University of Dundee

Modeling bivariate geyser eruption system with covariate-adjusted recurrent event process

Jin, Zhongnan; Lu, Lu; Bedair, Khaled; Hong, Yili

[Link to publication in Discovery Research Portal](#)

Modeling Bivariate Geyser Eruption System with Covariate-Adjusted Recurrent Event Process

Zhongnan Jin[1], Lu Lu[2], Khaled Bedair[3,4], and Yili Hong[1]

[1]Department of Statistics, Virginia Tech, Blacksburg, VA, USA

[2]Department of Mathematics & Statistics, University of South Florida, Tampa, FL, USA

[3]Department of Statistics & Mathematics, Faculty of Commerce, Tanta University, Egypt

[4]School of Medicine, University of Dundee, Dundee, DD1 9SY, UK

Abstract

Geyser eruption is one of the most popular signature attractions at the Yellowstone National Park. The interdependence of geyser eruptions and impacts of covariates are of interest to researchers in geyser studies. In this paper, we propose a parametric covariate-adjusted recurrent event model for estimating the eruption gap time. We describe a general bivariate recurrent event process, where a bivariate lognormal distribution and a Gumbel copula with different marginal distributions are used to model an interdependent dual-type event system. The maximum likelihood approach is used to estimate model parameters. The proposed method is applied to analyzing the Yellowstone geyser eruption data for a bivariate geyser system and offers a deeper understanding of the event occurrence mechanism of individual events as well as the system as a whole. A comprehensive simulation study is conducted to evaluate the performance of the proposed method.

Key Words: Competing risks; Copula; Event dependence; Gap time; Recurrent events; Yellowstone National Park.

1 Introduction

1.1 Background

Geyser eruption is one of the signature attractions at the Yellowstone National Park, which is home to two-thirds of the worlds' geysers. Tourists around the world crave to witness this fascinating natural phenomenon. Many researchers are interested in studying geyser eruptions and the underlying mechanisms. Fournier (1969) built a physical model to describe the time interval between eruptions for the Old Faithful geyser, which is one of the most famous geysers in the Yellowstone National Park. Rinehart (1972) showed that the Old Faithful Geyser activities are affected by earth tidal forces, barometric pressure, and tectonic stresses. However, geyser eruptions have not been studied by statistical methods. This paper develops a statistical model for analyzing geyser eruption data from a bivariate geyser system. This work will benefit the geyser study community for understanding and effectively modeling geyser eruption activities.

Typical geyser eruption is a repeating process and hence can be modeled with a recurrent process for events repeatedly occurring over time. Recurrent processes have had broad applications in diverse areas. For example, they have been widely used for studying vehicle failures in warranty studies (Lawless, 1995), relapse biomarkers in cancer research (Schaubel and Cai, 2004), and sports injury analysis (Ullah et al., 2014). Typically the time interval between two consecutive events, which is also referred to as the gap time, is studied to model the event frequency in a recurrent process. The proportional intensity models (Cox, 1972, and Andersen and Gill, 1982) are popular for modeling event occurrences of a single type of event. In more sophisticated studies, there are multiple types of recurrent events observed in a single system. The occurrence of any type of event will result in a system event. In addition, in a multi-type recurrent event process, the gap time for different event types could be correlated. For example, the occurrence of one type of event could cause other types of events to occur more frequently. In this case, a multivariate recurrent process should be considered to model the interdependence of multiple event types in the same system.

In many event analyses, covariates are found to be useful for modeling the event occurrence time and frequency. Many recurrent processes are affected by process conditions. For example, some mechanical failures could occur at a higher frequency under a higher temperature, humidity, or pressure. Incorporating covariates into the recurrent process models could improve the model performance and provide a more precise estimation of the event time and frequency. Models of this type are referred to as the covariate-adjusted recurrent event process models.

In this paper, we focus on modeling and analysis of geyser eruptions for a two geyser system in the Yellowstone National Park. In a multi-type recurrent event system, the system events can result from either type of events, and hence any consecutive events could be associated with the same or different event types. In order to describe this bivariate recurrent event process, we need to not only understand the marginal behavior of each type of event, but also understand the interdependence between the two types of events. We consider a bivariate distribution for the gap times between successive events for a bivariate event system. To improve the estimation, we leverage the covariate information on the eruption duration and develop a covariate-adjusted bivariate recurrent event process model for estimating the eruption gap time of a two geyser system formed by West Triplet Geyser and Grotto Geyser in the Yellowstone National Park.

1.2   Related Literature

Recurrent event processes are extensively studied in the areas of reliability, public health, and medical studies. The nonparametric estimation of gap time distribution based on multivariate failure time data was introduced in Schaubel and Cai (2004). Dauxois and Sencey (2009) considered the risks of two nosocomial infections for patients admitted to hospitals. Bouaziz et al. (2013) provided a nonparametric method to estimate the intensity function of a recurrent process. Other than hazard functions and intensity functions, survival status in time is often of interest as well. Huang and Liu (2007) investigated the disease free survival rate in a recurrent heart failure study. Zeng and Lin (2009) and Garre et al. (2008) focused on studying the terminal events in recurrent systems. Meyer and Romeo (2015) presented Bayesian analysis of recurrent event using copulas. An earlier review regarding recurrent events can be found in Lawless (1995) and a review on modeling of repairable systems can be found in Lindqvist (2006). Classical books on recurrent event data analysis include Daley and Vere-Jones (2003), Cook and Lawless (2007), and Duchateau and Janssen (2008).

In medical research, Liu and Huang (2009) presented repeated measurements of biomarker to determine the HIV survival status in a recurrent event system. Sun et al. (2006) applied covariate-adjusted additive hazard model for the data, which is involving recurrent gap times. Prasad and Rao (2002) used a proportional hazard function with covariate adjustment in a repairable system. In another application of recurrent event data, Huzurbazar and Williams (2010) incorporated covariates in a flowgraph model. Yang et al. (2013) introduced multivariate lognormal assumption on event gap times of different event types. Yang et al. (2017) considered a parametric model for the multi-type event recurrent event data without

covariates and developed copula function on gap times for the recurrent process in a car body manufacturing process.

Existing research does not consider correlated renewal process with covariate adjustments. Motivated by the geyser data, we propose the CARP model to study the eruption gap time for a two-geyser system. Application-wise, geyser eruption is rarely studied by statistical models. For the geyser eruption study, the modeling and analysis presented in this paper are new to geyser research.

1.3   Overview

The rest of this paper is organized as follows. Section 2 provides more details on the geyser eruption data from the Yellowstone National Park. Section 3 discusses the model formulation for the multi-type recurrent event system. Section 4 describes the maximum likelihood approach for estimating the model parameters. A simulation study is described in Section 5 to evaluate the proposed method with model comparisons made under different parameter settings. The modeling and analysis of the Yellowstone geyser data are detailed in Section 6. Section 7 contains some concluding remarks.

2   Geyser Data

We use the publicly available Yellowstone geyser eruption data, which were collected in 2008 by the Geyser Observation and Study Association (GOSA). By using underground sensors, water levels were measured continuously, and occurrences of geyser eruptions were detected automatically. For each geyser, the GOSA data include the starting time and duration of each eruption. We choose to analyze the data from the West Triplet Geyser and the Grotto Geyser during the study period between June 2008 and November 2008. This particular dataset and study period were chosen to ensure that completely uninterrupted recurrent event data are available over a relatively long time span to allow for the dependence modeling for the two geysers.

We merge their eruption records in a temporal order as illustrated in Table 1. The data include the date and time when an eruption occurred, the eruption duration, and which geyser had the eruption. During the study period, the West Triplet Geyser erupted more frequently than the Grotto Geyser. Also, the eruptions of the West Triplet Geyser also lasted longer than those of the Grotto Geyser on average. More specifically, the average eruption gap time for the West Triplet Geyser is 6.8 hours with a standard deviation of 2.8 hours, while

the Grotto Geyser had an average eruption gap time of 9.3 hours with a standard deviation of 8.6 hours. For the eruption duration, the Grotto Geyser eruptions lasted generally longer than the West Triplet Geyser, with the average duration of the Grotto Geyser eruptions being 3.5 hours with a standard deviation of 5.2 hours and the eruption duration of the West Triplet Geyser averaged at 0.7 hours with a 0.5 hours standard deviation.

Figures 1a and 1b display the side-by-side boxplots of the time between eruptions and the duration of eruptions, respectively, for the West Triplet and Grotto geysers. We can see compared with the West Triplet Geyser, the Grotto Geyser has a much larger variation of the eruption gap time with a number of extremely long gaps between eruptions and also more variation in the eruption frequency. On the other hand, the West Triplet Geyser has considerably shorter and less variable eruption durations than the Grotto Geyser. In addition, Figure 2 shows the plots of the time between eruptions versus the duration time of the previous eruption for both geysers. We can observe a moderate correlation between these two variables for the West Triplet Geyser and a strong correlation for the Grotto geyser. Hence, we decided to utilize the duration of the previous eruption for both geysers to help model the eruption gap time.

Figure 3 illustrates the data obtained for the two-geyser system. In this figure, W_{kj} denotes the gap time of the kth eruption (i.e., the time interval between the $(k-1)$th and kth eruptions) for geyser j where $j = 1$ for the West Triplet Geyser and $j = 2$ for the Grotto Geyser. The covariate x_i denotes the eruption duration for the ith eruption in the bivariate geyser system, where $i = 1, \cdots, n$, $n = n_1 + n_2$ is the total number of eruptions for both geysers, and n_j denotes the number of eruptions for the jth geyser. In particular, $n_1 = 580$, $n_2 = 421$, and $n = 1001$ for the dataset we analyzed. Note here the time between eruptions are labeled separately for each individual geyser. To model the bivariate geyser system, we will introduce new notation for jointly describing the event time, event type and covariate information in Section 3.

3 Data Setup and Model

3.1 Data Setup

Suppose that in a bivariate recurrent process with n total events, the systematic event time is described by the variable T_i for $i = 1, \cdots, n$. We use T_0 to denote the starting time or the system installation time, and T_n denotes the last event time in the system. In a bivariate system, there are two types of events, and hence we use an indicator variable $\Delta_i \in \{1, 2\}$ to

Table 1: Sample observations from the West Triplet and Grotto Geysers.

Eruption time	Duration (hours)	Geyser
2008-06-20 16:58:00	0.93	Grotto
2008-06-20 20:46:00	0.75	West Triplet
2008-06-20 21:31:00	2.05	Grotto
2008-06-21 02:51:00	1.08	West Triplet
2008-06-21 04:48:00	2.63	Grotto
2008-06-21 11:15:00	1.72	Grotto
2008-06-21 13:02:00	0.48	West Triplet
2008-06-21 17:56:00	3.58	Grotto
2008-06-21 18:20:00	0.73	West Triplet
2008-06-22 00:11:00	0.93	Grotto
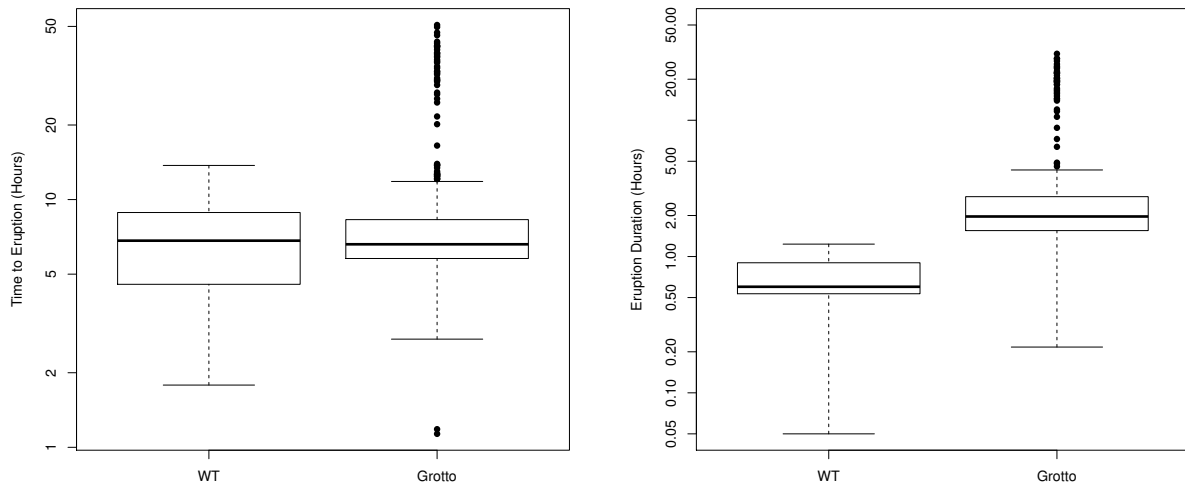
(a) Time to eruption (Hour)

(b) Eruption duration (Hour)

Figure 1: Boxplots for the time to eruption and the eruption duration for the West Triplet Geyser and Grotto Geyser from June to November in 2008. Note that the $y-$axis is on log scale.
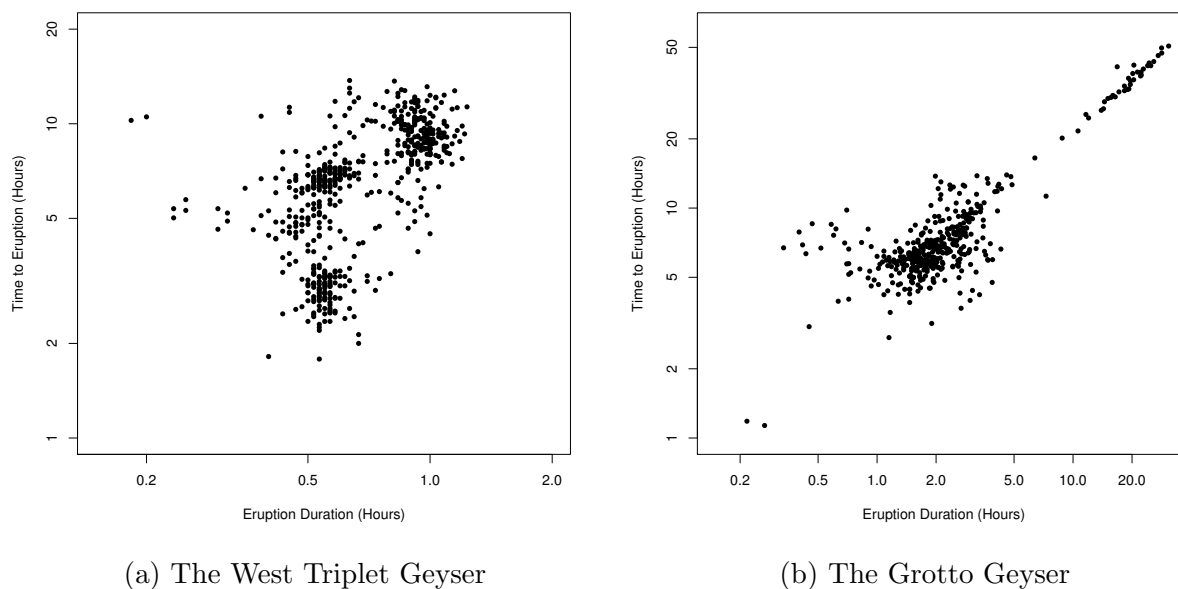
(a) The West Triplet Geyser

(b) The Grotto Geyser

Figure 2: The plot of the time between eruptions versus the eruption duration for (a) the West Triplet Geyser and (b) the Grotto Geyser from June to November in 2008. Note that both the $x-$axis and the $y-$axis are on log scale.
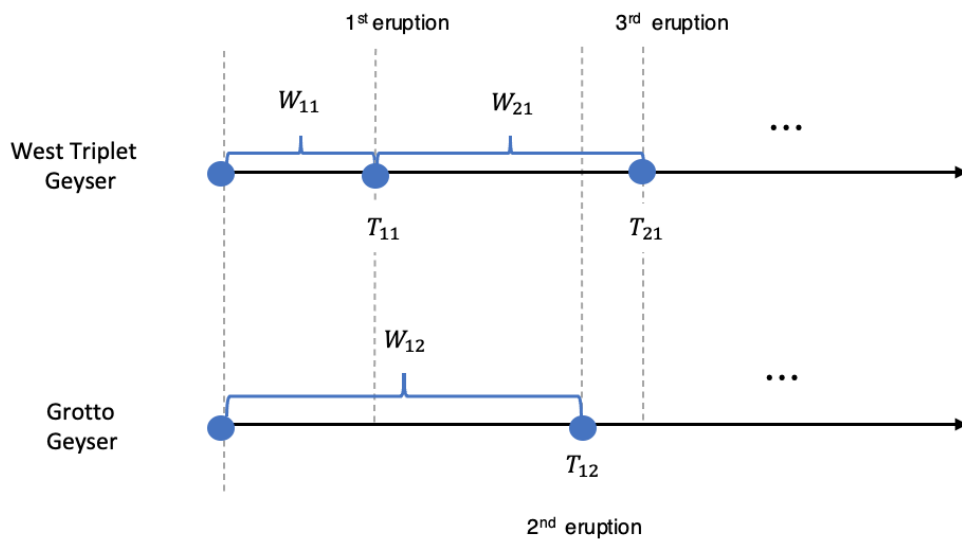

Figure 3: Illustration of geyser eruptions with West Triplet and Grotto Geysers.

represent the type of event. For an event that occurs at time t_i, the covariate vector on the event duration is denoted by $\boldsymbol{X}_i$, which is a vector of the previous duration for both types of events. Therefore, each event can be represented by $\{T_i, \Delta_i, \boldsymbol{X}_i\}$, where $i = 1, \cdots, n$. We use $\{t_i, \delta_i, \boldsymbol{x}_i\}$ to denote the observations of the ith event which occurs at time t_i and it is from event type δ_i with covariates measured as $\boldsymbol{x}_i$. Note in an observed recurrent process with n total events, the last event is observed at time t_n, where $t_n \leq \kappa$ with κ being the pre-defined study termination time.

This sequence of bivariate events can be expressed as a counting process $\{N(t) : t \geq 0\}$, where $N(t)$ denotes the cumulative number of events at time t regardless of the event type. Similarly, we define the counting process for each individual event type j as $\{N_j(t) : t \geq 0\}$, where $j = 1$ or 2. In addition, we denote the event history of the system up to a time point $s \leq \kappa$ by $\mathcal{H}_s = \{N(t) : t \leq s\}$. Similarly, the covariate history can be expressed as $\mathcal{X}_s = \{\boldsymbol{x}_t : t \leq s\}$. For further discussion, we use $\mathcal{F}_s = \{\mathcal{H}_s, \mathcal{X}_s\}$ to denote the history including both event and covariate information.

For a bivariate recurrent process system, event time variable T_i defined above satisfies

$$0 = T_0 < T_1 < \cdots < T_i < \cdots < T_n \leq \kappa.$$

Similarly, for events of type j, the event time variables are defined as T_{lj}, where $l = 1, \ldots, n_j$ and n_j is the number of events from type j. In a bivariate recurrent system, we have the relation $n_1 + n_2 = n$. The time variable T_{lj} is also presented in temporal order. For the events of type j, $j = 1$ or 2, we have

$$0 = T_{0j} < T_{1j} < \cdots < T_{lj} < \cdots < T_{n_j,j} \leq \kappa.$$

Based on the ordered event times, the event gap time for the jth type as the interval between two consecutive events can be calculated as $W_{lj} = T_{l+1,j} - T_{lj}$, where $l = 0, 1, \ldots, n_j - 1$.

Now consider the joint bivariate recurrent system. For the ith event in the system, the two-dimensional event gap time variable is defined as

$$\boldsymbol{W}_i = (W_{l_{i1},1}, W_{l_{i2},2})',$$

where $l_{ij} = N_j(t_i)$ is the cumulative number of events for jth event type by time t_i. For example, at the system installation time where $t = 0$, the event gap time vector is $\boldsymbol{W}_0 = (W_{01}, W_{02})'$. If the first event occurs at time t_1 and is from event type 1, then the event gap time vector is denoted as $\boldsymbol{W}_1 = (W_{11}, W_{02})'$. In this case, $l_{11} = 1$ since one event of type 1

has occurred as of time t_1, while $l_{12} = 0$ as no event of type 2 has occurred by that time point.

To link the event gap time $\boldsymbol{W}_i$ with the system event time T_i, we introduce an age variable at time t which is defined as the time between the time point t and the time of the latest event of each event type. In a bivariate recurrent process, the age variable is a vector of two components denoted as

$$\boldsymbol{A}_i = [A_1(T_i), A_2(T_i)]',$$

where $A_j(t) = t - T_{N_j(t),j}$ for $j = 1$ and 2. For an observed event at time t_i, the observed age vector is denoted as $\boldsymbol{a}_i = [a_1(t_i), a_2(t_i)]'$, where $a_j(t) = t - t_{N_j(t),j}$. For example, at the system installation time at $t = 0$, the age vector is $\boldsymbol{a}_0 = (0,0)'$. If the first event occurs from type 1 at time t_1, then $\boldsymbol{a}_1 = (0, t_1)'$. The age variable then connects between the system event time and the event gap time $\boldsymbol{W}_i$ through the relation $\boldsymbol{W}_i \geq \boldsymbol{A}_i$. Here, the vector comparison is defined to be an element-wise comparison. In other words, $\boldsymbol{W}_i \geq \boldsymbol{A}_i$ suggests $W_{l_{ij},j} \geq A_j(T_i)$ for both $j = 1$ and 2.

3.2 Model

In this section, we introduce the proposed covariate-adjusted recurrent process (CARP). After an event occurred at time t_i (or after the system installation at time $t_0 = 0$), the distribution of the event gap time is given by

$$\boldsymbol{W}_i | \mathcal{F}_{t_i} \sim F_{\boldsymbol{W}}(\boldsymbol{v} | \boldsymbol{a}_i, \boldsymbol{x}_i), \quad i = 0, 1, 2, \cdots, n. \tag{1}$$

Here, $\boldsymbol{W} = (W_1, W_2)'$, $\boldsymbol{v} = (v_1, v_2)'$, and

$$F_{\boldsymbol{W}}(\boldsymbol{v} | \boldsymbol{a}_i, \boldsymbol{x}_i) = \Pr(\boldsymbol{W}_i \leq \boldsymbol{v} | \boldsymbol{W}_i \geq \boldsymbol{a}_i, \boldsymbol{x}_i) \tag{2}$$

is the joint cumulative distribution function (cdf) of the event gap time conditioned on the age and covariates. At any event time, the information we have about the event type that has not yet occurred is captured through its age and the conditional probability that is conditioned on the event eruption time is greater than or equal to the age since the last eruption. At each event time, the age is set to be zero for the occurred event type. The event gap time variable $\boldsymbol{W}_i$ is adjusted by covariates $\boldsymbol{x}_i$ which is further discussed in Section 3.3 in more detail.

3.3   Dependence Modeling and Covariate Adjustment

Dependence between events from different types in a bivariate system is modeled by implementing distributional assumptions on variable $\boldsymbol{W}_i = (W_{i1}, W_{i2})'$. In this paper, we use a bivariate lognormal distribution and a copula function to model the random vector $\boldsymbol{W}_i$, where in both models, covariates $\boldsymbol{x}_i$ are used for adjustment. We refer to the CARP models under these assumptions as the CARP-MLN and CARP-copula models, respectively. Here MLN is short for multivariate lognormal.

The CARP-MLN Model

For the multivariate lognormal distribution,

$$\boldsymbol{W}_i \sim \text{MLN}[\boldsymbol{\mu}(\boldsymbol{x}_i), \boldsymbol{\Sigma}],$$

where the location parameter in the bivariate lognormal distribution is expressed as a linear form of covariates $\boldsymbol{x}_i$. That is,

$$\boldsymbol{\mu}(\boldsymbol{x}_i) = \boldsymbol{\mu}_0 + \mathbf{B}\boldsymbol{x}_i. \tag{3}$$

In the linear expression above, $\boldsymbol{\mu}_0$ is a vector of baseline location parameters and $\mathbf{B}$ is a 2 $\times$ 2 coefficient matrix. In the bivariate lognormal assumption, we use a covariance matrix $\boldsymbol{\Sigma}$ to capture the event dependence between events from two event types. Specifically, the diagonal elements in $\boldsymbol{\Sigma}$ represent marginal variances while the off diagonal elements stand for the covariances. When using the bivariate lognormal distribution, the covariance matrix is defined as $\boldsymbol{\Sigma} = \mathbf{CC}'$ to ensure $\boldsymbol{\Sigma}$ to be positive definite, where

$$\mathbf{C} = \begin{pmatrix} \sigma_1 & 0 \\ \eta & \sigma_2 \end{pmatrix}. \tag{4}$$

The correlation ρ is determined by σ_2 and η as

$$\rho = \frac{\eta}{\sqrt{\sigma_2^2 + \eta^2}}.$$

Note the above covariance matrix offers great flexibility to model different correlation relationships of varied size and direction. The sign of the η value determines if the two types of events have a positive or negative correlation. In a bivariate lognormal distribution, the marginal distribution of each dimension also follows a lognormal distribution. Therefore,

when the observed marginal distributions do not seem to follow the lognormal distributions, or their dependency cannot be characterized by the covariance matrix $\mathbf{\Sigma}$, the assumed bivariate lognormal distribution is not appropriate. In this case, the alternative strategy is to define W_1 and W_2 by separate distributions and combine them through a more flexible copula function. This is referred to as the CARP-copula model, which will be introduced in the next section.

The CARP-copula Model

Let F_1 and F_2 denote the marginal cdfs of W_1 and W_2. There always exists a copula function C such that the joint cdf of the two dimensional variable $(W_1, W_2)'$ can be written as

$$F(v_1, v_2) = C[F_1(v_1), F_2(v_2)], \tag{5}$$

where for any unitary uniform variable $U_j, j = 1, 2$, the bivariate copula is defined as

$$C(u_1, u_2) = \Pr(U_1 \leq u_1, U_2 \leq u_2).$$

This is also known as the Sklar's Theorem. By using a copula function, we have the flexibility to choose marginal distributions separately for each event type. For instance, a Gamma distribution and a Weibull distribution can be used as marginal distributions for the two types of events, respectively. With selected F_1 and F_2, one can combine the marginal distributions by using different copula functions.

Note the copula approach allows us to model the marginal distributions separately and then combine them through an appropriate copula function for modeling the dependence structure. For the geyser system, we choose to use parametric distributions for modeling the marginal distributions. However, the method can be easily generalized to using nonparametric methods for modeling the marginal distributions and hence offers great flexibility to be adapted for broad applications.

In the literature, a variety of copula functions has been introduced to capture different dependence patterns among the marginal distributions. A Gaussian copula uses a multivariate normal distribution of transformed marginal distributions based on the inverse cdf of the standard normal distribution. The Archimedean copulas are an associative class of copula functions that model the multivariate dependence through a single parameter. The power variance function copulas including Clayton, Gumbel and Inverse Gaussian are among the most popular ones that are flexible for modeling various dependence structures (e.g., Romeo et al., 2018). In this paper, we use the Gumbel copula function which is popular for modeling

stronger dependence in the positive tail. We refer to the CARP model with the use of Gumbel copula as CARP-copula model for the rest of the paper. In fact, the CARP-MLN model is a special case of the CARP-copula model, where in CARP-MLN, the Gaussian copula is applied and lognormal marginal distributions are selected.

The CARP-MLN model characterizes the dependence among event types through the covariance matrix $\boldsymbol{\Sigma}$, while the CARP-copula model quantifies dependence using the copula parameter. Particularly, in the Gumbel copula model, the parameter is denoted as α. As a result, in the CARP-MLN and CARP-copula models, we have different parameters to characterize event dependence. In order to compare dependence from different models, we introduce the Kendall's tau.

In the CARP-copula model, the Kendall's tau is expressed as $\tau = 1 - 1/\alpha$, where α is the copula coefficient in the Gumbel copula. In the CARP-MLN model, the Kendall's tau is calculated as $\tau = (2/\pi)\arcsin(\eta/\sqrt{\sigma_2^2 + \eta^2})$, where η and σ_2 can be found in (4).

Similar to the CARP-MLN model, for the CARP-copula model, we also use a linear form of the covariates $\boldsymbol{x}_i$ to represent location parameters in marginal distributions as in (3). In the CARP-MLN model, we use a two dimensional vector $\boldsymbol{\mu}(\boldsymbol{x}_i) = [\mu_1(\boldsymbol{x}_i), \mu_2(\boldsymbol{x}_i)]'$ to represent the location parameter of the lognormal distribution. While in the CARP-copula, $\mu_1(\boldsymbol{x}_i)$ and $\mu_2(\boldsymbol{x}_i)$ stand for the location parameters for the first and second marginal distributions, respectively.

3.4   Properties of CARP

For event gap time variable $\boldsymbol{W}_i$, we define the survival function (sf), cdf and hazard function as follows. We denote the joint sf of $\boldsymbol{W}_i$ as,

$$S(\boldsymbol{v}) = \Pr(W_{i1} > v_1, W_{i2} > v_2). \tag{6}$$

The joint cdf is

$$F_{\boldsymbol{W}}(\boldsymbol{v}) = \Pr(W_{i1} \leq v_1, W_{i2} \leq v_2),$$

and the corresponding joint probability density function (pdf) is denoted by $f_{\boldsymbol{W}}(\boldsymbol{v})$. According to (2), the joint pdf of the event gap time variable given all historical events $\boldsymbol{W}_i|\mathcal{F}_{t_i}$ is given by

$$f_{\boldsymbol{W}}(\boldsymbol{v}|\boldsymbol{a}_i, \boldsymbol{x}_i) = \frac{f_{\boldsymbol{W}}[a_1(c_1), a_2(c_2)]}{S(\boldsymbol{a}_i)}, \quad v_j > a_j(t_i), \quad j = 1, 2, \tag{7}$$

where $c_j = t_{l_{ij},j} + W_j, j = 1, 2$. The denominator in (7) takes age condition into account, while the numerator builds the relationship among the event gap time W_j, the event time $t_{l_{ij},j}$ and the age $a_j(c_j)$. Covariates $\boldsymbol{x}_i$ are used to adjust the gap time $\boldsymbol{W}_i$ as discussed in Section 3.3.

For further discussions, let $\mathcal{F}_{t^-}$ be the event history up to time t. Note that for event type j, $T_{N_j(t^-),j}$ gives the most recent event time by time t, and the age variable prior to time t is denoted as $A_j^-(t) = t - T_{N_j(t^-),j}$, which calculates the cumulative running time upon time t since the last event. We use $\boldsymbol{a}^-(t) = [a_1^-(t), a_2^-(t)]'$ to denote the age vector for the two event processes prior to time t. Hence, prior to an event time t_i, the age vector is denoted as $\boldsymbol{a}_i^- = \boldsymbol{a}^-(t_i)$. For example, at the initial time 0, the vector is $\boldsymbol{a}_0^- = (0,0)'$. If the event type is $\delta_1 = 1$ at t_1, then $\boldsymbol{a}_1^- = (t_1, t_1)'$. Note that $\boldsymbol{a}_1$ updates the age to be zero for the corresponding event type at an event time, while $\boldsymbol{a}_1^-$ does not. This notation is used to derive the likelihood in Section 4, is also used to define the hazard function below.

In the literature, the sub-intensity function (i.e., cause-specific event intensity function) is often used to characterize an event process. In particular, the sub-intensity function for event type j is defined as

$$h_j(t) = \lim_{\Delta t \to 0} \frac{\Pr[T \in (t, t + \Delta t), \Delta = \delta_j | \mathcal{F}_{t^-}]}{\Delta t}, \tag{8}$$

where T is the event time, and Δ is the event type. The sub-cumulative intensity function is $H_j(t) = \int_0^t h_j(s)ds$. The hazard function and cumulative hazard function for the system are calculated as the sum of corresponding functions for the two event types,

$$h(t) = \sum_{j=1}^2 h_j(t) \quad \text{and} \quad H(t) = \sum_{j=1}^2 H_j(t).$$

The sub-intensity function in (8) is calculated as

$$h_j(t) = \frac{D_j[\boldsymbol{a}^-(t)]}{S[\boldsymbol{a}^-(t)]}, \quad \text{where} \quad D_j(\widetilde{\boldsymbol{v}}) = -\left.\frac{\partial S(\boldsymbol{v})}{\partial v_j}\right|_{\boldsymbol{v}=\widetilde{\boldsymbol{v}}}, \tag{9}$$

and $\widetilde{\boldsymbol{v}} = (\widetilde{v}_1, \widetilde{v}_2)'$ is a vector with two components. More detail about the calculation of $D_j(\widetilde{\boldsymbol{v}})$ under different models will be discussed in Section 4.

4   Parameter Estimation

The maximum likelihood (ML) approach is used to estimate model parameters. Parameters in the model include parameters in the joint distribution function, the copula function and the linear covariate transformation function. Given all the event history $\mathcal{F}_\tau$, the likelihood function is constructed as follows:

$$L(\boldsymbol{\theta}|\mathcal{F}_\tau) = \prod_{i=1}^{n} L_i(\boldsymbol{\theta}|\boldsymbol{a}_i, \boldsymbol{x}_i), \tag{10}$$

where

$$L_i(\boldsymbol{\theta}|\boldsymbol{a}_i, \boldsymbol{x}_i) = \Pr[T_i \in (t_i, t_i + \Delta t), \Delta_i = \delta_i | \mathcal{F}_{t_{i-1}}], \quad \text{for} \quad i = 1, \cdots, n. \tag{11}$$

Let $\boldsymbol{\theta}$ denote the vector of all the parameters in the model. The estimated parameters $\widehat{\boldsymbol{\theta}}$ are asymptotically normally distributed based on the large sample ML theory (Casella and Berger, 2002). The calculation of the likelihood L_i for the proposed CARP models is shown below.

For any observed recurrent process with n total events, the likelihood contribution for $i = 1, \cdots, n$ in (11) is given by

$$L_i(\boldsymbol{\theta}|\boldsymbol{a}_i, \boldsymbol{x}_i) = \frac{D_{\delta_i}(\boldsymbol{a}_i^-)}{S(\boldsymbol{a}_{i-1})}. \tag{12}$$

In (12), the quantity $D_{\delta_i}(\boldsymbol{a}_i^-)$, which is introduced in (9), is the partial derivative of the bivariate sf $S(\boldsymbol{a}_i^-)$. The covariates $\boldsymbol{x}_i$ are used to adjust the distribution of $\boldsymbol{W}_i$. Likelihood calculations so far are the same for both the CARP-MLN and CARP-copula models. However, we need different ways to calculate $D_j(\boldsymbol{a}_i^-)$ for the two CARP models based on how the survival functions are calculated, which are detailed below.

For the CARP-MLN model, the sf is calculated in a closed form as discussed in (6), and the partial derivative with regard to the jth event type can be written as

$$D_j(\boldsymbol{a}_i^-) = -\frac{\partial S(v_j, v_{j'})}{\partial v_j}\bigg|_{\boldsymbol{v}=\boldsymbol{a}_i^-} = f_j[a_j^-(t_i)] \times \Pr[W_{j'} \geq a_{j'}^-(t_i); j' \neq j | W_j = a_j^-(t_i)], \tag{13}$$

where $f_j[a_j^-(t_i)]$ is the jth marginal density function from a bivariate lognormal distribution. In (13), the conditional probability can be calculated from the conditional normal distribution

with a logarithm transformation. Calculation details can be found in Appendix A.1.

For the CARP-copula model, the likelihood function is calculated based on the relationship between the bivariate sf $S(\boldsymbol{a}_i^-)$ and the cdf $F(\boldsymbol{a}_i^-)$. In bivariate cases, the sf and cdf have the following relationship:

$$S[a_1(t_i), a_2(t_i)] = 1 - F[a_1(t_i), \infty] - F[\infty, a_2(t_i)] + F[a_1(t_i), a_2(t_i)],$$

where the joint cdf can be calculated by the copula as in (5). The calculated likelihood will vary for different choices of the marginal distribution and the copula function. In our case, we chose the Gumbel copula as described in Section 3.3. The ML estimates $\widehat{\boldsymbol{\theta}}$ are obtained by maximizing the likelihood function in (10).

5 Simulation Study

This section describes the simulation study we conducted for evaluating the proposed method. We simulated bivariate recurrent system data based on different models and parameter values. The performance of the proposed method was then evaluated based on the simulated data. Different parameter values were considered in the simulation for both the CARP-MLN and the CARP-copula models to understand the impact of the model parameters. We evaluated the model goodness of fit by calculating the average AIC and the mean squared error (MSE) of the estimated model parameters. The goal is to demonstrate the performance of the proposed method and the improvement by using the covariate adjustment.

5.1 Simulation Setting

We used both the CARP-MLN and CARP-copula models to simulate the data. For each generated data set, both models were fitted to the data and results were evaluated. The model used to generate data is referred to as the true model, while the models used to fit the data are referred to as the fitted models. We explored changing the sample sizes (n), the Kendall's tau value, the scale parameters (σ_j, $j = 1, 2$), and the linear transformation matrix (**B**) to understand their impacts on the analysis.

First, to understand the impact of sample size, we varied the sample size at $n =$200, 500, 1000 and 2000 when generating the data using each model. For the CARP-copula model, the lognormal marginal distributions were used for both event types along with the Gumbel copula function. The location and scale parameters for lognormal marginal distributions in the true model were specified at $(\mu_1 = 1, \sigma_1 = 0.25)$ and $(\mu_2 = 1.5, \sigma_2 = 0.25)$. For the linear

transformation matrix $\mathbf{B}$, we used the 2×2 matrix below

$$\mathbf{B} = \begin{pmatrix} 1.5 & 0 \\ 0 & 0.1 \end{pmatrix}. \tag{14}$$

We chose to set the Gumbel copula parameter α at 1.5. When generating the data based on the CARP-MLN model, we used the same location and scale parameters for the marginal lognormal distributions, and adjusted the correlation parameter η to obtain the same Kendall's tau as in the CARP-copula model.

Second, to evaluate the effect of the Gumbel copula coefficient α, we varied its value at 1, 1.12, 1.5 and 2.22 by changing the Kendall's tau parameter in the true models, while keeping all other parameters the same with the sample size fixed at $n = 1000$. Considering that the CARP-MLN and CARP-copula models quantify the Kendall's tau differently, we used η at 0, 0.0443, 0.1445 and 0.299 for the CARP-MLN model, so that the Kendall's tau from both CARPs are varied at 0, 0.11, 0.33 and 0.55.

We also explored the impact of the scale parameters by letting σ_1 and σ_2 vary at 0.35, 0.3, 0.25 to 0.20 for the marginal lognormal distributions of the CARP-Copula model while keeping sample size at $n = 1000$ and the Kendall's tau was kept at 0.33. Similarly for the CARP-MLN model, covariance matrices were adjusted to align with the marginal distributions used in the CARP-Copula model.

Lastly, the effect of covariate adjustment was studied by using different types of $\mathbf{B}$. True models used both non-zero and zero $\mathbf{B}$ matrices. In the non-zero $\mathbf{B}$ case, we used the coefficient matrix in (14) to generate data. In the zero $\mathbf{B}$ case, $\mathbf{B} = \mathbf{0}$ was used. The Kendall's tau was set to be 0.33, while other parameters are the same as ones in the sample size case. A sample size of 1000 is used across true models.

5.2   Simulation Results

The simulation results under different true models and fitted models are summarized in this section. For each true model, the model performance was summarized over 1000 simulated data sets. The simulation size was chosen to ensure a reliable analysis result while balancing the computing time needed to evaluate a broad number of scenarios. Further increasing the simulation size would not result in a meaningful change in the evaluated summary statistics and the drawn conclusions. Under each fitted model, the average AIC was computed. In order to evaluate the performance of parameter estimates, we calculated the MSE of the location parameter μ, scale parameter σ and the linear transformation parameter $\mathbf{B}$ for

Table 2: Average AIC from CARP-MLN and CARP-copula calculated by 1000 repeated samples on true models generated by both CARP-MLN and copula models. Sample sizes from true models are changed from 200, 500, 1000 and 2000.

True Model		Copula Generation		MLN Generation	
Fitted Model		MLN	Copula	MLN	Copula
	200	702.0	722.5	722.9	721.9
Sample	500	1755.4	1751.3	1806.1	1808.7
size n	1000	3513.7	3504.9	3615.4	3622.3
	2000	7027.7	7014.3	7220.0	7234.8

Table 3: Average AIC by CARP-MLN and copula under different Kendall's tau. α and η are used to adjust the Kendall's tau in the true models for CARP-MLN and CARP-copula, respectively.

True Model		Copula Generation		MLN Generation	
Fitted Model		MLN	Copula	MLN	Copula
	0	3472.2	3459.5	3459.4	3455.7
Kendall's	0.11	3510.5	3506.9	3489.3	3490.8
tau (τ)	0.33	3513.7	3504.9	3611.7	3618.6
	0.55	3445.6	3439.7	3837.3	3852.3

different sample sizes.

Table 2 shows the average AIC of the fitted models summarized over the 1000 simulated data sets generated using both the CARP-copula and CARP-MLN models at different sample sizes. Both the CARP-MLN and CARP-copula models were applied to each simulated data set. We can see at each fixed sample size, the fitted CARP model that matches the true model used for simulating the data generally outperforms the other model by having a smaller average AIC value, except for the smallest size case at $n = 200$ where the difference in the average AIC values are extremely small. The advantage of using the true model becomes

Table 4: The Average AIC by CARP-MLN and copula under different scale parameters. σ_1 and σ_2 are set equal in the true models for CARP-MLN and CARP-copula.

True Model		Copula Generation		MLN Generation	
Fitted Model		MLN	Copula	MLN	Copula
Scale	0.35	4084.7	4071.7	4148.9	4158.1
parameter	0.30	3747.4	3736.1	3820.0	3826.7
(σ_1 and σ_2)	0.25	3513.7	3504.9	3615.4	3622.3
	0.20	2891.6	2882.7	2980.9	2983.3
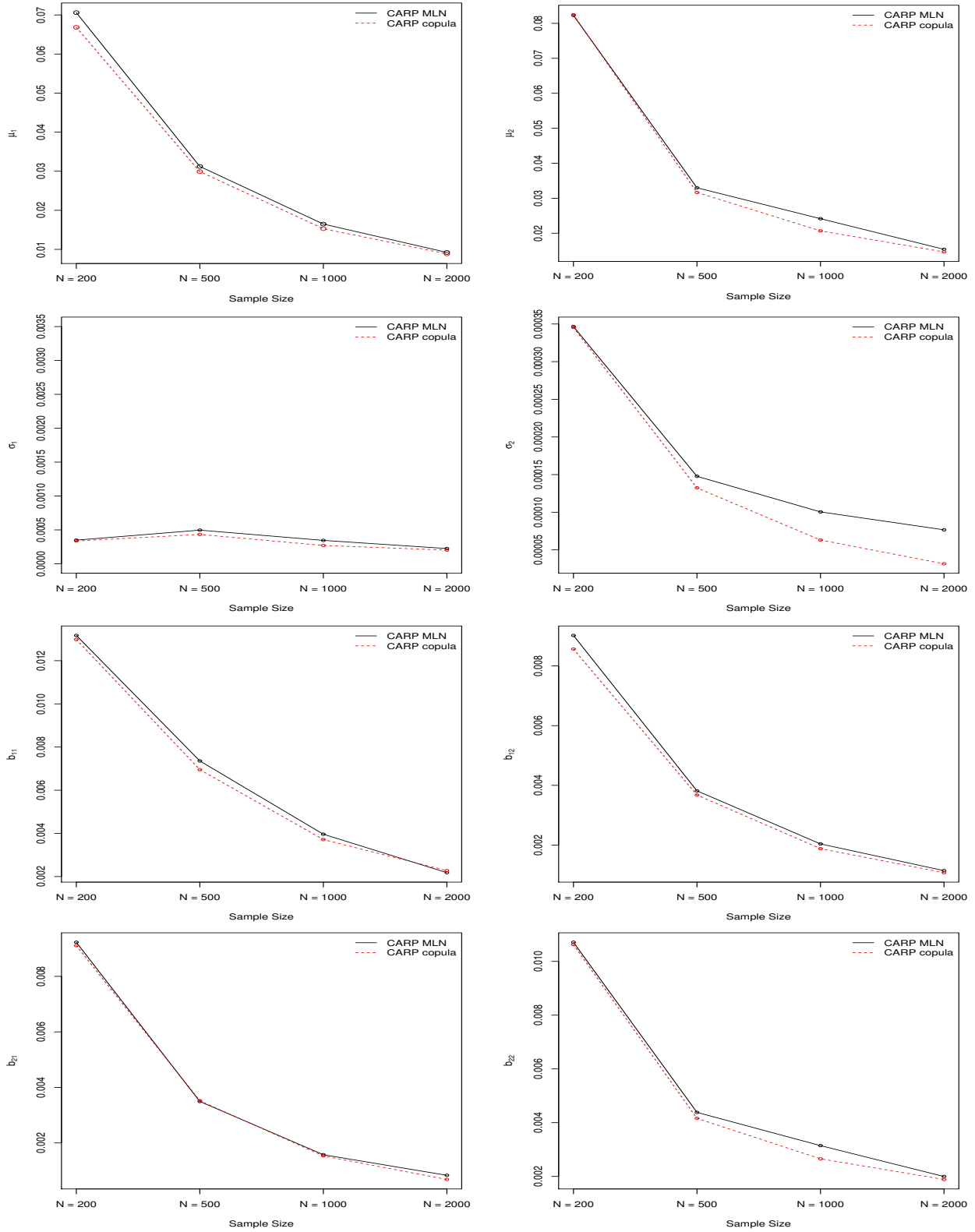
Figure 4: MSE for the location and parameter μ_1, μ_2, scale parameter σ_1, σ_2 and coefficient **B**, calculated by both CARP-MLN and copula model with different sample sizes. The true model is CARP-copula with lognormal marginal distributions.
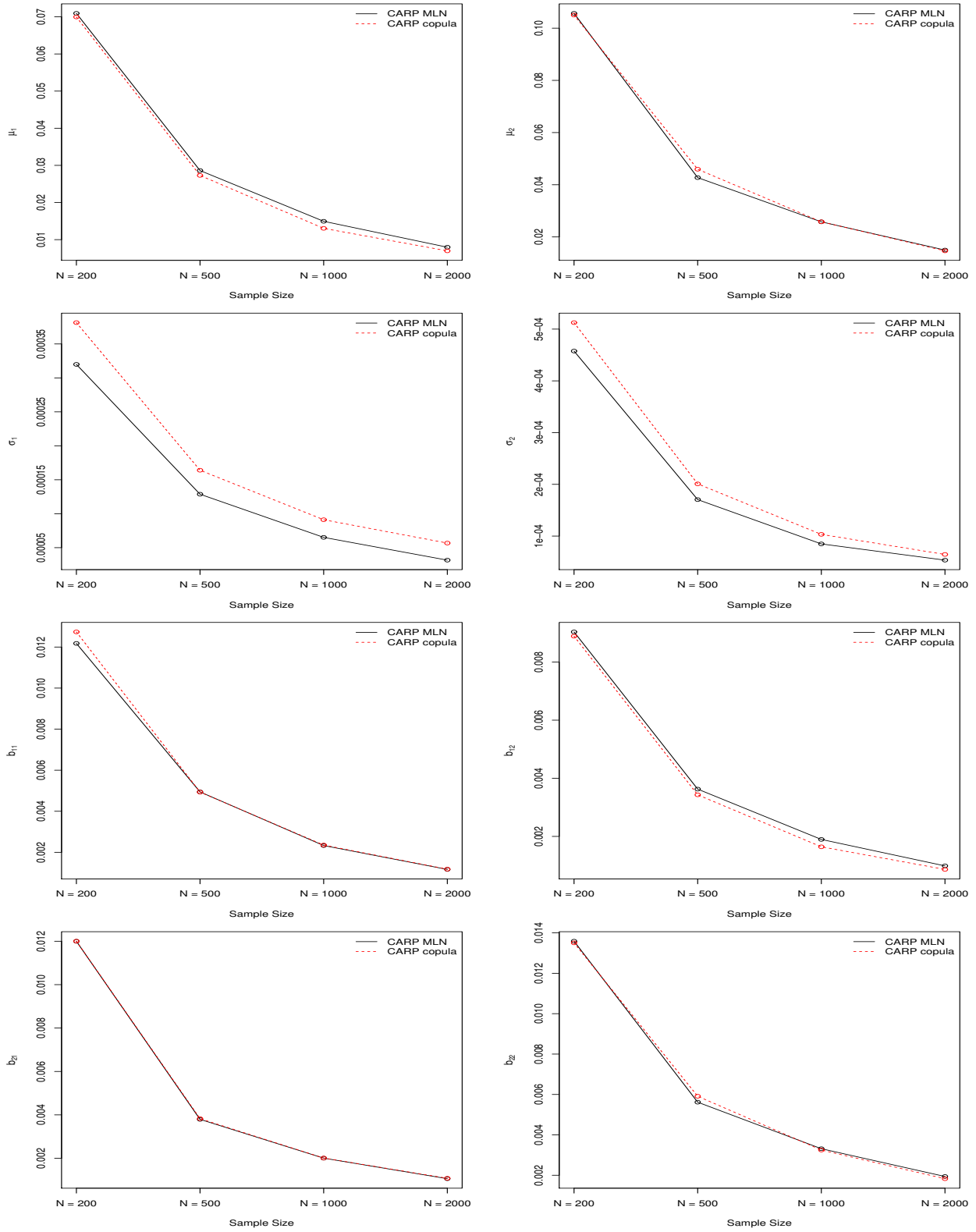
Figure 5: MSE for the location parameter μ_1, μ_2, scale parameter σ_1, σ_2 and $\mathbf{B}$, calculated by both CARP-MLN and copula model with different sample sizes. The true model used to generated data is CARP-MLN.

Table 5: Average AIC from different true models and fitted models to evaluate effect of covariate adjustment. True and fitted models are copula and MLN with or without coefficient **B**.

True Model	Fitted Model	Copula		MLN	
		Non-zero **B**	Zero **B**	Non-zero **B**	Zero **B**
Copula	Non-zero **B**	3505.6	4069.9	3514.4	4064.8
	Zero **B**	2480.5	2485.1	2492.9	2489.0
MLN	Non-zero **B**	3617.7	4186.2	3610.8	4180.0
	Zero **B**	2572.9	2568.7	2565.3	2561.1

more prominent as the sample size increases.

Figures 4 and 5 compare the MSE of the estimated parameters at different sample sizes between the two fitted models for the simulated data generated from the CARP-copula and the CARP-MLN models, respectively. The MSEs are evaluated for the location parameter μ, the scale parameter σ and linear transformation coefficient **B** which are included in both figures. The black solid lines represent the MSE based on the fitted CARP-MLN model, and the red dashed lines represent the estimates from the fitted CARP-copula model. Two major conclusions can be drawn from these two figures. First, the MSEs of all the parameters decrease with the sample size. The more data we generate and use to fit the models, the more reliable the estimated models are with more accurate estimates of model parameters regardless of the choice of the model. Second, the use of the correct model does lead to slightly more accurate estimate of the model parameters across all sample size and model parameters. However, the improvement is more prominent for the scale parameters σ than other model parameters. When the true underlying model is generated by CARP-copula and far from bivariate lognormal, CARP-copula fits data better.

Table 3 shows the average AIC values of the fitted models when the Kendall's tau for capturing the dependence between two event types varies at the levels of 0, 0.11, 0.33 and 0.55. When the Kendall's tau is not zero which suggests some level of dependence between two marginal variables, using a fitted model that matches the true model for generating the data will result in a smaller average AIC value with a bigger improvement achieved when stronger dependence exists between the two types of events. When the Kendall's tau is zero indicating the marginal distributions are independent, the average AIC values are generally similar regardless which model is used to fit the data.

Table 4 shows the comparison based on using different scale parameters in the true underlying model for both the CARP-copula and CARP-MLN models. Again, across all evaluated scale parameter values, we observe consistently smaller average AIC values when

the fitted model matches the true model used for data generation.

Lastly, Table 5 compares the results of using different coefficient matrix **B**. We simulated data with both non-zero and zero **B** using both the CARP-MLN and CARP-copula models. For each simulated data, we fitted both the CARP-MLN and CARP-copula models with non-zero and zero **B**. Similar patterns can be observed. When the true model was used as the fitted model, it results in the smallest AIC value. When non-zero **B** was used to generate the data, the use of covariate adjustment led to substantial improvement in the average AIC of the fitted model compared to using zero **B**. On the other hand, when the data were generated with zero **B**, using the non-zero **B** fitted models produced similar AIC values as the zero **B** models. Therefore, the covariate adjustment is generally recommended due to its potential to substantially improving the model performance by leveraging the additional covariate information.

6   Analysis of the Geyser Data

In this section, we present the analysis for the bivariate geyser system in the Yellowstone National Park using the proposed CARP models. Two adjacent geysers including the West Triplet and the Grotto Geyser are considered for our analysis. Geyser eruptions are highly related to underground water levels, which can be affected by a nearby geyser eruption. Also, it is believed that the eruption duration could affect the gap time until the next eruption. In particular, the longer the current eruption lasts, the longer it will take for the next eruption to occur. This is because a longer eruption usually indicates more water consumption during the eruption and hence a longer water gathering time is expected to reach the next eruption.

We use both the CARP-MLN and the CARP-copula models with lognormal marginal distributions to analyze the geyser data. For the CARP-MLN model, the parameters are $\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma_1, \sigma_2, \eta, b_{11}, b_{12}, b_{21}, b_{22})'$, where $\mu_1, \mu_2, \sigma_1, \sigma_2$ and η define the baseline location parameters and scale parameters in the bivariate lognormal distribution. For the CARP-Copula model, the parameters are $\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma_1, \sigma_2, \alpha, b_{11}, b_{12}, b_{21}, b_{22})'$, where μ_1, μ_2, σ_1 and σ_2 are the location and scale parameters in marginal lognormal distributions, and α is the coefficient parameter in the Gumbel copula. In both cases, the 2×2 matrix **B** is the linear coefficient, and it can be denoted as

$$\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}.$$

Table 6: Parameter estimates and 95% confidence intervals from CARP-MLN model for the geyser data.

Parameter	Estimates	95% lower	95% upper
μ_1	1.881	1.839	1.922
μ_2	2.126	2.072	2.180
b_{11}	0.880	0.720	1.041
b_{21}	0.058	-0.166	0.283
b_{12}	-0.003	-0.008	0.003
b_{22}	0.063	0.050	0.077
η	-0.053	-0.093	-0.014
σ_1	0.416	0.391	0.440
σ_2	0.493	0.459	0.527
τ	-0.069	-0.070	-0.067

Table 6 shows estimation results for the CARP-MLN model, while Table 7 presents estimation results for the CARP-copula model. For simplicity, we use $\widehat{\boldsymbol{\theta}}_{\mathrm{MLN}}$ and $\widehat{\boldsymbol{\theta}}_{\mathrm{CP}}$ to represent parameter estimates based on the CARP-MLN and CARP-copula models, respectively.

The estimated linear coefficients $\widehat{\mathbf{B}}$ for both models are shown as follows,

$$\widehat{\mathbf{B}}_{\mathrm{MLN}} = \begin{pmatrix} 0.880 & -0.003 \\ 0.058 & 0.063 \end{pmatrix} \quad \text{and} \quad \widehat{\mathbf{B}}_{\mathrm{CP}} = \begin{pmatrix} 0.862 & -0.017 \\ 0.081 & 0.062 \end{pmatrix}. \tag{15}$$

The estimates in (15) indicate that in general the longer the previous eruption duration is, the longer waiting time it takes until the next eruption. The eruption duration impact is substantial for the West Triplet Geyser as shown by $\widehat{b}_{11} = 0.880$ in $\widehat{\mathbf{B}}_{\mathrm{MLN}}$, which suggests the marginal previous eruption duration effect of West Triplet Geyser on its location parameter is 0.880 on average.

For the CARP-MLN model, the estimated covariance matrix is

$$\widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} 0.172 & -0.022 \\ -0.022 & 0.246 \end{pmatrix},$$

where the correlation estimation is calculated as $\widehat{\eta}/\sqrt{\widehat{\sigma}_2^2 + \widehat{\eta}^2} = -0.05$. This indicates a small negative correlation between the event gap times of the West Triplet and Grotto Geysers. In other words, a longer eruption gap time for West Triplet Geyser could be associated with a shorter time interval for the next eruption of the Grotto Geyser. The Kendall's tau provides

Table 7: Parameter estimates and the corresponding 95% confidence intervals from CARP-copula model using lognormal marginal distributions for the geyser data.

Parameter	Estimates	95% lower	95% upper
μ_1	1.847	1.809	1.886
μ_2	2.102	2.051	2.152
b_{11}	0.862	0.700	1.024
b_{21}	0.081	-0.147	0.311
b_{12}	-0.017	-0.007	0.004
b_{22}	0.062	0.047	0.076
σ_1	0.417	0.392	0.442
σ_2	0.497	0.459	0.535
α	1.000	0.958	1.042
τ	0.000	-0.001	0.001

a measure on event dependence in both models. The estimates from the CARP-MLN and the CARP-copula models are $\widehat{\tau}_{\mathrm{MLN}} = -0.07$ and $\widehat{\tau}_{\mathrm{CP}} = 0$, respectively. The confidence intervals for τ is calculated by the Delta method in Appendix A.2. The calculated AICs for the CARP-MLN and CARP-copula models are 5113.1 and 5120.5, respectively, indicating the CARP-MLN model is a slightly better fit for the geyser data.

One way to measure the goodness of fit is to compare the estimated cumulative intensity function with the observed. The estimated and observed cumulative intensity functions from both models are shown in Figures 6 and 7, with the black solid lines and the red dashed lines representing the estimated and the observed cumulative intensity functions respectively. We can see a better agreement between the estimated and the observed cumulative intensity functions based on the CARP-copula model compared with the CARP-MLN model.

7 Concluding Remarks

This paper introduces two CARP models, CARP-MLN and CARP-Copula, for modeling a bivariate recurrent process. With the covariate adjustment, the CARP models provide improvements over traditional models for capturing the interdependence of the bivariate event system. When the underlying data are close to a bivariate lognormal distribution, both models work similarly well. However, when the real data are inconsistent with a bivariate lognormal distribution, the CARP-copula model is recommended to improve the model performance.
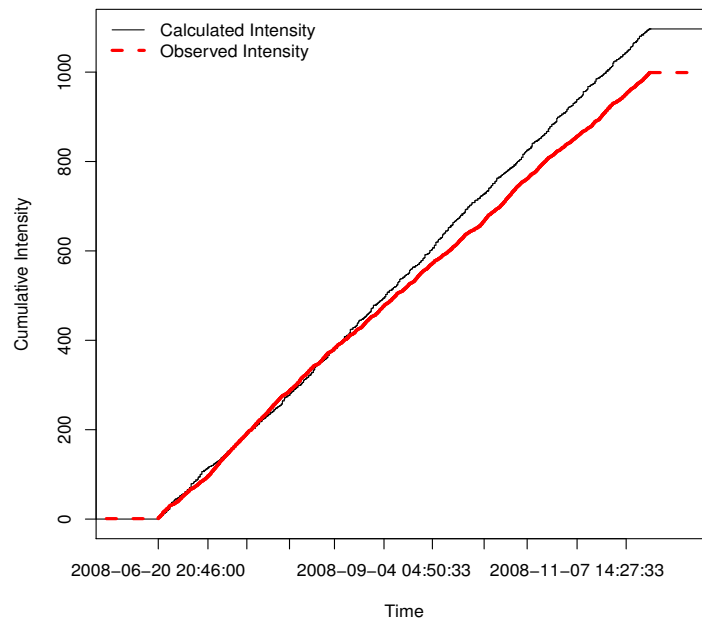
Figure 6: Cumulative intensity function from fitted CARP-MLN model using the geyser data.
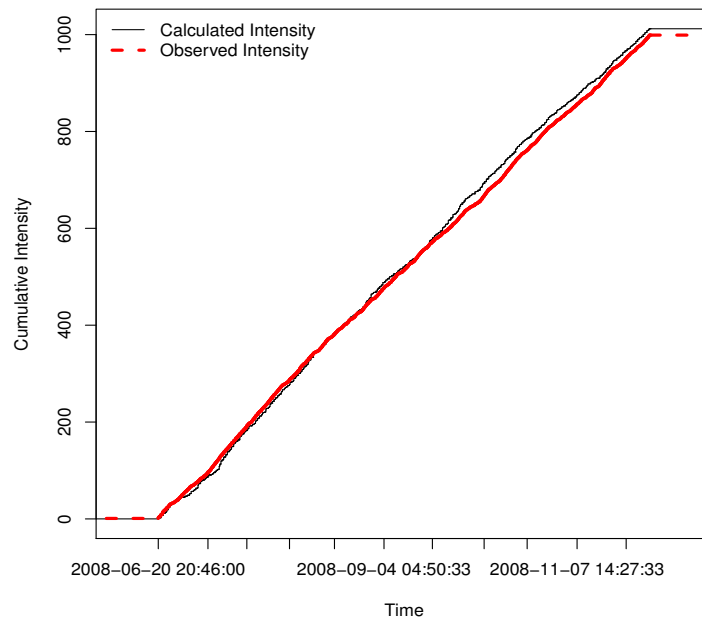

Figure 7: Cumulative intensity function calculated from the CARP-copula model with the lognormal marginal distributions.

A special case of Archimedean copulas, the Gumbel copula function, is considered in this paper. However, the general method can easily be adapted for other copula functions available for the chosen models. One example is the Frank copula. Compared to the Gumbel copula where only positive dependence can be quantified, the Frank copula can accommodate both positive and negative dependence between two event types. On the other hand, the Gumbel copula has an asymmetric dependence structure where correlations on the tail can be very different than both the Gaussian and Frank copulas. In addition, differences of the joint density functions between the Frank and Gaussian copulas are negligible when the marginal distributions are the same.

The adjustment by using effective covariates significantly improves the model performance. We have shown through the simulation study, when the true model includes a significant covariate effect, that using the covariate adjustment significantly improves model fitting. On the other hand, if there is no significant covariate effect, using the covariate adjustment does not raise the AIC values. Therefore, in general we recommend using the covariate adjustment given the true underlying model is unknown.

The choice of the marginal distributions in the CARP-copula model was made based on the AIC value. Marginal distributions achieving the minimum AIC value were selected in our model. One advantage of using the CARP-copula model is that different marginal distributions can be easily used to model different event types. This offers tremendous flexibility and broader generality to the CARP-copula model.

A multivariate CARP model will be considered for future work when there are more than two event types in the recurrent system. For the geyser application, other covariates in addition to the eruption duration would also be explored to further improve the performance of the CARP models. In engineering applications when testing the reliability of systems is often of interest. The loading-sharing system is related to the CARP model (e.g., Smith, 1983, Tierney, 1982, Sutar and Naik-Nimbalkar, 2014, and Zhang et al., 2020). In the future, it will be interesting to apply the CARP to model the reliability of loading sharing systems.

Acknowledgments

A Appendix

A.1 Conditional Lognormal Probability

If a bivariate random variable $\boldsymbol{y} = (y_1, y_2)'$ follows a lognormal distribution $\mathrm{MLN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where

$$\boldsymbol{\mu} = (\mu_1, \mu_2)' \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix},$$

then the conditional distribution of $\log(y_1)|\log(y_2)$ follows a normal distribution $\mathrm{N}(\mu_c, \sigma_c)$ with location and scale parameter as

$$\mu_c = \mu_1 + \sigma_{12}\sigma_{22}^{-1}[\log(y_2) - \mu_2] \quad \text{and} \quad \sigma_c = \sigma_{11} - \sigma_{12}\sigma_{22}^{-1}\sigma_{21},$$

respectively.

A.2 Confidence Interval for Kendall's tau

For lognormal cases, the Kendall's tau estimator can be expressed as

$$\widehat{\tau} = \frac{2}{\pi} \arcsin\left(\frac{\widehat{\eta}}{\sqrt{\widehat{\sigma}_2^2 + \widehat{\eta}^2}} \right),$$

where the asymptotic distribution is known from the ML estimator. Using the Delta method,

$$\mathrm{Var}(\widehat{\tau}) = \left(\frac{\partial \widehat{\tau}}{\partial \widehat{\eta}}\right)^2 \mathrm{Var}(\widehat{\eta}) + \left(\frac{\partial \widehat{\tau}}{\partial \widehat{\sigma}_2}\right)^2 \mathrm{Var}(\widehat{\sigma}_2) + 2\left(\frac{\partial \widehat{\tau}}{\partial \widehat{\eta}}\right)\left(\frac{\partial \widehat{\tau}}{\partial \widehat{\sigma}_2}\right) \mathrm{cov}(\widehat{\eta}, \widehat{\sigma}_2),$$

where

$$\frac{\partial \widehat{\tau}}{\partial \widehat{\eta}} = \frac{2}{\pi} \frac{1}{\sqrt{1 - \dfrac{\widehat{\eta}^2}{\widehat{\sigma}_2^2 + \widehat{\eta}^2}}} \left[(\widehat{\eta}^2 + \widehat{\sigma}_2^2)^{-\frac{1}{2}} - \frac{1}{2}\widehat{\eta}(\widehat{\eta}^2 + \widehat{\sigma}_2^2)^{-\frac{3}{2}} \right],$$

and

$$\frac{\partial \widehat{\tau}}{\partial \widehat{\sigma}_2} = -\frac{2}{\pi} \frac{1}{\sqrt{1 - \dfrac{\widehat{\eta}^2}{\widehat{\sigma}_2^2 + \widehat{\eta}^2}}} \left[-\widehat{\eta}\widehat{\sigma}_2(\widehat{\eta}^2 + \widehat{\sigma}_2^2)^{-\frac{3}{2}} \right].$$

For the Gumbel copula case,

$$\widehat{\tau} = 1 - \frac{1}{\widehat{\alpha}}.$$

Similarly with the lognormal case, the variance for the estimator can be written by using the Delta method. That is,

$$\mathrm{Var}(\widehat{\tau}) = \Big[\frac{\partial g(\widehat{\alpha})}{\partial \widehat{\alpha}}\Big]^2 \mathrm{Var}(\widehat{\alpha}),$$

where $g(x) = 1 - 1/x$.

References

P. K. Andersen and R. D. Gill. Cox's regression model for counting processes: a large sample study. *The Annals of Statistics*, 10:1100–1120, 1982.

O. Bouaziz, F. Comte, and A. Guilloux. Nonparametric estimation of the intensity function of a recurrent event process. *Statistica Sinica*, 23:635–665, 2013.

G. Casella and R. L. Berger. *Statistical Inference*, volume 2. Duxbury Pacific Grove, CA, 2002.

R. J. Cook and J. F. Lawless. *The Statistical Analysis of Recurrent Events*. Springer-Verlag, New York, 2007.

D. R. Cox. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*, 34:187–220, 1972.

D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. Springer, 2003.

J.-Y. Dauxois and S. Sencey. Non-parametric tests for recurrent events under competing risks. *Scandinavian Journal of Statistics*, 36:649–670, 2009.

L. Duchateau and P. Janssen. *The Frailty Model*. Springer-Verlag, New York, 2008.

R. O. Fournier. Old faithful: A physical model. *Science*, 163:304–305, 1969.

F. G. Garre, A. H. Zwinderman, R. B. Geskus, and Y. W. Sijpkens. A joint latent class changepoint model to improve the prediction of time to graft failure. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171:299–308, 2008.

X. Huang and L. Liu. A joint frailty model for survival and gap times between recurrent events. *Biometrics*, 63:389–397, 2007.

A. V. Huzurbazar and B. J. Williams. Incorporating covariates in flowgraph models: applications to recurrent event data. *Technometrics*, 52:198–208, 2010.

J. Lawless. The analysis of recurrent events for multiple subjects. *Applied Statistics*, 44: 487–498, 1995.

B. Lindqvist. On the statistical modeling and analysis of repairable systems. *Statistical Science*, 21:532–551, 2006.

L. Liu and X. Huang. Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58:65–81, 2009.

R. Meyer and J. S. Romeo. Bayesian semiparametric analysis of recurrent failure time data using copulas. *Biometrical Journal*, 57:982–1001, 2015.

P. Prasad and K. Rao. Reliability models of repairable systems considering the effect of operating conditions. In *Proceedings of Annual of Reliability and Maintainability Symposium*, pages 503–510. IEEE, 2002.

J. S. Rinehart. Fluctuations in geyser activity caused by variations in earth tidal forces, barometric pressure, and tectonic stresses. *Journal of Geophysical Research*, 77:342–350, 1972.

J. S. Romeo, R. Meyer, and D. I. Gallardo. Bayesian bivariate survival analysis using the power variance function copula. *Lifetime Data Analysis*, 24:355–383, 2018.

D. E. Schaubel and J. Cai. Non-parametric estimation of gap time survival functions for ordered multivariate failure time data. *Statistics in Medicine*, 23:1885–1900, 2004.

R. L. Smith. Limit theorems and approximations for the reliability of load-sharing systems. *Advances in Applied Probability*, pages 304–330, 1983.

L. Sun, D.-H. Park, and J. Sun. The additive hazards model for recurrent gap times. *Statistica Sinica*, 16:919–932, 2006.

S. S. Sutar and U. Naik-Nimbalkar. Accelerated failure time models for load sharing systems. *IEEE Transactions on Reliability*, 63:706–714, 2014.

L. Tierney. Asymptotic bounds on the time to fatigue failure of bundles of fibers under local load sharing. *Advances in Applied Probability*, pages 95–121, 1982.

S. Ullah, T. J. Gabbett, and C. F. Finch. Statistical modelling for recurrent events: an application to sports injuries. *British Journal of Sports Medicine*, 48:1287–1293, 2014.

Q. Yang, N. Zhang, and Y. Hong. Statistical reliability analysis of repairable systems with dependent component failures under partially perfect repair assumption. *IEEE Transactions on Reliability*, 62:490–498, 2013.

Q. Yang, Y. Hong, N. Zhang, and J. Li. A copula-based trend-renewal process model for analysis of repairable systems with multitype failures. *IEEE Transactions on Reliability*, 66:590–602, 2017.

D. Zeng and D. Lin. Semiparametric transformation models with random effects for joint analysis of recurrent and terminal events. *Biometrics*, 65:746–752, 2009.

J. Zhang, Y. Zhao, and X. Ma. Reliability modeling methods for load-sharing k-out-of-n system subject to discrete external load. *Reliability Engineering & System Safety*, 193: 106603, 2020.