Journal of Sciences, Islamic Republic of Iran 34(1): 35 - 42 (2023) http://jsciences.ut.ac.ir
University of Tehran, ISSN 1016-1104

Copula Functions for Spatial Survival Data Analysis

N. Ebrahimi[1], M. Mohammadzadeh[1*], G. Cortese[2]

[1] *Department of Statistics, Tarbiat Modares University, Tehran, Islamic Republic of Iran*
[2] *Department of Statistical Sciences, Padua University, Italy*

Received: 24 October 2022 / Revised: 27 January 2022 / Accepted: 27 February 2022

Abstract

Many survival data analyses aim to assess the effect of different risk factors on survival time. In some studies, the survival times are correlated, and the dependence between survival times is related to their spatial locations. Identifying and considering the dependence structure of data is essential in survival modeling. The copula functions are helpful tools for incorporating data dependencies. So, one may use these functions for modelling spatial survival data. This paper presents a model for spatial survival data by the Gumbel-Hougaard copula function. A two-stage estimator using a composite likelihood function is used to estimate regression and dependence parameters. A simulation study investigates the performance of the model. Finally, the proposed model is applied to model a set of COVID-19 data.

Keywords: Spatial Survival Data; Copula Function; Composite Likelihood; Frailty Models; Two-Stage Estimator.

Introduction

Analyzing the joint pattern of diseases in specific population groups has become a popular research field in many epidemiological studies. These groups could be family members, residents of a geographic region, patients of a clinical centre, etc. Analyzing dependent survival data requires a model for joint survival function. Such multivariate survival data analysisaims, especially in biomedical and epidemiological research, to determine the effects of some risk factors on the joint hazard (or survival) function. Ordinary survival models, such as the Cox proportional hazard model, can explain the effect of known factors as covariates in the model. Sometimes some unknown or unmeasurable factors affect the hazard function and cause some dependencies between survival times. We should consider any dependence structure of the data in survival modeling; Otherwise, it can lead to misleading results. Frailty and copula models are two commonly used approaches for modelling dependent survival data. These models take the association between survival times into account in two manners. The frailty models consider the dependency in the model using a latent random effect term (so-called frailty) and consider a known distribution for this latent variable. From the conditional distribution point of view, frailty models assume that the observations are independent conditionally on the frailty term (1). Used frailty models in analyzing univariate survival data. They extended the ordinary Cox regression model by multiplying it by a frailty term, W, as follows

$$\lambda(t|Z) = \mathrm{W} \cdot \lambda_0(\mathrm{t})\exp(\beta' Z). \qquad (1)$$

There are different types of dependency on survival data. In some cases, their dependencies are related to

[*] Correspomding author: Tel: +989122066712; Email: mohsen_m@modares.ac.ir

their spatial locations. When the survival data's dependency is due to their spatial locations, they will be called spatial survival data (2). Modeled this type of data by adding a spatial random field as frailty to the proportional hazards model as follows

$$\lambda\big(t, s\big|Z(s)\big) = \lambda_0(t)\exp\big(\beta' Z + X(s)\big), \qquad (2)$$

where X(s) is a Gaussian random field to explain the spatial dependence between subjects. They used this model to analyze the childhood asthma data (3). considered a non-Gaussian random effect model and proved the model parameters' identifiability in this case (4). Investigated the spatial structure of leukemia survival data using a spatial survival model (5). Assumed the proportional odds model as the marginal hazard and used frailties to model spatial dependency structure (6). Considered random effects from a CAR model as frailties in accelerated failure time, proportional hazards, and proportional odds models (7). Used the spatial survival model for interval-censored data. They used the Bayesian approach to estimate the model parameters. Frailty models are appropriate when within-cluster inferences are desired because the covariate effects in these models are interpreted as being conditional on the frailties and are cluster-specific (8).

On the other hand, copula functions are powerful tools for constructing multivariate distributions based on their marginal distributions and dependence structure. The variety of dependency structures that copulas can model, copula families' flexibility, and the copulas' relative mathematical simplicity make them a popular tool for modeling dependencies between random variables.

The word copula was first employed mathematically or statistically by (9) in the theorem describing the functions that combine one-dimensional distribution functions to form multivariate distribution functions (10). From one point of view, copulas are functions that join or couple multivariate distribution functions to their one-dimensional marginal distribution functions. Alternatively, copulas are multivariate distribution functions whose one-dimensional margins are uniform at intervals (0,1) (10). We can use the copula functions to model joint survival functions based on assumed marginals (11). Considered a bivariate association model for an ordered pair of individuals in constructing bivariate life tables. Without knowing the copula concept, his model uses a copula to create the bivariate survival function (12). Used the copula function to analyze spatial dependent data (13) and (14) modeled drought data with copula functions (15). Assumed a marginal proportional hazard model and used the Gaussian copula function to model the geostatistical survival data's spatial dependence structure (16). Used a copula-based approach to evaluate the effect of dependence in stress-strength models in the context of reliability (17). Assumed the Farlie-Gumbel-Morgenstern copula and modeled the dependence parameter as a function of geographic and demographic pairwise distances. For the estimation of the dependence parameters, they presented pairwise composite likelihood equations (18). Modeled the marginal survival function with a Bayesian nonparametric model and applied the Gaussian copula function to describe the survival data's spatial dependence structure. They also considered the Bayesian version of the (15) model and assumed a piecewise exponential prior for the baseline hazard function (19). Extended the Archimedean copula methodology to model multivariate survival data grouped in clusters.

The methodology in this paper is motivated by modeling spatially dependent survival data. We use the Gumbel-Hougaard bivariate copula to model pairwise survival functions. We assume that the copula's parameter is a function of Euclidian distances between subjects to capture the spatial dependence structure of survival data. A two-stage estimation method is used to estimate the model parameters. First, the marginal model parameters are estimated using a marginal likelihood, and then we use the composite likelihood approach for the spatial dependence parameters. The rest of the paper is structured as follows: Section 2 describes the model in detail. Section 3 presents the estimation procedure. In Section 4, the model's performance is evaluated via a simulation study. Section 5 contains the application of the model on a real data set. Finally, a brief discussion is presented in Section 6.

Model Specification

Let T_i and C_i, *i=1,…,n*, be the unobserved failure and right censoring time, respectively, and Z_i be the p-vector of covariates. Conditional on Z, T, and C are assumed to be independent. For the i-th subject, the observed quantity is $\widetilde{T_i} = \min(T_i, C_i)$. In order to know whether the event happened or the time has been censored, the indicator function $\delta_i = I(T_i \leq C_i)$ is defined. For subject *i*, one can observe $\big(\widetilde{T_i}, \delta_i, Z_i\big)$ and geographic coordinates x_i and y_i.

We model the dependence structure between subjects by modeling the bivariate survival function. For this purpose, assume that the marginal survival function of T follows the Weibull model given by

$$S(t|Z) = \exp(-\lambda t^{\alpha}\exp(\beta' Z)) \qquad (3)$$

where α and λ are location and scale parameters, respectively, and β is a regression coefficient vector. Model (3) is a marginal model for T_is, and the estimated β from this model has a population-average interpretation.

When the survival data are spatially correlated, one may use the copula functions to model the survival data's spatial dependence structure. Spatial data are correlated, and the dependence decreases as the spatial distance between units increases. So, a copula function is considered a valid spatial copula if it covers positive dependencies between variables. In addition, the copula results in independence and maximum dependency, when the spatial distance between units, d, goes to infinity and zero, respectively. For every copula C and every $(u, v) \in I^2$,

$$W(u,v) = \max(u + v - 1,0) \leq C(u,v) \leq M(u,v) = \min(u,v),$$

where $M$and $W$are referred to Fréchet-Hoeffding upper and lower bound, respectively. Another important copula is product copula $\Pi(u,v) = uv$. So, in the context of copula functions, for a valid spatial copula, $d \to \infty$ results in the product copula, and $d \to 0$ leads to Fréchet-Hoeffding upper bound.

To model the spatial dependence structure of survival data, functions should be used that, in addition to having the conditions of a valid spatial copula, also match the characteristics of the survival data. Survival data are always positive, and their distributions are often skewed. So, like Frank copula, symmetric copula functions could not be the appropriate choice to fit survival data. While, Gumbel-Hougaard and Clayton copulas can model upper and lower tail dependence, respectively. In the context of joint survival models, a copula with upper tail dependence exhibits the association between early event times. At the same time, a copula with lower tail dependence is used to represent the dependency of late event times. Commonly, late event times are influenced by right censoring. So, a lower tail dependence copula is affected more by the right censoring and has a weak performance in inference (20).

o model the dependence structure of the data, we assumed the Gumbel-Hougaard (GH) copula function, given by

$$S_{ij}(t_i, t_j) = C_{ij}\left(S_i(t_i), S_j(t_j)\right) = \exp\left\{-\left[(-\ln S_i(t_i))^{\theta_{ij}} + \left(-\ln S_j(t_j)\right)^{\theta_{ij}}\right]^{1/\theta_{ij}}\right\} \quad (4)$$

where $S_{ij}(t_i, t_j) = P(T_i > t_i,\ T_j > t_j)$ is the joint survival function, $S_i(t_i)$ and $S_j(t_j)$ are the marginal survival functions for subjects i and j, respectively, and $\theta_{ij} \in (1, \infty)$ is the copula function parameter that measures the dependency of two margins. A GH copula meets the requirements of a valid spatial copula. It covers only the positive dependencies between variables. Moreover, when $\theta_{ij} = 1$, $C_1 = \Pi$, and when $\theta_{ij} \to \infty$, $C_\infty = \mathrm{M}$. Furthermore, it is able to model upper tail dependence, which is consistent with the distribution of the right censored survival data.

There is a one-to-one relation between the copula function parameter, $\theta_{ij} \in (1, \infty)$, and the global measure of dependence, Kendall's correlation, because $\theta_{ij} = \frac{1}{1-\tau_{ij}}$. To consider the spatial dependence in the model, τ_{ij} can be viewed as a function of the geographic distance between subjects i and j. (21) introduced a new method to build spatio-temporal covariance functions using Archimedean copula generators and listed some Archimedean copulas and their corresponding stationary spatial covariance functions. Exponential covariance function can be constructed by the generator of the GH copula. So, we assume an exponential model for τ_{ij} given by

$$\tau_{ij} \equiv \tau_{ij}(d_{ij}; \psi) = \frac{2\exp\left(-\frac{d_{ij}}{\psi}\right)}{1+\exp\left(-\frac{d_{ij}}{\psi}\right)} \quad (5)$$

where d_{ij} is the Euclidean distance between the spatial location of subjects i and j, and ψ is the spatial dependence parameter to be estimated. Considering model (5) for τ_{ij}, it is seen that the GH copula meets the requirements of a spatial copula function, since when $d_{ij} \to 0$, $\tau_{ij} \to 1$, $\theta_{ij} \to \infty$ and $C_\infty = \mathrm{M}$. Similarly, when $d_{ij} \to \infty$, $\tau_{ij} \to 0$, $\theta_{ij} = 1$ and $C_1 = \Pi$.

Estimation Procedure

To determine the joint bivariate survival function, we should estimate the marginal survival functions' parameters and the copula function. Since the copula approach separately models the marginal functions and dependence structure, we use a two-stage estimation method. First, β, λ and α are estimated from the marginal survival model using the marginal maximum likelihood procedure. The log-likelihood for the assumed marginal model has the following form

$$\ell_M(\beta, \lambda, \alpha | \tilde{T}, \delta, Z) = \sum_{i=1}^{n} \delta_i [\ln \lambda + \ln \alpha + (\alpha - 1) \ln \tilde{t_i} + z_i\beta] - [\lambda \tilde{t_i}^{\alpha} exp(z_i\beta)] \quad (6)$$

(22) showed that this estimator is consistent for marginal parameters if the marginal model is correctly specified. However, the estimator's variance is affected by the dependency of the observations.

In the second stage, the calculated estimates from the first stage, $\hat{\beta}$, $\hat{\lambda}$ and $\hat{\alpha}$, are inserted into the composite likelihood function, and $\max_{\psi} \mathcal{L}_C(\hat{\beta}, \hat{\lambda}, \hat{\alpha}, \psi)$ is solved for $\hat{\psi}$. A composite likelihood is formed by multiplying a collection of component likelihoods; the context often determines the particular collection used. For an m-dimensional random variable Y with probability density function $f(y; \theta)$,consider a set of marginal or conditional events $\{\mathcal{A}_1, \dots, \mathcal{A}_K\}$ with associated likelihoods $\mathcal{L}_k(\theta; y) \propto f(y \in \mathcal{A}_K; \theta)$. A composite likelihood is a weighted product

$$\mathcal{L}_C(\theta; y) = \prod_{k=1}^{K} \mathcal{L}_k(\theta; y)^{w_k}, \quad (7)$$

where w_ks are nonnegative weights to be determined. The weights can be equal and ignored (23). The pairwise composite likelihood for survival data has the following general form

$$\mathcal{L}_C = \prod_{\substack{i \le j \\ i,j \in D_n}} S_{ij}(t_i, t_j)^{(1-\delta_i)(1-\delta_j)} S_{ij}^{(1)}(t_i, t_j)^{\delta_i(1-\delta_j)} S_{ij}^{(2)}(t_i, t_j)^{(1-\delta_i)\delta_j} f_{ij}(t_i, t_j)^{\delta_i\delta_j} \quad (8)$$

where $S_{ij}^{(1)}(t_i, t_j) = -\frac{\partial}{\partial t_i} S_{ij}(t_i, t_j)$, $S_{ij}^{(2)}(t_i, t_j) = -\frac{\partial}{\partial t_j} S_{ij}(t_i, t_j)$, $f_{ij}(t_i, t_j) = \frac{\partial^2}{\partial t_i \partial t_j} S_{ij}(t_i, t_j)$ and D_n is a set of all pairs *i* and *j* that are in a specific distance from each other. For our proposed joint model:

$$S_{ij}^{(1)}(t_i, t_j) = -\frac{\partial}{\partial t_i} S_{ij}(t_i, t_j) = \frac{P_{ij}^{1/\theta_{ij}} \times (-\ln S_i(t_i))^{\theta_{ij}} \times (-f_i(t_i)) \times \exp\left(-P_{ij}^{1/\theta_{ij}}\right)}{S_i(t_i) \times \ln S_i(t_i) \times P_{ij}}$$

$$S_{ij}^{(2)}(t_i, t_j) = -\frac{\partial}{\partial t_j} S_{ij}(t_i, t_j) = \frac{P_{ij}^{1/\theta_{ij}} \times (-\ln S_j(t_j))^{\theta_{ij}} \times (-f_j(t_j)) \times \exp\left(-P_{ij}^{1/\theta_{ij}}\right)}{S_j(t_j) \times \ln S_j(t_j) \times P_{ij}}$$

$$f_{ij}(t_i, t_j) = \frac{\partial^2}{\partial t_i \partial t_j} S_{ij}(t_i, t_j) = P_{ij}^{1/\theta_{ij}} \times f_i(t_i) \times (-\ln S_i(t_i))^{\theta_{ij}} \times f_j(t_j) \times (-\ln S_j(t_j))^{\theta_{ij}} \times \frac{\exp\left(-P_{ij}^{1/\theta_{ij}}\right)\left(\theta_{ij} - 1 + P_{ij}^{2/\theta_{ij}}\right)}{S_i(t_i) \times \ln S_i(t_i) \times S_j(t_j) \times \ln S_j(t_j) \times P_{ij}^2}$$

where $P_{ij} = (-\ln S_i(t_i))^{\theta_{ij}} + (-\ln S_j(t_j))^{\theta_{ij}}$.

Simulation Study

We investigate the efficacy of the proposed model via a simulation study. For this purpose, we generated a spatial survival data set using a method proposed by (24) For a spatial survival function, we have

$$S(t|Z, s) = \exp(-\Lambda_0(t)\exp(\beta'Z + R(s))) \quad (9)$$

where R(s) is a Gaussian random field with a valid covariance function, and *s* denotes each subject's geographic location. Since $S(t|Z, s)$ has a standard uniform distribution, according to the probability integral transform theorem, the spatial survival time

$$T = \Lambda_0^{-1}\left(-\ln U \times \exp(-(\beta'Z + R(s)))\right), \quad U \sim U(0,1) \quad (10)$$

has equation (3) as its survival function. For our proposed model, with the Weibull baseline hazard function, a spatial survival time can be generated by

$$T = -\left[\frac{\ln U}{\lambda \exp(\beta'Z + R(s))}\right]^{1/\alpha} \quad (11)$$

Following this method, two elements should be generated randomly: *U* from $U(0,1)$ and a Gaussian random field $R(s)$ with a specific covariance function.

First, *n* locations were simulated using the uniform distribution $U(0,1)$ .After calculating the distance matrix of these locations, the Gaussian random field, $R(s)$, with an stationary exponential covariogram, $\sigma(d) = \exp\left(-\frac{d}{\psi}\right)$, were generated. We consider $\alpha = 1$, $\lambda = 0.9$ and one covariate from Bernoulli(0.5) distribution. Using this method ensures that the presented simulations had marginal survival times from Weibull model (3). We consider a 10% censoring rate for simulated data. The results for β and ψ are summarized in Table 1 for n=50, 100 and 500. The MSEs of the estimators $\hat{\beta}$ and $\hat{\psi}$ in this table are defined by $\text{MSE}(\hat{\beta}) = \frac{1}{M}\sum_{i=1}^{M}(\hat{\beta_i} - \beta)^2$ and $\text{MSE}(\hat{\psi}) = \frac{1}{M}\sum_{i=1}^{M}(\hat{\psi_i} - \psi)^2$, where *M* is the number of iterations.

In Table 1, the estimates of ψ have a slight bias in all simulations. The MSE criterion decreases slightly by increasing *n*; This is because only the observations in a specific neighbourhood participate in the estimation of ψ, not all of them. The length of the interval $(Q_{0.025}, Q_{0.975})$ decreases as *n* increases. On the other hand, since the β estimation method assumes the observations are independent, the β estimator's precision slightly reduces by increasing the data dependency by increasing ψ.

Table 1. Simulation results for regression and spatial dependency terms

$\boldsymbol{\beta}$	$\boldsymbol{n}$	$\boldsymbol{\psi}$	$\widehat{\boldsymbol{\beta}}$	$\boldsymbol{SE(\widehat{\beta})}$	$\boldsymbol{MSE(\widehat{\beta})}$	$\widehat{\boldsymbol{\psi}}$	$\boldsymbol{MSE(\widehat{\psi})}$	$(\boldsymbol{Q_{0.025}, Q_{0.975}})$
0.5	50	0.5	0.60	0.171	0.237	0.54	0.638	(0.302,1.185)
		1.0	0.57	0.171	0.246	1.25	0.923	(0.670,1.814)
		1.5	0.53	0.175	0.297	1.61	1.187	(0.978,2.033)
	100	0.5	0.53	0.083	0.106	0.46	0.627	(0.366,0.950)
		1.0	0.56	0.084	0.110	1.09	0.745	(0.788,1.660)
		1.5	0.52	0.084	0.147	1.36	1.132	(1.052,1.935)
	500	0.5	0.51	0.017	0.018	0.46	0.593	(0.316,0.662)
		1.0	0.51	0.017	0.024	1.01	0.697	(0.813,1.391)
		1.5	0.46	0.017	0.024	1.44	1.034	(1.374,1.689)
1.0	50	0.5	1.09	0.171	0.235	0.65	0.759	(0.385,1.090)
		1.0	1.08	0.174	0.281	1.17	0.832	(0.694,1.796)
		1.5	0.99	0.179	0.288	1.38	1.042	(0.861,1.857)
	100	0.5	0.97	0.086	0.094	0.59	0.704	(0.344,0.969)
		1.0	1.01	0.086	0.088	1.20	0.894	(0.823,1.258)
		1.5	0.98	0.088	0.156	1.44	1.017	(1.257,1.680)
	500	0.5	0.99	0.017	0.018	0.44	0.516	(0.346,0.688)
		1.0	1.01	0.017	0.021	0.98	0.657	(0.859,1.074)
		1.5	1.01	0.017	0.027	1.56	0.935	(1.266,1.740)
1.5	50	0.5	1.61	0.165	0.190	0.79	0.909	(0.340,1.176)
		1.0	1.49	0.174	0.205	1.13	0.782	(0.628,1.394)
		1.5	1.50	0.179	0.270	1.37	1.024	(0.824,2.121)
	100	0.5	1.47	0.086	0.085	0.57	0.682	(0.291,1.029)
		1.0	1.52	0.085	0.116	1.09	0.707	(0.769,1.178)
		1.5	1.45	0.087	0.154	1.48	1.132	(1.119,1.903)
	500	0.5	1.51	0.017	0.024	0.51	0.661	(0.317,0.704)
		1.0	1.50	0.017	0.017	0.98	0.657	(0.865,1.090)
		1.5	1.49	0.017	0.026	1.49	0.110	(1.324,1.767)

Application: Analysis of COVID-19 Data

The COVID-19 pandemic is an ongoing pandemic of coronavirus disease 2019 that was first identified in December 2019. Globally, as of 10 January 2021, there have been 88,120,981 confirmed cases of COVID-19, including 1,914,378 deaths, reported to WHO (https://covid19.who.int/). COVID-19 is thought to spread mainly through close contact from person to person, including between people who are physically near each other. So it seems that people infected by this disease will desire more spatial dependence, which copula functions can model. To determine the effective factors in the survival time of COVID-19 patients and also to test if there is a spatial dependence between patients, we applied the proposed model to a dataset from the Philippines on GitHub (25).

Philippines is an archipelagic country in Southeast Asia, near China. As of 10 January 2021, more than 490,000 confirmed cases, including more than 9,500 deaths, have been reported to WHO in Philippines (https://covid19.who.int/region/wpro/country/ph). The available data set on GitHub included laboratory-confirmed COVID-19 individuals who observed symptoms in the first three months after the outbreak and were followed up until recovery or death. For each patient, in addition to geographical coordinates, longitude and latitude, age, sex (1: male, 0: female), onset symptoms date, date of death or recovery, and the outcome (death, recovery or discharging) have been recorded. Since the dataset included patients with acute conditions, out of 363 cases, death was recorded for 321 patients. So The censoring percentage is about 11%. Figures 1 and 2 show the geographical dispersion of 363 patients and geo-plots of this dataset.

First, based on a parametric approach, we fitted a marginal Weibull model on survival times and considered age and sex as covariates. The results showed that patients' sex does not have a significant effect on survival time. A summary of the final fitted model from 'survreg' function in survival package in R is provided in Table 2.

The summary of 'survreg' includes the coefficients of each covariate, standard errors and P-values. The output scale corresponds to the Weibull distribution's shape parameter is shown in the last line of this table. The result of the log-likelihood test indicates the fitted model is significantly better than the null model: Loglik(model)=-1167.3, Loglik(intercept only)= -1173.6, Chisq= 12.56 on 2 degrees of freedom, and P-value=0.0019.

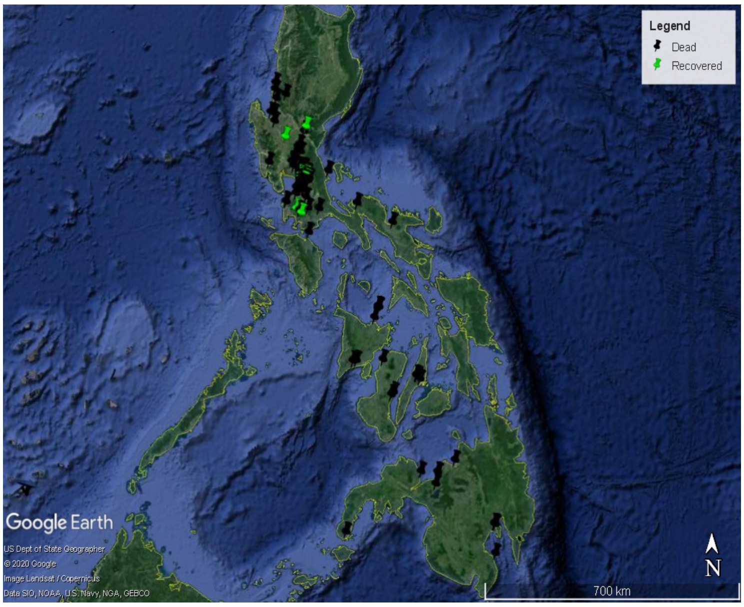
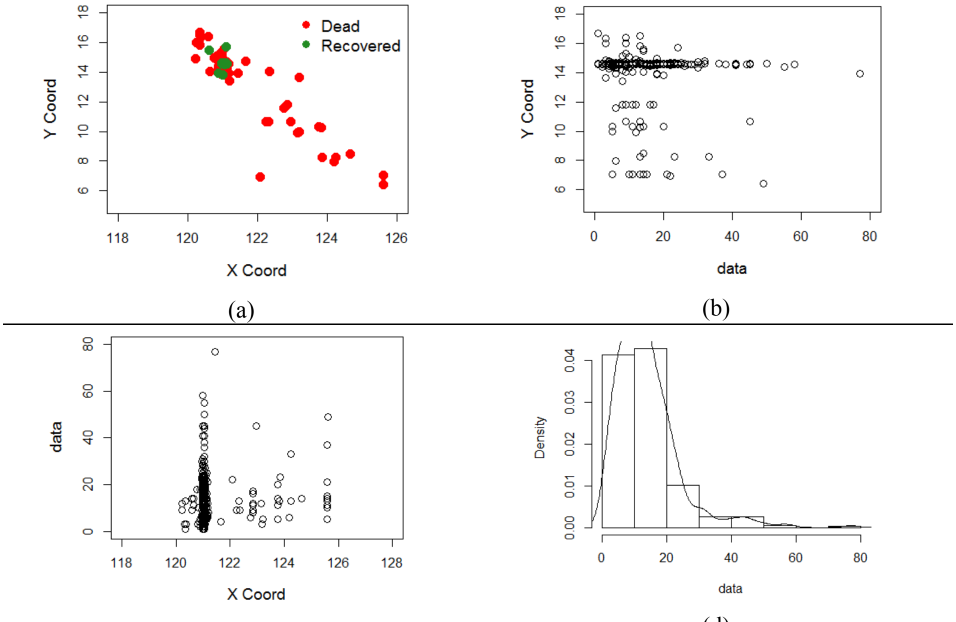

Legend
Dead
Recovered
Google Earth
US Dept of State Geographer
© 2020 Google
Image Landsat / Copernicus
Data SIO, NOAA, U.S. Navy, NGA, GEBCO
700 km
N

Figure 1. Geographical dispersion of patients in Philippines



(a)

(b)

(c)

(d)

Figure 2. Geoplots of Philippines COVID-19 data (a) data locations, (b) data against the Y coordinate, (c) data against the X coordinate, and (d) the histogram of the data values.

Despite the convenience of fitting the Weibull model and the flexibility of Weibull distribution, it has not been used widely in medical research because the estimated coefficients are not clinically meaningful (26). So the function 'ConvertWeibull' in package SurvRegCensCov was used to convert the summary of 'survreg' to more clinically relevant statistics such as hazard ratio. Using this function, the Hazard Ratio for variable age is calculated as 1.014 with (1.0059,1.0221) as its 95% confidence intervaln. It seems that the variable age has a small contribution to the hazard ratio, as an additional year of patients' age increases the risk of death by 1% (HR=1.01) and decreases the survival time by around 1%.

To model the dependence between survival times, we assumed a Gumbel-Hougaard survival copula for the

Table 2. Estimates for marginal hazard using Weibull model

	Value	Std	z	P-value
Intercept	3.4040	0.1841	18.49	<2e-16
Age	-0.0100	0.0028	-3.49	0.0005
Sex	0.0549	0.0817	0.67	0.5016
Log (scale)	-0.3629	0.0422	-8.60	<2e-16

bivariate joint survival function, allowing the copula's dependence parameter to be a function of geographic distance via its relation to Kendall-tau. Plugging the estimated regression coefficient from the marginal Weibull model into the composite likelihood function and maximizing it as a function of ψ obtained $\hat{\psi} = 0.73$. This result indicates that the dependence is reducing as the geographic distance increases. According to equation (5), a 10-km distance between two patients leads to $\hat{\tau} = 0.93$, and a 500 km distance causes $\hat{\tau} = 0.001$, which seems reasonable for a country like the Philippines.

Results and Discussion

This paper proposed a copula-based approach to model the spatially correlated survival data. We used a two-stage estimation method to estimate the regression coefficients of the marginal hazard model and the spatial dependence parameter of the joint survival function. Estimating the spatial dependence parameter was obtained by maximizing the composite likelihood function based on bivariate survival copulas. The Farlie-Gumbel-Morgenstern copula, used by (17), can only project the small dependencies. On the other hand, members of the FGM family are symmetric. Because of the skewness of the survival data and a positive dependency between the spatial data, we used a bivariate Gumbel-Hougaard copula function to consider the spatial dependence between data. MSE criteria in the simulation study indicated the proper performance of the proposed model and the estimation method based on the composite likelihood function. Applying the proposed model to Philippine's individual-level COVID-19 data allowed us to model the spatial dependence from the patients' geographic distance.

Acknowledgements

The authors thank the editorial team and the anonymous reviewers for their comments, helpful suggestions and encouragement, which helped improve the final version of this paper. Receiving support from the Center of Excellence in Analysis of Spatio-Temporal Correlated Data at Tarbiat Modares University is acknowledged.

References

1. Vaupel JW, Manton KG and Stallard E. The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. Demography. 1979; 16: 439-454.
2. Li Y and Ryan L. Modeling Spatial Survival Data Using Semiparametric Frailty Models. Biometrics. 2002; 58: 287-297.
3. Motarjem K, Mohammadzadeh M and Abyar A. Bayesian Analysis of Spatial Survival Model with Non-Gaussian Random Effect. Journal of Mathematical Sciences. 2019; 237: 692-701.
4. Henderson R, Shimakura S, Gorst D. Modeling Spatial Variation in Leukemia Survival Data. Journal of the American Statistical Association. 2002; 97: 965-972.
5. Banerjee S and Dey DK. Semiparametric Proportional Odds Models for Spatially Correlated Survival Data. Lifetime Data Analysis. 2005; 11: 175-191.
6. Zhao L, Hanson TE. and Carlin BP. Mixtures of Polya Trees for Flexible Spatial Frailty Survival Modelling. Biometrika. 2009; 96: 263-276.
7. Pan C, Cai B, Wang L and Lin X. Bayesian Semi Parametric Model for Spatial Interval-Censored Survival Data. Computational Statistics and Data Analysis. 2014; 74: 198-209.
8. Liu D, Kalbfleisch JD, Schaubel DE. A Positive Stable Frailty Model For Clustered Failure Time Data With Covariate-Dependent Frailty. Biometrics. 2014; 67(1): 8-17.
9. Sklar A. Fonctions de Répartition à n Dimensions et Leurs Marges. Publications de l'Institut de Statistique de l'Universite de Paris. 1959; 8: 229-231.
10. Nelsen RB. An Introduction to Copulas. 2005; Second Edition. Springer Series in Statistics .
11. Clayton D. A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence. Biometrika. 1978; 65: 141-151.
12. Bárdossy A. Copula-based Geostatistical Models for Groundwater Quality Parameters. Water Resources Research. 2006; 42: 1-12.
13. Shiau JT. Fitting Drought Duration and Severity with Two-Dimensional Copulas. Water Resources Research. 2006; 20: 795-815.
14. Omidi M, Mohammadzadeh M and Morid S. The Probabilistic Analysis of Drought Severity-Duration in Tehran Province using Copula Functions. Iranian Journal of Agricultural Sciences. 2010; 41: 95-102.
15. Li Y and Lin X. Semiparametric Normal Transformation Models for Spatially Correlated Survival Data. Journal of the American Statistical Association. 2006; 101: 591-603.

16.Domma F and Giordano S. A Copula-Based Approach to Account for Dependence in Stress-Strength Models. Statistical Papers. 2013; 54: 807-826.

17.Paik J, Ying Z. A Composite Likelihood Approach for Spatially Correlated Survival Data. Computational Statistics and Data Analysis. 2013; 56(1): 209-216.

18.Zhou H, Hanson T and Knapp R. Marginal Bayesian Nonparametric Model for Time to Disease Arrival of Threatened Amphibian Populations. Biometrics. 2015; 71: 1101-1110.

19.Prenen L and Braekers R. Extending the Archimedean Copula Methodology to Model Multivariate Survival Data Grouped in Clusters of Variable Size. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2017; 79: 483-505.

20.Geerdens C, Acar EF and Janssen P. Conditional Copula Models For Right-Censored Clustered Event Time Data. Biostatistics. 2018;19(2): 247-262.

21.Omidi, M and Mohammadzadeh, M. A New Method to Build Spatio-Temporal Covariance Functions: Analysis of Ozone Data. Statistical Papers. 2015; 57(3): 689–703

22.Lee EW, Wei LJ and Amato DA. Cox-Type Regression Analysis for Large Numbers of Small Groups of Correlated Failure Time Observations. Survival Analysis: State of the Art. J. P. Klein and P. Goel, eds. Boston: Kluwer Academic Publishers. 1992; 237–248.

23.Varin C, Reid N and Firth D. An Overview of Composite Likelihood Methods. Statistica Sinica. 2011; 21(1): 5-42.

24.Motarjem K, Mohammadzadeh M and Abyar A. Geostatistical Survival Model with Gaussian Random Effect. Statistical Papers. 2020; 21(1):85-107.

25.Xu B, Gutierrez B, Mekaru S, Sewalk K, Goodwin L, Loskill A, Cohn E L, Hswen Y, Hill S C, Cobo M M, Zarebski A E, Li S, Wu C, Hulland E, Morgan J D, Wang L, O'Brien K, Scarpino S V, Brownstein J S, Pybus O G, Pigott D M and Kraemer M U G. Epidemiological Data From the COVID-19 Outbreak, Real-Time Case Information. Scientific Data. 2020; 7 (1): 106.

26.Zhang Z. Parametric Regression Model for Survival Data: Weibull Regression Model as an Example. Annals of translational medicine. 2016; 4 (24): 484.