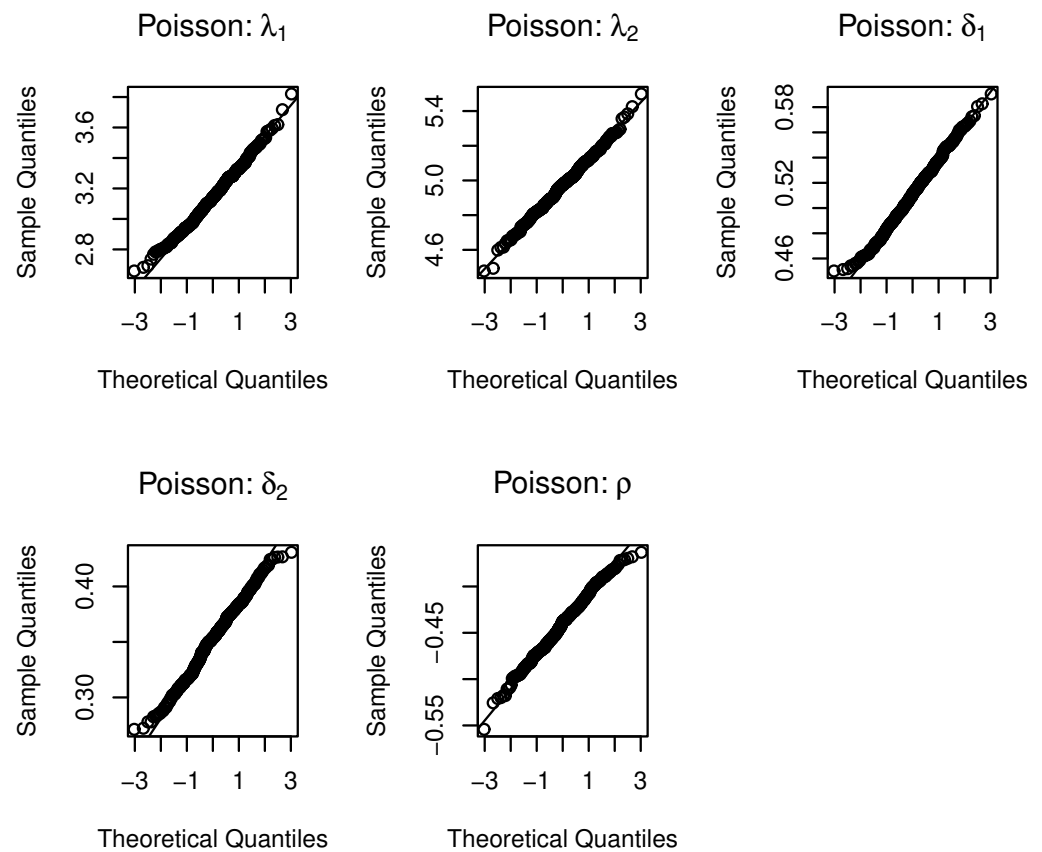


Figure 2. Q-Q plots of the ML estimates for $n = 500$ with negative cross-correlation.

6. Real-Data Application

We apply our proposed bivariate model to analyze bivariate weather data acquired from Mobile Regional Airport Station in Alabama at <https://www.wunderground.com/history/monthly/us/al/mobile/KMOB>, accessed on 6 April 2023. The data consist of average daily temperature and humidity values from January 2022 and ending in February 2023. These averages are computed from daily 15 min interval readings. The daily temperature averages are given on the degree Fahrenheit scale ($^{\circ}\text{F}$), while the daily humidity average is expressed as a percent. Both variables are on a continuous scale and converted to discrete count data. The data conversion is considered to fit counts that are defined by levels ranging from 0 to 5 as shown in Table 4. We categorize the data in this way because we usually define values like that on any given day or time to fall into one of those groups. Erhardt et al. [20] considered a similar transformation of the temperature data into copula data in their paper. Figure 3 shows the daily level of the humidity and temperature for the first three months of 2022, from January to March.

Table 4. Data count converted in scale.

| Interval | Level | Temperature | Humidity |
|------------------|-------|-------------|----------|
| $x < 45$ | 0 | 34 | 12 |
| $45 \leq x < 55$ | 1 | 61 | 31 |
| $55 \leq x < 65$ | 2 | 73 | 69 |
| $65 \leq x < 75$ | 3 | 117 | 111 |
| $75 \leq x < 85$ | 4 | 129 | 126 |
| $x \geq 85$ | 5 | 10 | 75 |

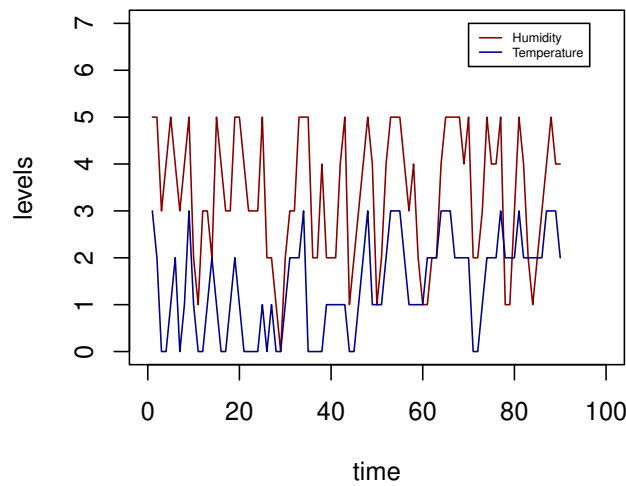


Figure 3. greenPlot of humidity and temperature levels for the first 3 months of 2022.

Figure 4 shows the relationship between temperature and humidity for the weather data. There appears to be a relationship between the temperature and humidity for the given months but, apparently, the relationship is not linear. In Figure 5, we observe that the distributions of temperature and humidity are relatively the same except when the two atmospheric variables are at levels zero and five.

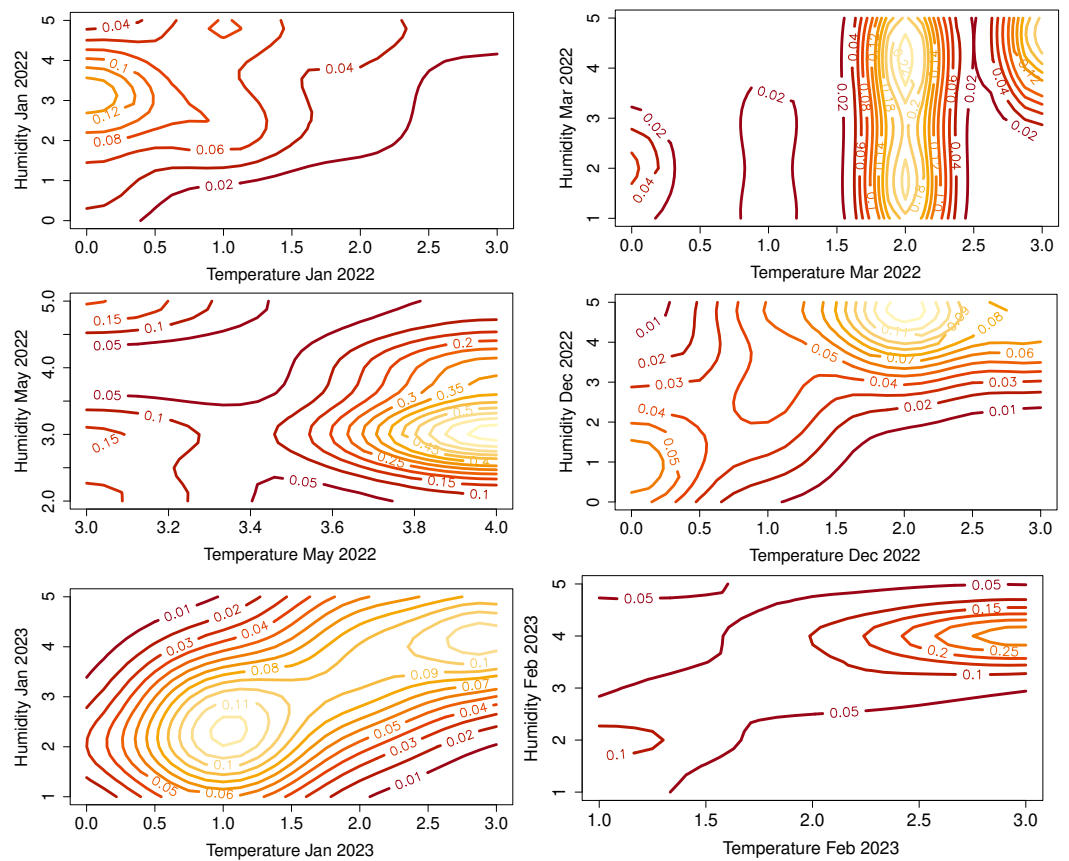


Figure 4. Contour plots of the relationship between temperature and humidity for selected months of 2022 and 2023.

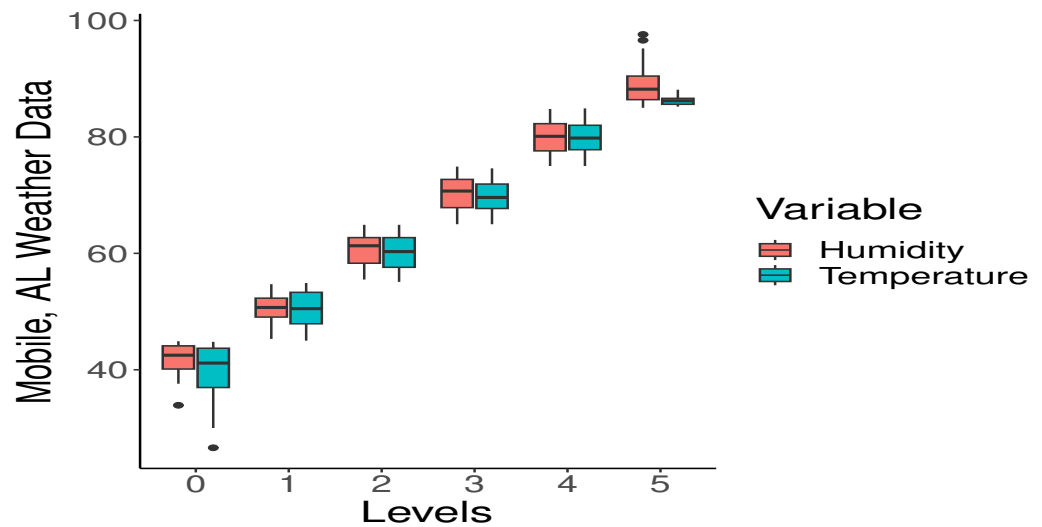


Figure 5. Boxplot of temperature and humidity data.

Table 5 represents the parameter estimates for the fitted univariate model for temperature and humidity, choosing the Gaussian copula as the candidate copula family with a Poisson marginal distribution. Here, λ_1 and λ_2 denote the means of two marginal distributions; δ_1 and δ_2 measure the serial dependence within each time series for temperature and humidity, respectively.

Table 5. Gaussian copula marginal regression model for temperature and humidity.

| Parameter | Estimate | SE |
|-------------|----------|-------|
| λ_1 | 1.020 | 0.118 |
| δ_1 | 0.902 | 0.009 |
| λ_2 | 1.169 | 0.068 |
| δ_2 | 0.745 | 0.021 |

These results suggest that the estimated Poisson mean temperature and humidity are around the 50s in degrees Fahrenheit.

Table 6 represents the parameter estimates for the fitted bivariate model choosing Gaussian copula as the candidate copula family with Poisson marginals.

Table 6. Parameter estimates with Poisson marginals.

| Parameter | Estimate | SE |
|-------------|----------|-------|
| λ_1 | 2.677 | 0.121 |
| λ_2 | 3.293 | 0.067 |
| δ_1 | 0.920 | 0.009 |
| δ_2 | 0.754 | 0.020 |
| ρ | 0.501 | 0.044 |

Table 6 also displays the standard errors associated with the parameter estimates for the bivariate copula model fitted using the Gaussian copula as the chosen copula family. Notably, both marginal and copula parameter estimates exhibit robust standard errors but more reliable parameter estimates. The temperature range around 60 degrees Fahrenheit holds greater relevance than the temperature around the 50s in the city of Mobile, AL. The same can be said for the city’s humidity. Hence, this alternate structure better underscores the joint relationship indicator that was ignored. The inclusion of other

pertinent variables would enhance the climate and ecosystems models, making it more compelling for preparedness efforts.

7. Conclusions

In this manuscript, we propose a bivariate count time series model, which was built using copula theory. The Gaussian copula is used as the candidate copula family to capture serial dependence as well as the cross-correlation between the two time series. The model performs equally well on modeling both positive and negative cross-correlations. Simulated examples are conducted to evaluate the likelihood-based estimation method with importance sampling to evaluate the bivariate normal integral. To prove the effectiveness of the proposed model, bivariate counts of temperature and humidity are analyzed.

Author Contributions: Methodology, D.F., O.A. and N.D.; Software, D.F. and O.A.; Formal analysis, D.F. and O.A.; Investigation, N.D.; Writing—original draft, D.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Acknowledgments: The authors thank the Editor and Reviewers whose comments have significantly improved the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Chaudhari, P.; Rana, D.P.; Mehta, R.G.; Mistry, N.J.; Raghuvanshi, M.M. Discretization of temporal data: A survey. *arXiv* **2014**, arXiv:1402.4283.
- Márquez-Grajales, A.; Acosta-Mesa, H.G.; Mezura-Montes, E.; Graff, M. A multi-breakpoints approach for symbolic discretization of time series. *Knowl. Inf. Syst.* **2020**, *62*, 2795–2834. [[CrossRef](#)]
- Mordvanyuk, N.; López, B.; Bifet, A. TA4L: Efficient temporal abstraction of multivariate time series. *Knowl.-Based Syst.* **2022**, *244*, 108554. [[CrossRef](#)]
- Teuling, A.; Stöckli, R.; Seneviratne, S.I. Bivariate colour maps for visualizing climate data. *Int. J. Climatol.* **2011**, *31*, 1408–1412. [[CrossRef](#)]
- Teuling, A.; Hirschi, M.; Ohmura, A.; Wild, M.; Reichstein, M.; Ciais, P.; Buchmann, N.; Ammann, C.; Montagnani, L.; Richardson, A.; et al. A regional perspective on trends in continental evaporation. *Geophys. Res. Lett.* **2009**, *36*, 2. [[CrossRef](#)]
- Agrawal, S. The Effectiveness of Copulas for Modeling Compound Climate Extreme Events in Boulder County, Colorado. Ph.D. Thesis, UCLA, Los Angeles, CA, USA, 2022.
- Li, W.; Isberg, J.; Chen, W.; Engström, J.; Waters, R.; Svensson, O.; Leijon, M. Bivariate joint distribution modeling of wave climate data using a copula method. *Int. J. Energy Stat.* **2016**, *4*, 1650015. [[CrossRef](#)]
- Lee, T.; Modarres, R.; Ouarda, T.B. Data-based analysis of bivariate copula tail dependence for drought duration and severity. *Hydrol. Process.* **2013**, *27*, 1454–1463. [[CrossRef](#)]
- Li, M.; Zhang, T.; Feng, P. Bivariate frequency analysis of seasonal runoff series under future climate change. *Hydrol. Sci. J.* **2020**, *65*, 2439–2452. [[CrossRef](#)]
- Yin, J.; Guo, S.; He, S.; Guo, J.; Hong, X.; Liu, Z. A copula-based analysis of projected climate changes to bivariate flood quantiles. *J. Hydrol.* **2018**, *566*, 23–42. [[CrossRef](#)]
- Barma, S.D.; Uttarwar, S.B.; Mahesha, A. Probabilistic Assessment of the Interaction between Weather, COVID-19 and Exchange rate of Mumbai City in India using Archimedean Copulas. *Res. Sq.* **2022**. [[CrossRef](#)]
- George, J.; Letha, J.; Jairaj, P. Daily rainfall prediction using generalized linear bivariate model—A case study. *Procedia Technol.* **2016**, *24*, 31–38. [[CrossRef](#)]
- Nelsen, R.B. *An Introduction to Copulas*; Springer: Berlin/Heidelberg, Germany, 2007.
- Joe, H. *Dependence Modeling with Copulas*; CRC press: Boca Raton, FL, USA, 2014.
- Masarotto, G.; Varin, C. Gaussian copula marginal regression. *Electron. J. Statist.* **2012**, *6*, 1517–1549. [[CrossRef](#)]
- Panagiotelis, A.; Czado, C.; Joe, H. Pair copula constructions for multivariate discrete data. *J. Am. Stat. Assoc.* **2012**, *107*, 1063–1072. [[CrossRef](#)]
- Genz, A.; Bretz, F. *Computation of Multivariate Normal and t Probabilities*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2009; Volume 195.
- Hothorn, T.; Bretz, F.; Genz, A. On multivariate t and Gauss probabilities in R. *Sigma* **2001**, *1000*, 3.

19. Silva, A.; Diniz, A. Fisher Information Matrix for Crovelli's Andgamma Beta II Bivariate Distributions. *Rev. Bras. Biom* **2021**, *39*, 350–361. [[CrossRef](#)]
20. Erhardt, T.M.; Czado, C.; Schepsmeier, U. R-vine models for spatial time series with an application to daily mean temperature. *Biometrics* **2015**, *71*, 323–332. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.