Journal of Data Science 19 (1), 37–55
January 2021
DOI: 10.6339/21-JDS997
Statistical Data Science

A Vine Copula Model for Climate Trend Analysis using Canadian Temperature Data

Haoxin Zhuang[1], Liqun Diao[1,*], and Grace Y. Yi[2]

[1]*Department of Statistics and Actuarial Science, University of Waterloo, ON, Canada*
[2]*Department of Statistical and Actuarial Science, Department of Computer Science, University of Western Ontario, ON, Canada*

Abstract

Climate change is widely recognized as one of the most challenging, urgent and complex problem facing humanity. There are rising interests in understanding and quantifying climate changing. We analyze the climate trend in Canada using Canadian monthly surface air temperature, which is longitudinal data in nature with long time span. Analysis of such data is challenging due to the complexity of modeling and associated computation burdens. In this paper, we divide this type of longitudinal data into time blocks, conduct multivariate regression and utilize a vine copula model to account for the dependence among the multivariate error terms. This vine copula model allows separate specification of within-block and between-block dependence structure and has great flexibility of modeling complex association structures. To release the computational burden and concentrate on the structure of interest, we construct composite likelihood functions, which leave the connecting structure between time blocks unspecified. We discuss different estimation procedures and issues regarding model selection and prediction. We explore the prediction performance of our vine copula model by extensive simulation studies. An analysis of the Canada climate dataset is provided.

Keywords *climate change; composite likelihood; longitudinal data; prediction*

1 Introduction

Climate change is widely recognized as one of the most challenging, urgent and complex problem facing humanity (OBrien, 2010). Its impacts are global in scope and unprecedented in scale, its fingerprints are across natural systems (Parmesan and Yohe, 2003), and it exposes human-being under the risks of but not limited to global food security (Wheeler and von Braun, 2013) and forced immigration (Hugo, 2013). Climate change features global warming since the mid-20th century and it is believed that human activities are responsible for the observed waming (Stocker et al., 2013). Therefore, it is increasingly important to understand and quantify climate change and the magnitude and the speed of global warming, which provides a basis for policy makers and financial institutions to respond smartly (Lim et al., 2004; Fang et al., 2019).

The objective of our research is to develop a new statistical model to better characterize and forecast the trend of temperature changing. The temperature data is usually longitudinal in nature and features long time span, which imposes challenges to conventional method for longitudinal data analysis.

Longitudinal data analysis, which studies the change of repeated observations of the same subjects over time, has long been a thriving topic in statistical research. There have been sev-

*Corresponding author. Email: l2diao@uwaterloo.ca.

Received July, 2020; Accepted December, 2020

eral approaches in this field, including multivariate analysis, linear and generalized linear mixed and mixture models, generalized estimating equations, structural equation models, transition methods, Bayesian methods and so on. A large body of books, papers and reviews discussed and summarized the aforementioned topics, and for comprehensive summaries of different approaches, we can refer to Diggle et al. (2002), Hedeker and Gibbons (2006), Fitzmaurice et al. (2009), Verbeke and Molenberghs (2009), Verbeke et al. (2014), for example. Analysis of intensive longitudinal data (Walls and Schafer, 2006) is also an emerging area.

Copula (Joe, 1997; Nelsen, 2007) is a powerful and flexible tool to model the multivariate distribution and it allows separate models for marginal distribution and dependence structure. To cope with the restrictions of multivariate copula, a graphical model, vine copula, (Joe, 1997; Bedford and Cooke, 2002; Aas et al., 2009) was developed based on density decomposition and bivariate copulas and it can model the multivariate distribution flexibly.

The applications of copulas and vine copulas to longitudinal data are limited. Lambert and Vandenhende (2002) introduced copula to model multivariate non-normal longitudinal data. Smith et al. (2010) considered using D-Vine copula to model the serial dependence in time series, but they focused more on the estimation of the vine copulas and did not include covariates into the model. Ruscone and Osmetti (2016) and Smith (2015) consider using copula and vine copula to model the multivariate time series. Killiches and Czado (2018) considered modeling the unbalanced longitudinal data with a homogeneous vine copula model. Each bivariate copula in the vine structure is assumed to have the form of Gaussian copula, so that the model can be used to make prediction easily. Other studies include Frees and Wang (2006); Shen and Weissfeld (2006); Domma et al. (2009); Madsen and Fang (2011); Shi and Yang (2018). Most of these references considered a short time span, or used model selection methods to create sparse vine structure, such as Smith et al. (2010).

In this paper, we use vine copula model to describe the dependence structure of longitudinal data with possible long time span by dividing data into different time blocks. The temporal length of the longitudinal data determines the number of parameters in regular vine model, which increases quadratically as time length increases. Thus directly using vine copula model for the longitudinal data on a large time-span will introduce a large number of parameters and hence create difficulties for parameter estimation. As a result, we consider using the composite likelihood (Lindsay, 1988; Varin, 2008; Varin et al., 2011; Lindsay et al., 2011; Yi, 2017) to simplify the likelihood function and concentrate on the parameters of primary interest. We also compare different estimation procedures, simultaneous estimation and two-stage estimation, to further facilitate the fast inference of our proposed model. Moreover, we find out in simulation studies that the composite likelihood provides robustness against misspecification on structure linking between time blocks, accurate selection of the (conditional) bivariate copulas and convenient structure for prediction. The proposed model yields promising prediction results in terms of subject and time extrapolations in both simulation studies and analysis of Canadian temperature data.

The rest of the paper is organized as follows. In Section 2, we discuss the model formulation, including marginal model and association model. In Section 3, we describe how to estimate the parameters, and in Section 4, we give the procedure for copula selection and prediction based on our model. In Section 5 and 6, simulation studies and analysis of Canadian temperature data are provided, respectively.

2 Model Formulation

Suppose that we are interested in modeling longitudinal data collected over a long period of time, for instance, monthly temperature data over years. Such data usually features with natural period (e.g., years) or exhibits a periodic pattern. To feature the periodic patterns, we examine the data by periods, called time blocks in what follows, and let b denote the number of time points in each time blocks. Suppose that we have a time blocks, let $m = ab$ denote the total number of observed occasions, and n subjects are observed at the m occasions. For longitudinal data with no periodic pattern, we set $a = 1$. Let Y_{ikl} be the continuous response for the ith subject at the lth time point in the kth time block, and let x_{ikl} be the associated covariate matrices. Let $Y_{ik} = (Y_{ik1}, \ldots, Y_{ikb})^{\mathrm{T}}$ be the vector of responses of the ith subject in the kth time block, and let $Y_i = (Y_{i1}^{\mathrm{T}}, \ldots, Y_{ia}^{\mathrm{T}})^{\mathrm{T}}$ be the full vector of responses of subject i for $i = 1, \ldots, n$ and $k = 1, \ldots, a$. Let lower case letters y_{ik} and y_i denote the realizations of of Y_{ik} and Y_i, respectively, and let x_{ik} and x_i denote the corresponding covariates.

We now introduce the joint model for Y_i which shows the dependence of Y_i on x_i. It is difficult to directly specify a meaningful joint distribution of Y_i, given x_i, to facilitate the dependence structure of the components of Y_i. To come up with an interpretable joint model for Y_i given x_i, we take two steps. In the first step, we characterize the dependence of Y_i on x_i via regression models, which contain random errors; in the second step, we further delineate the dependence structures of the components of Y_i by characterizing the dependence structures of the random errors resulted from the first step.

Specifically, for $i = 1, \ldots, n$, $k = 1, \ldots, a$, and $l = 1, \ldots, b$, we assume that

$$Y_{ikl} = \mu_{ikl} + \varepsilon_{ikl}, \tag{1}$$

where $\mu_{ikl} = E(Y_{ikl}|x_{ikl})$, and ε_{ikl} is the associated random error term. We further assume that

$$g_l(\mu_{ikl}) = x_{ikl}^{\mathrm{T}}\beta_l,$$

where $g_l(\cdot)$ is the link function and β_l is the parameter vector associated with time l. Let $\beta = (\beta_1^{\mathrm{T}}, \ldots, \beta_b^{\mathrm{T}})^{\mathrm{T}}$. For $i = 1, \ldots, n$ and $k = 1, \ldots, a$, we let $\varepsilon_{ik} = (\varepsilon_{ik1}, \ldots, \varepsilon_{ikb})^{\mathrm{T}}$ and $\varepsilon_i = (\varepsilon_{i1}^{\mathrm{T}}, \ldots, \varepsilon_{ia}^{\mathrm{T}})^{\mathrm{T}}$.

To reflect that responses from the same subject across time points are possibly associated, in the next step, we focus on characterizing the dependence structure among the components of ε_i using vine copula models.

2.1 Joint Distribution of ε_i

2.1.1 Marginal Distribution of ε_i

For $l = 1, \ldots, b$, we assume that marginally, the random errors $\{\varepsilon_{ikl} : i = 1, \ldots, n; k = 1, \ldots, a\}$ share the same distribution function and let $F_l(\cdot; \omega_l)$ and $f_l(\cdot; \omega_l)$, respectively, denote their cumulative distribution function (CDF) and the density function indexed by parameter vector ω_l, i.e.,

$$\varepsilon_{ikl} \sim F_l(\varepsilon_{ikl}; \omega_l),$$

for $i = 1, \ldots, n$; $k = 1, \ldots, a$. Let $\omega = (\omega_1^{\mathrm{T}}, \ldots, \omega_b^{\mathrm{T}})^{\mathrm{T}}$ and let $\eta = (\beta^{\mathrm{T}}, \omega^{\mathrm{T}})^{\mathrm{T}}$ denote the parameter vector associated with the marginal distribution of the Y_{ikl}.

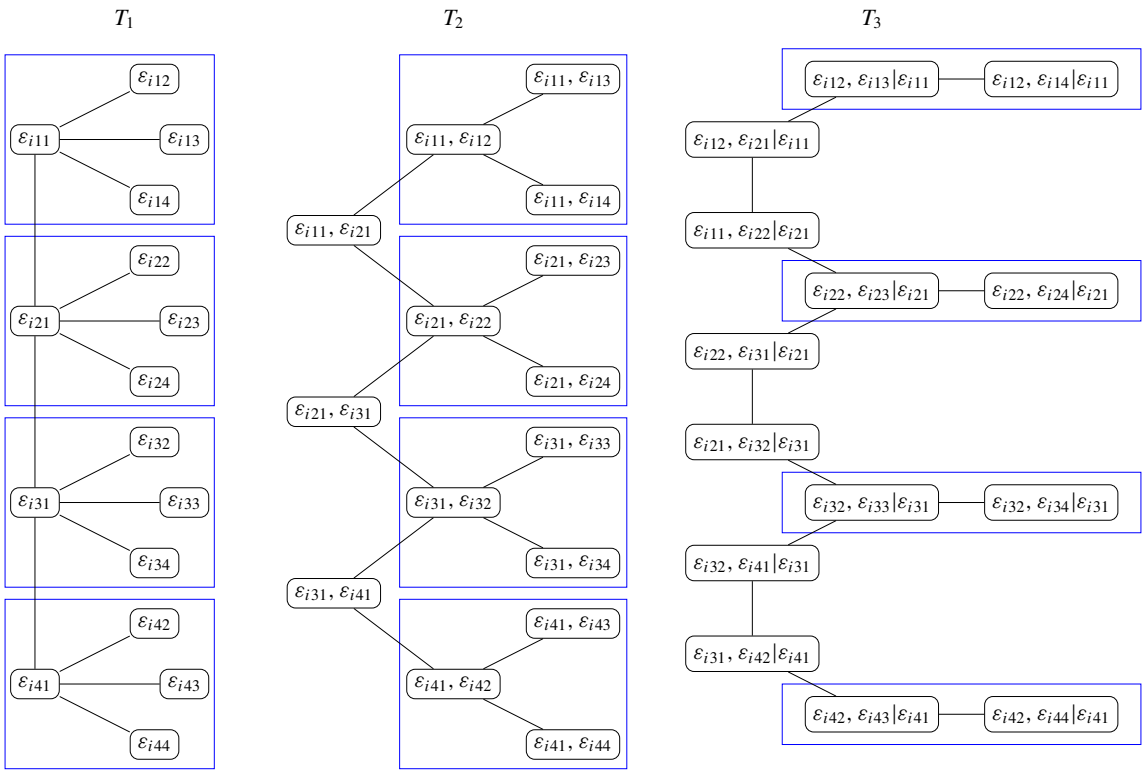

Figure 1: A R-Vine structure for 4 time blocks and 4 time points within each block.

2.1.2 Dependence Structure of ε_i

We employ vine copula models (Bedford and Cooke, 2002) to delineate the dependence structures of the random vector ε_i. In particular, D-Vine and Canonical vine (C-Vine) are two useful cases of regular vine copula models, which pertain to pair-copula constructions (Aas et al., 2009). A brief review of the idea of vine copula model is provided in Section 1 of the supplementary materials.

As longitudinal data has a natural temporal order, Smith et al. (2010) and Killiches and Czado (2018) both considered modeling the longitudinal data using a D-Vine structure under different settings. However, in the second or higher levels of D-Vine trees, describing the stochastic behavior of the current responses needs to be conditional on future responses, which creates difficulties in interpreting the copula parameters. A good property of C-Vine is that if a non-dominating variable (i.e., a variable that is not the root of the tree T_1) is dropped, the remaining variables still follow a C-Vine structure. As a result, we adopt a C-Vine to model the dependence structure between different time points within a block to avoid this problem and yield an interpretable model.

Specifically, we propose to use an R-Vine structure (Bedford and Cooke, 2002) to model the dependence structures within ε_i. Within each time block, the dependence structure between time points is assumed to be identical and modeled with a C-Vine structure; and different time blocks are connected by a D-Vine structure. To illustrate this idea, in Figure 1 we present an example with 4 time blocks and 4 time points within each block, where T_1, T_2 and T_3 represent the first three levels of trees in the vine copula model, and the nodes in the (blue) boxes represent the error terms of time points within time blocks, which have a C-Vine model structure.

We first introduce necessary notation before we give the mathematical form of the R-Vine structure. For $c = 1, \ldots, a$ and $d = 2, \ldots, b+1$, let $\mathcal{G}_{icd} = \{\varepsilon_{icl} : l = 1, \ldots, d-1\}$. For $s, g \in \{1, \ldots, a\}$ and $h, r \in \{1, \ldots, b\}$, let

$$\mathcal{D}_{ish,igr} = \begin{cases} \left\{ \bigcup_{c=s+1}^{g-1} \mathcal{G}_{ic(b+1)} \right\} \cup \mathcal{G}_{ish} \cup \mathcal{G}_{igr}, & \text{if } s < g-1; \\ \mathcal{G}_{ish} \cup \mathcal{G}_{igr}, & \text{if } s = g-1; \\ \mathcal{G}_{ish}, & \text{if } s = g \text{ and } h < r. \end{cases}$$

Furthermore, for a random variable Z_1, a random vector $Z_2 = (Z_{21}, \ldots, Z_{2d_1})^{\mathrm{T}}$ and a random vector $Z_3 = (Z_{31}, \ldots, Z_{3d_2})^{\mathrm{T}}$ with $1 + d_1 + d_2 = m$, let $F_{\mathrm{Z}_1\mathrm{Z}_2\mathrm{Z}_3}(z_1, z_2, z_3)$ denote the joint CDF of Z_1,Z_2 and Z_3, with $f_{\mathrm{Z}_1\mathrm{Z}_2\mathrm{Z}_3}$ as the corresponding density function. As a result, the joint density of the random vector Z_2 is derived as $f_{\mathrm{Z}_2}(z_2) = \int\int f_{\mathrm{Z}_1\mathrm{Z}_2\mathrm{Z}_3}(z_1, z_2, z_3) dz_1 dz_3$, and the conditional CDF of Z_1, given Z_2 is

$$F_{\mathrm{Z}_1|\mathrm{Z}_2}(z_1|z_2) = \frac{\partial^{d_1} F_{\mathrm{Z}_1\mathrm{Z}_2}(z_1, z_2)}{\partial z_{21} \ldots \partial z_{2d_1}} \frac{1}{f_{\mathrm{Z}_2}(z_2)}, \tag{2}$$

where $F_{\mathrm{Z}_1\mathrm{Z}_2}(z_1, z_2) = \lim_{z_3 \to \infty} F_{\mathrm{Z}_1\mathrm{Z}_2\mathrm{Z}_3}(z_1, z_2, z_3)$ is the joint CDF of Z_1 and Z_2.

For ε_{ikh} and ε_{ikr} with $h < r$ in the same time block k, let $c_{kh,kr}(\cdot, \cdot)$ denote the conditional copula density function between ε_{ikh} and ε_{ikr}, given the conditioning set $\mathcal{D}_{ikh,ikr}$, where the first and second arguments in the copula density are given by $u_{ikh|\mathcal{D}_{ikh,ikr}} = F_{\varepsilon_{ikh}|\mathcal{D}_{ikh,ikr}}(\varepsilon_{ikh}|\mathcal{D}_{ikh,ikr})$ and $u_{ikr|\mathcal{D}_{ikh,ikr}} = F_{\varepsilon_{ikr}|\mathcal{D}_{ikh,ikr}}(\varepsilon_{ikr}|\mathcal{D}_{ikh,ikr})$ respectively, and $F_{\varepsilon_{ikh}|\mathcal{D}_{ikh,ikr}}$ and $F_{\varepsilon_{ikr}|\mathcal{D}_{ikh,ikr}}$ are the conditional CDFs of ε_{ikh} and ε_{ikr}, given the conditioning set $\mathcal{D}_{ikh,ikr}$ respectively, which are obtained from (2) by letting $Z_1 = \varepsilon_{ikh}$ or ε_{ikr}, $Z_2 = \mathcal{D}_{ikh,ikr}$ and $Z_3 = \varepsilon_i \backslash \{\varepsilon_{ikh} \cup \mathcal{D}_{ikh,ikr}\}$ or $\varepsilon_i \backslash \{\varepsilon_{ikr} \cup \mathcal{D}_{ikh,ikr}\}$.

For ε_{ish} and ε_{igr} in different time block with $s < g$, let $c_{sh,gr}(\cdot, \cdot)$ denotes the conditional copula density function between ε_{ish} and ε_{igr}, given the conditioning set $\mathcal{D}_{ish,igr}$, where the first and second arguments in the copula density are given by $u_{ish|\mathcal{D}_{ish,igr}} = F_{\varepsilon_{ish}|\mathcal{D}_{ish,igr}}(\varepsilon_{ish}|\mathcal{D}_{ish,igr})$ and $u_{igr|\mathcal{D}_{ish,igr}} = F_{\varepsilon_{igr}|\mathcal{D}_{ish,igr}}(\varepsilon_{igr}|\mathcal{D}_{ish,igr})$ respectively, and $F_{\varepsilon_{ish}|\mathcal{D}_{ish,igr}}$ and $F_{\varepsilon_{igr}|\mathcal{D}_{ish,igr}}$ are the conditional CDFs of ε_{ish} and ε_{igr}, given the conditioning set $\mathcal{D}_{ish,igr}$ respectively, which are obtained from (2) by letting $Z_1 = \varepsilon_{ish}$ or ε_{igr}, $Z_2 = \mathcal{D}_{ish,igr}$ and $Z_3 = \varepsilon_i \backslash \{\varepsilon_{ish} \cup \mathcal{D}_{ish,igr}\}$ or $\varepsilon_i \backslash \{\varepsilon_{igr} \cup \mathcal{D}_{ish,igr}\}$.

Combining the marginal model and the dependence structures specified, we write the joint density function of ε_i as

$$\begin{aligned} f(\varepsilon_i; \omega, \theta, \psi) = & \left\{ \prod_{k=1}^{a} \prod_{l=1}^{b} f_l(\varepsilon_{ikl}; \omega_l) \right\} \\ & \times \left\{ \prod_{k=1}^{a} \prod_{h=1}^{b-1} \prod_{r=h+1}^{b} c_{kh,kr}(u_{ikh|\mathcal{D}_{ikh,ikr}}, u_{ikr|\mathcal{D}_{ikh,ikr}}; \theta_{kh,kr}) \right\} \\ & \times \left\{ \prod_{s=1}^{a-1} \prod_{g=s+1}^{a} \prod_{h=1}^{b} \prod_{r=1}^{b} c_{sh,gr}(u_{ish|\mathcal{D}_{ish,igr}}, u_{igr|\mathcal{D}_{ish,igr}}; \psi_{sh,gr}) \right\}, \end{aligned} \tag{3}$$

where the product in the first set of brackets corresponds to the marginal densities of the ε_{ikl}, the product in the second set of brackets corresponds to the C-Vine structure within time blocks

indexed by the dependence parameter vector $\theta = \{\theta_{kh,kr} : k = 1, \ldots, a; h = 1, \ldots, b-1; r = (h+1), \ldots, b\}$, and the product in the third set of brackets corresponds to the D-Vine structure connecting the time blocks indexed by the dependence parameter vector $\psi = \{\psi_{sh,gr} : s = 1, \ldots, a-1; g = s+1, \ldots, a; h, r = 1, \ldots, b\}$. Let $\vartheta = (\theta^{\mathrm{T}}, \psi^{\mathrm{T}})^{\mathrm{T}}$ denote the vector of dependence parameters.

2.2 Joint Model of the Responses Y_i

Applying the one-to-one transformation to the random variables defined by (1) in combination with the joint density function (3) for ε_i, we obtain the joint distribution of responses Y_i, given by

$$\begin{aligned} f(y_i; \eta, \vartheta) = & \prod_{k=1}^{a}\prod_{l=1}^{b} f_l(y_{ikl} - g^{-1}(x_{ikl}^{\mathrm{T}}\beta_l); \omega_l) \\ & \times \prod_{k=1}^{a}\prod_{h=1}^{b-1}\prod_{r=h+1}^{b} c_{kh,kr}(u_{ikh|\mathcal{D}_{ikh,ikr}}, u_{ikr|\mathcal{D}_{ikh,ikr}}; \theta_{kh,kr}) \\ & \times \prod_{s=1}^{a-1}\prod_{g=s+1}^{a}\prod_{h=1}^{b}\prod_{r=1}^{b} c_{sh,gr}(u_{ish|\mathcal{D}_{ish,igr}}, u_{igr|\mathcal{D}_{ish,igr}}; \psi_{sh,gr}), \end{aligned} \tag{4}$$

where $u_{ish|\mathcal{D}_{ish,igr}} = F_{\varepsilon_{ish}|\mathcal{D}_{ish,igr}}(\varepsilon_{ish}|\mathcal{D}_{ish,igr})$ in (3) is now expressed as

$$u_{ish|\mathcal{D}_{sh,gr}} = F_{\varepsilon_{ish}|\mathcal{D}_{ish,igr}}\left(y_{ish} - g^{-1}(x_{ish}^{\mathrm{T}}\beta_h)|\mathcal{D}_{ish,igr}\right)$$

by using (1).

3 Estimation Methods

Given the availability of the joint distribution of Y_i, it is natural to use the likelihood method to estimate the marginal parameters η and dependence parameters ϑ simultaneously. Let

$$L_i(\eta, \vartheta) = f(y_{i11}, \ldots, y_{iab}; \eta, \vartheta)$$

be the likelihood contributed from subject i. Then the full likelihood is

$$L(\eta, \vartheta) = \prod_{i=1}^{n} L_i(\eta, \vartheta). \tag{5}$$

Maximizing the likelihood function (5) with respect to η and ϑ gives the maximum likelihood estimator of $(\eta^{\mathrm{T}}, \vartheta^{\mathrm{T}})^{\mathrm{T}}$, denoted by $(\hat{\eta}^{\mathrm{T}}, \hat{\vartheta}^{\mathrm{T}})^{\mathrm{T}}$.

The likelihood method is conceptually easy to implement, and it yields consistent and efficient estimators if the associated models are correctly specified. However, this method has two major limitations. Computationally, when the dimension of Y_i increases, the number of parameters in the likelihood function will increase dramatically, and thus, using the likelihood for estimation can be computationally prohibitive. Theoretically, the validity of the maximum likelihood estimator hinges on the correctness of all the assumed models. Any model misspecification may result in biased results.

To overcome the weakness of the likelihood method, we explore the alternative estimation methods using the composite likelihood framework (Lindsay, 1988; Varin, 2008; Varin et al., 2011; Lindsay et al., 2011; Yi, 2017), of which a review is provided in Section 2 of the supplementary materials and the details of formulation are elaborated in following sections.

3.1 Simultaneous Estimation with Composite Likelihood

Rather than working with the joint distribution of Y_i in (4), we ignore the dependence structure between time blocks. This ignorance is driven by the fact that the parameters ψ, which consists mostly of the parameters in high levels of R-Vine tree, are not of primary interest (Brechmann et al., 2012).

Let $\phi = (\eta^{\mathrm{T}}, \theta^{\mathrm{T}})^{\mathrm{T}}$, we consider the joint distribution of Y_{ik} for subject i within the kth time block

$$f(y_{ik1}, \ldots, y_{ikb}; \phi) = \prod_{l=1}^{b} f_l(y_{ikl} - g_l^{-1}(x_{ikl}^{\mathrm{T}}\beta_l); \omega_l) \times \prod_{h=1}^{b-1} \prod_{r=h+1}^{b} c_{kh,kr}(u_{ikh|\mathcal{D}_{kh,kr}}, u_{ikr|\mathcal{D}_{ikh,ikr}}; \theta_{kh,kr}), \tag{6}$$

for $i = 1, \ldots, n$ and $k = 1, \ldots, a$. This distribution form is simpler than (4).

Next, we formulate a composite likelihood for the parameters ϕ using (6) and ignoring the dependence among different time blocks:

$$L_c(\phi) = \prod_{i=1}^{n} L_{ci}(\phi), \tag{7}$$

where $L_{ci}(\phi) = \prod_{k=1}^{a} f(y_{ik1}, \ldots, y_{ikb}; \phi)$. Maximizing (7) with respect to ϕ yields a composite maximum likelihood estimator of ϕ, denoted by $\hat{\phi}_{\mathrm{CS}}$.

Under regularity conditions (Varin, 2008; Varin et al., 2011; Yi, 2017), $\hat{\phi}_{\mathrm{CS}}$ is a consistent estimator of ϕ, and $\sqrt{n}(\hat{\phi}_{\mathrm{CS}} - \phi)$ has the asymptotic normal distribution with mean zero and covariance matrix $H_{\mathrm{CS}}(\phi)J_{\mathrm{CS}}^{-1}(\phi)H_{\mathrm{CS}}\phi)$, where

$$H_{\mathrm{CS}}(\phi) = E\left(\frac{\partial^2 L_{ci}(\phi)}{\partial\phi\partial\phi^{\mathrm{T}}}\right), \text{ and } J_{\mathrm{CS}}(\phi) = E\left[\left(\frac{\partial L_{ci}(\phi)}{\partial\phi}\right)\left(\frac{\partial L_{ci}(\phi)}{\partial\phi}\right)^{\mathrm{T}}\right].$$

Matrices $H_{\mathrm{CS}}(\phi)$ and $J_{\mathrm{CS}}(\phi)$ can be estimated by their empirical counterparts with $\hat{\phi}_{cs}$ plugged in, when using the asymptotic distribution to conduct inference about ϕ.

3.2 Two-Stage Estimation with Composite Likelihood

To further ease computation burdens, we treat η and θ differently when employing (7) for estimation. Specifically, we estimate η using a simpler formulation than (7) and then use (7) to estimate θ only.

We now describe a two-stage estimation procedure. In the first stage, for $l = 1, \ldots, b$, we construct the marginal likelihood functions for marginal parameters $\eta_l = (\beta_l^{\mathrm{T}}, \omega_l^{\mathrm{T}})^{\mathrm{T}}$,

$$L_l(\eta_l) = \prod_{i=1}^{n} L_{il}(\eta_l), \tag{8}$$

where $L_{il}(\eta_l) = \prod_{k=1}^{a} f_l\big(y_{ikl} - g_l^{-1}(x_{ikl}^{\mathrm{T}}\beta_l); \omega_l\big)$. Maximizing (8) with respect to η_l yields an estimator of η_l, denoted by $\hat{\eta}_l$, for $l = 1, \ldots, b$. Let $\hat{\eta}_{\mathrm{CT}} = (\hat{\eta}_1^{\mathrm{T}}, \ldots, \hat{\eta}_b^{\mathrm{T}})^{\mathrm{T}}$. In the second stage, we plug $\hat{\eta}_{\mathrm{CT}}$ into (7) and obtain $L_c(\hat{\eta}_{\mathrm{CT}}, \theta)$. Then maximizing $L_c(\hat{\eta}_{\mathrm{CT}}, \theta)$ with respect to θ provides an estimator of θ, denoted by $\hat{\theta}_{\mathrm{CT}}$. Let $\hat{\phi}_{\mathrm{CT}} = (\hat{\eta}_{\mathrm{CT}}^{\mathrm{T}}, \hat{\theta}_{\mathrm{CT}}^{\mathrm{T}})^{\mathrm{T}}$.

Let $Q_i(\eta) = \frac{\partial}{\partial \eta} \sum_{l=1}^{b} \log[L_{il}(\eta_l)]$ and $U_i(\eta, \theta) = \frac{\partial}{\partial \theta} \log[L_{ci}(\eta, \theta)]$. Define

$$H_{\mathrm{CT}}(\phi) = E\begin{pmatrix} \frac{\partial}{\partial \eta^{\mathrm{T}}} Q_i(\eta) & 0 \\ \frac{\partial}{\partial \eta^{\mathrm{T}}} U_i(\eta, \theta) & \frac{\partial}{\partial \theta^{\mathrm{T}}} U_i(\eta, \theta) \end{pmatrix} \quad \text{and } J_{\mathrm{CT}}(\phi) = E[W_i(\eta, \theta) W_i(\eta, \theta)^{\mathrm{T}}],$$

where $W_i(\eta, \theta) = (Q_i(\eta)^{\mathrm{T}}, U_i(\theta, \eta)^{\mathrm{T}})^{\mathrm{T}}$. Similarly, by the results of Varin (2008); Varin et al. (2011) and Yi (2017), under regularity conditions, $\hat{\phi}_{\mathrm{CT}}$ is a consistent estimator of ϕ, and $\sqrt{n}(\hat{\phi}_{\mathrm{CT}} - \phi)$ has the asymptotic normal distribution with mean zero and covariance matrix $H_{\mathrm{CT}}(\phi) J_{\mathrm{CT}}^{-1}(\phi) H_{\mathrm{CT}}(\phi)$. When conducting inference about ϕ using this asymptotic distribution, $H_{\mathrm{CT}}(\phi)$ and $J_{\mathrm{CT}}(\phi)$ are estimated by their empirical counterparts with $\hat{\phi}_{ct}$ plugged in.

4 Copula Selection and Prediction

Dissmann et al. (2013) proposed a sequential procedure which selects copula forms for each of the (conditional) bivariate copulas level by level, where the selection is carried out with a prespecified vine structure from a set of candidate copula functions. The sequential procedure facilitates a fast model selection process by considering each (conditional) pair separately. In the same spirit of the composite likelihood formulation (7), we assume the same dependence structures within time blocks and ignore the dependence between blocks. Pretending to have $n \times a$ independent time blocks, we apply sequential selection procedure of Dissmann et al. (2013) to select copula functions in the C-Vine structure within blocks.

We are interested in predicting the observations for a subject in the study for a future time point (i.e., time extrapolation) or for some new subjects at a given time point (i.e., subject extrapolation). Please see supplementary materials for our discussion on subject extrapolation through simulation studies and data analysis. We focus on the time extrapolation in this subsection.

Suppose that for subject i, at time block k, the observations for all time points $j \leqslant h$ have been observed, and we would like to predict the observation at time $(h+1)$, where h is a given time point. First, the estimate of the mean for the marginal model is calculated as

$$\hat{\mu}_{ikl} = g_l^{-1}(x_{ikl}^{\mathrm{T}} \hat{\beta}_l)$$

for $l = 1, \ldots, (h+1)$. Then, the error terms of the h observed time points can be calculated and transformed as "pseudo-observations", i.e., for $l = 1, \ldots, h$,

$$\hat{\varepsilon}_{ikl} = y_{ikl} - \hat{\mu}_{ikl} \quad \text{and} \quad \hat{u}_{ikl} = F_l(\hat{\varepsilon}_{ikl}; \hat{\omega}_l).$$

Next, the conditional distribution of the error term at time $(h+1)$ can be approximated as

$$f(\varepsilon_{ik(h+1)} | \hat{\varepsilon}_{ik1}, \ldots, \hat{\varepsilon}_{ikh}) = \frac{f(\hat{\varepsilon}_{ik1}, \ldots, \hat{\varepsilon}_{ikh}, \varepsilon_{ik(h+1)})}{f(\hat{\varepsilon}_{ik1}, \ldots, \hat{\varepsilon}_{ikh})},$$

which by (3), is equal to

$$\frac{f(\hat{\varepsilon}_{ik1}, \ldots, \hat{\varepsilon}_{ikh}) f_{h+1}(\varepsilon_{ik(h+1)}) \prod_{r=1}^{h} c_{kr,k(h+1)}(\hat{u}_{ikr|\mathcal{D}_{ikr,ik(h+1)}}, u_{ik(h+1)|\mathcal{D}_{ikr,ik(h+1)}})}{f(\hat{\varepsilon}_{ik1}, \ldots, \hat{\varepsilon}_{ikh})}$$

$$= f_{h+1}(\varepsilon_{ik(h+1)}; \hat{\omega}_{h+1}) \prod_{r=1}^{h} c_{kr,k(h+1)}(\hat{u}_{ikr|\mathcal{D}_{ikr,ik(h+1)}}, u_{ik(h+1)|\mathcal{D}_{ikr,ik(h+1)}}), \tag{9}$$

where the conditional terms $\hat{u}_{ikr|\mathcal{D}_{ikr,ik(h+1)}}$ and $u_{ik(h+1)|\mathcal{D}_{ikr,ik(h+1)}}$ are calculated by applying the formulas $u_{p|q} = \frac{\partial c_{pq}(u_p,u_q)}{\partial u_q}$ and $u_{q|p} = \frac{\partial c_{pq}(u_p,u_q)}{\partial u_p}$ iteratively, in which p and q can be any unconditional label, such as ikr, or conditional label, such as $ikr|\mathcal{D}_{ikr,ik(h+1)}$. As a result, the predicted outcome $\hat{y}_{ik(h+1)}$ for subject i at time point $(h+1)$ in time block k is given by

$$\begin{aligned}\hat{y}_{ik(h+1)} &= E(\varepsilon_{ik(h+1)}|\hat{\varepsilon}_{ik1}, \ldots, \hat{\varepsilon}_{ikh}) + \hat{\mu}_{ik(h+1)} \\ &= \int_{-\infty}^{\infty} \varepsilon_{ik(h+1)} f(\varepsilon_{ik(h+1)}|\hat{\varepsilon}_{ik1}, \ldots, \hat{\varepsilon}_{ikh}) d\varepsilon_{ik(h+1)} + \hat{\mu}_{ik(h+1)}\end{aligned}$$

with $f(\varepsilon_{ik(h+1)}|\hat{\varepsilon}_{ik1}, \ldots, \hat{\varepsilon}_{ikh})$ determined by (9). The prediction variance of $\hat{y}_{ik(h+1)}$ is calculated as

$$\mathrm{Var}(\hat{y}_{ik(h+1)}) = \mathrm{Var}(\varepsilon_{ik(h+1)})/(k-1) + \mathrm{Var}(\varepsilon_{ik(h+1)}|\hat{\varepsilon}_{ik1}, \ldots, \hat{\varepsilon}_{ikh}),$$

where the first component is related to the marginal model at time $h+1$, and the second component can be calculated from the conditional density $f(\varepsilon_{ik(h+1)}|\hat{\varepsilon}_{ik1}, \ldots, \hat{\varepsilon}_{ikh})$.

5 Simulation Studies

We conduct extensive simulation studies to examine the finite sample performance of the proposed composite likelihood under simultaneous and two-stage estimation procedures. To save space, the details of the simulation designs, evaluation measures, simulation results concerning efficiency, robustness, model selection and subject extrapolation (making prediction for a sample out of training set) are reported in supplementary materials. Here we summarize the simulation results which confirm that the proposed methods yield consistent estimates, with negligible empirical biases, fairly agreeable empirical standard errors and asymptotic standard errors and good coverage rates of 95% confidence intervals. The simultaneous procedure incurs moderate efficiency loss, compared to the full likelihood method, as expected. It is more efficient than the two-stage estimation procedure, at the price of a longer computational time. The composite likelihood provides robustness against misspecification on structure linking between time blocks and accurate selection of the (conditional) bivariate copulas. The prediction performance for subject extrapolation is similar to that for time extrapolation, which is discussed in detail as follows.

We elaborate the studies to evaluate the time extrapolation (predicting for a future time) using the proposed R-Vine model and compare it to that of the conventional regression models and time-series models here. We consider various settings in Section 5.1, and report our findings in Section 5.2, respectively.

5.1 Simulation Settings

We simulate 200 datasets of the sample size $n = 500$. The covariates x_{ikl} are generated independently from the uniform distribution on $[0, 5]$ for $i = 1, \ldots, n$; $k = 1, 2, 3, 4, 5$; and $l = 1, 2, 3, 4$. The marginal model is

$$Y_{ikl} = \beta_{0l} + \beta_{1l} x_{ikl} + \beta_{2l} k + \varepsilon_{ikl}, \tag{10}$$

where $\varepsilon_{ikl} \sim N(0, \sigma_l^2)$, for $i = 1, \ldots, n$, $k = 1, 2, 3, 4, 5$ and $l = 1, 2, 3, 4$.

In this subsection, we assume the error terms bear the R-Vine structure as demonstrated in Figure 1 and we further assume the conditional independence in tree structure T_4 and beyond for simplicity. We consider two scenarios where the dependence is either strong or moderate. For the scenario of strong or moderate dependence, the (conditional) bivariate copulas connecting the time blocks in T_1, T_2 and T_3 are all Gaussian(0.8) or Gaussian(0.5), a Gaussian copula with the parameter value shown in the brackets. More specifically, the bivariate copula functions and their corresponding parameter values for the C-Vine structure within each time block are given in Table 1 of the supplementary materials. In the scenario of strong dependence, the Kendall's Taus of the bivariate copulas in T_1, T_2 and T_3 are set to be 0.7, 0.6 and 0.5, respectively; in that of moderate dependence, they are set to be 0.4, 0.3 and 0.2, respectively. The values of the dependence. We consider the following six scenarios:

- Scenario 1: The marginal parameters are set as $\eta_l = (\beta_{0l}, \beta_{1l}, \beta_{2l}, \sigma_l)^{\mathrm{T}} = (l, l+1, l+2, 2)^{\mathrm{T}}$ for $l = 1, 2, 3, 4$.
- Scenario 2: We restrict the marginal parameters across different time points to be the same. Specifically, we set $\eta_l = (\beta_{0l}, \beta_{1l}, \beta_{2l}, \sigma_l) = (2.5, 3.5, 4.5, 2)$ for $l = 1, 2, 3, 4$.
- Scenario 3: The dependence structures within each time block previously assumed to be the same are allowed to be different from block to block. More specifically, the bivariate copulas and the value of dependence parameters for the strong and the moderate dependence settings are given in Table 11 in the supplementary materials.
- Scenario 4: We consider the same settings as those of Scenarios 3, except that

$$\eta_l = (\beta_{0l}, \beta_{1l}, \beta_{2l}, \sigma_l) = (2.5, 3.5, 4.5, 2), \quad l = 1, 2, 3, 4.$$

- Scenario 5: The error terms ε_i are simulated from an $AR(1)$ structure instead of an R-Vine. We set $\rho = 0.5$ for $m = ab = 20$ time points. The marginal model is assumed to be

$$y_{ij} = 2.5 + 3.5x_{ij} - 50\sin\left(\frac{\pi j}{2}\right) + 50\cos\left(\frac{\pi j}{2}\right) + \varepsilon_{ij},$$

  where ε_{ij} are independently generated from $N(0, 1)$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. The sine and cosine functions are used to model the periodic trend.
- Scenario 6: We consider the same setting as that of Scenario 5, except that the marginal model does not contain the periodic sine and cosine functions but is of the form

$$y_{ij} = 2.5 + 3.5x_{ij} + 4.5j + \varepsilon_{ij},$$

  where ε_{ij} are independently generated from $N(0, 1)$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$.

Scenarios 3 and 4 are designed to evaluate the prediction performance when the dependence structures within each time block are not identical. Scenario 5 and 6 are designed when the true dependence structure is not a vine.

We fit the following models and compare their prediction performance.

- VINE1 : The proposed R-Vine copula model is fitted using the proposed composite likelihood method. For Scenarios 1-4, the (conditional) bivariate copula functions are assumed to follow the correct forms of the settings. For Scenarios 5 and 6, the (conditional) bivariate copula functions are all assumed to be the Gaussian copula. The parameters are estimated using simultaneous estimation presented in Section 3.1.

- VINE2 : The same as VINE1 except that the parameters are estimated using two-stage estimation procedure presented in Section 3.2.
- VINE3 : The (conditional) bivariate copulas are selected using the methods presented in Section 4 and the parameters are estimated under simultaneous estimation.
- VINE4 : The same as VINE3 except that the parameters are estimated under two-stage estimation.
- MRM: We assume that the marginal model for the lth time point is identical across time blocks. A marginal regression model of the form (10) is fitted. The dependence structure is completely ignored.
- LRM: A linear regression model is fitted, which takes both time block k and time point l as covariates and is of the form

$$y_{ikl} = \beta_0 + \beta_1 x_{ikl} + \beta_2 k + \beta_3 l + \varepsilon_{ikl},$$

  where ε_{ikl} are assumed to follow $N(0, \sigma^2)$, for $i = 1, \ldots, n; k = 1, 2, 3, 4, 5; l = 1, 2, 3, 4$
- AR: An autoregressive (AR) model in time series analysis is considered. The model form and the time lag are determined from the data.

We are interested in predicting the response value for a subject at a future time point. We partition the data by time points, use the time points from the first four blocks as the training set, denoted by $\{(y_{ikl}^{\mathrm{T}}, x_{ikl}^{\mathrm{T}})^{\mathrm{T}} : i = 1, \ldots, 500; k = 1, 2, 3, 4; l = 1, 2, 3, 4\}$, and reserve the time points in the fifth block as the test set, denoted by $\{(y_{ikl}^{\mathrm{T}}, x_{ikl}^{\mathrm{T}})^{\mathrm{T}} : i = 1, \ldots, 500; k = 5; l = 1, 2, 3, 4\}$. The training set is used to fit a model, which is utilized to predict y_{ikl} for a time point in time block $k = 5$, based on the covariate information and the first $l - 1$ time points in the 5th time block. We report the results in terms of Mean Absolute Error (MAE), the mean of the absolute difference between the predicted value and the true value over all time points in the test set across 200 simulations, and the associated prediction standard errors. We also provide the results in terms of "percentage outperformance" in Table 13 in supplementary materials, of which the definition is given in Section 3.4.3 in supplementary materials.

5.2 Prediction Results

We report simulation results for time extrapolations using all candidate models in terms of MAEs in Table 1. We also provide the boxplots of overall MAEs and MAEs by time points in Section 3.4.4 and Section 3.4.5 in the supplementary materials.

The four vine-based methods perform similarly across all the considered scenarios. They all provide smaller and less variant MAEs, suggesting superiority in prediction performance compared to other models. In Scenarios 1-4, it is not surprising that the vine-based models outperform the other ones, since the true models hold a vine structure. But the vine-based models still slightly outperform the AR model when the true model holds an AR(1) structure in Scenarios 5-6 and the dependence structures based on vine copulas are completely misspecifed. AR performs either comparably to MRM and LRM or a lot worse (e.g., in scenarios 1 and 3). The four vine-based models have smaller MAEs when the dependence is stronger while the MAEs are comparable in the strong and moderate settings when using MRM, LRM and AR models. The aforementioned prediction results are backed up by those measured by percentage outperformance, the percentage of one model outperforming the other, which is formally defined in Section 3.4.3 in the supplementary materials.

VINE1 and VINE3 yield smaller prediction standard errors than VINE2 and VINE4, because the simultaneous estimation tends to be more efficient than the two-stage estimation.

Table 1: MAEs of different models for time extrapolation under the proposed scenarios. S: strong dependence setting; M: moderate dependence setting.

Scenario	VINE1	VINE2	VINE3	VINE4	MRM	LRM	AR
1(S)	0.760 (1.076)	0.760 (1.082)	0.765 (1.076)	0.765 (1.083)	1.596 (1.954)	2.999 (2.823)	11.356 (7.681)
1(M)	1.145 (1.344)	1.145 (1.352)	1.146 (1.345)	1.146 (1.352)	1.598 (1.963)	3.002 (2.832)	11.360 (7.682)
2(S)	0.760 (1.076)	0.760 (1.083)	0.765 (1.076)	0.765 (1.083)	1.596 (1.953)	1.596 (1.953)	1.597 (1.951)
2(M)	1.145 (1.344)	1.145 (1.352)	1.146 (1.344)	1.146 (1.353)	1.598 (1.963)	1.597 (1.963)	1.598 (1.962)
3(S)	0.847 (0.663)	0.865 (0.675)	0.837 (0.664)	0.888 (0.675)	1.596 (1.942)	3.000 (2.818)	11.356 (7.665)
3(M)	1.219 (1.168)	1.222 (1.190)	1.230 (1.169)	1.232 (1.190)	1.599 (1.951)	3.002 (2.827)	11.359 (7.781)
4(S)	0.847 (0.663)	0.865 (0.675)	0.837 (0.664)	0.888 (0.675)	1.596 (1.942)	1.596 (1.942)	1.597 (2.976)
4(M)	1.219 (1.168)	1.222 (1.190)	1.230 (1.169)	1.232 (1.190)	1.599 (1.951)	1.598 (1.951)	1.599 (2.981)
5	0.830 (1.040)	0.830 (1.040)	0.830 (1.040)	0.830 (1.040)	0.922 (1.154)	0.922 (1.154)	0.920 (1.153)
6	0.830 (1.040)	0.831 (1.040)	0.831 (1.040)	0.831 (1.040)	0.923 (1.156)	0.923 (1.156)	0.922 (1.154)

However, factoring in the computation cost, the improvement of using the former method over the latter one seems marginal; in applications, it may not always be worthwhile to pursue the simultaneous estimation method due to its computation cost. Incorporating the observation history can greatly reduce the prediction standard errors. Moreover, prediction standard errors decrease as the strength of dependence increases.

From the boxplots of MAEs by time points in the supplementary materials, we find the MAEs for a later time point are always smaller and less variant when using the vine models, which is the benefit of taking into account the dependence structure within time blocks.

6 Data Analysis

6.1 Description of the Dataset

We consider the climate data available publicly on the website of Government of Canada. It is homogenized Canadian surface air temperature data (Vincent et al., 2012). The data is available at https://www.canada.ca/en/environment-climate-change/services/climate-change/science-research-data/climate-trends-variability/adjusted-homogenized-canadian-data.html. The dataset we use contains monthly mean of daily mean temperature in Celsius degree at 47 Ontarian observation stations from January 1978 to December 2018. Figure 2 is a run chart of the monthly temperature of the 47 stations from January 1978 to December 2018, which obviously exhibits a yearly periodic pattern and a mild overall increasing trend.

6.2 Statistical Models

In our analysis, the monthly `temperature` is used as the response variable, and the geographical information, `latitude`, `longitude` and `elevation`, and the time variables `year` are covariates. It is natural to select a year as a time block, yielding $a = 40$ time blocks (years) in total and $b = 12$ time points (months) in each block. We partition the 47 stations into a training group with 42 stations, and a test group with 5 stations, and we make a division in time by letting January 1978 to December 2008 be the training period and January 2009 to December 2018 as the testing period. The station information and the division of stations into training and test groups are given in Table 14 in supplementary materials. We use the data of the 42 stations from January 1978 to December 2008 to fit a model.

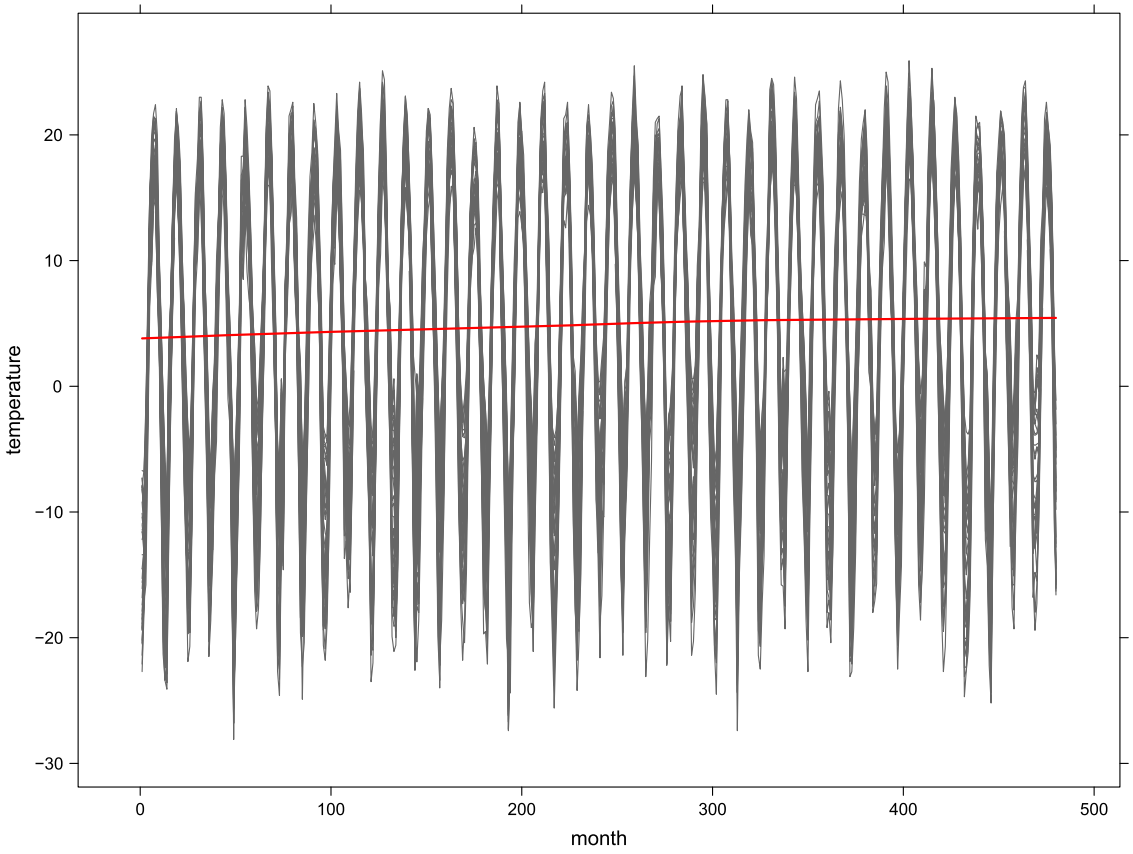

Figure 2: Monthly temperatures of all 47 stations from January 1978 to December 2018 (in grey) overlaid by that of the yearly average temperatures of the 47 stations (in red).

6.2.1 Marginal Model

The `temperature` highly depends on the geographical information, i.e., `latitude`, `longitude` and `elevation`, and tends to have an increasing trend with respect to `year` in some months. Preliminary marginal regression analysis (not shown here) suggests that the four covariates all have linear or quadratic relation with the responses, and the identity link function seems to be adequate, and the error terms of each month are appropriate to be modeled by a normal distribution with mean 0.

We assume that the marginal model for the lth month is of the following form: for $l = 1, 2, 10, 11, 12$,

$$\begin{aligned} Y_{ikl} =& \beta_{0l} + \beta_{1l} \cdot \texttt{latitude} + \beta_{2l} \cdot \texttt{longitude} + \beta_{3l} \cdot \texttt{elevation} \\ &+ \beta_{4l} \cdot \texttt{year} + \varepsilon_{ikl}; \end{aligned} \tag{11}$$

and for $l = 3, 4, 5, 6, 7, 8, 9$,

$$\begin{aligned} Y_{ikl} =& \beta_{0l} + \beta_{1l} \cdot \texttt{latitude} + \beta_{2l} \cdot \texttt{longitude} + \beta_{2l2} \cdot \texttt{longitude}^2 \\ &+ \beta_{3l} \cdot \texttt{elevation} + \beta_{4l} \cdot \texttt{year} + \varepsilon_{ikl}, \end{aligned} \tag{12}$$

Table 2: Summary of the selected bivariate copula functions for the C-Vine structure within each year. Cl=Clayton, Fr=Frank, Ga=Gaussian, Gu=Gumbel, In=Independent, Jo=Joe, T=Student t, T1=Tawn type 1, T2=Tawn Type 2, CG=Clayton-Gumbel mixed, JC=Joe-Clayton mixed, JF=Joe-Frank mixed. R means rotated with rotated degree in the bracket and S means survival copula.

	2	3	4	5	6	7	8	9	10	11	12	$\min(\hat{\tau})$	$\max(\hat{\tau})$
1	RT1(180)	T2	Ga	Cl	In	SCl	Cl	Fr	Ga	In	In	−0.151	0.186
2		JF	SCl	In	SCl	SCl	SCl	T1	Jo	T2	T2	0.000	0.215
3			T	Cl	In	Cl	RT1(90)	In	SJC	RT2(180)	T	−0.054	0.179
4				RT1(180)	T	Ga	In	In	RCl(90)	SJF	Ga	−0.089	0.165
5					In	SCl	RT2(180)	In	SCl	In	Jo	0.000	0.076
6						Ga	JC	Fr	Fr	RJo(90)	In	−0.048	0.371
7							SJC	SCl	RJo(90)	Gu	RGu(90)	−0.081	0.178
8								In	RT2(180)	In	T	−0.109	0.111
9									SGu	RT2(180)	In	0.000	0.180
10										RT2(180)	RCl(90)	−0.077	0.053
11											JF	0.208	0.208

where the ε_{ikl} are marginally distributed as $N(0, \sigma_l^2)$ for $l = 1, \ldots, 12$, $i = 1 \ldots, n$ is the index of observation stations and $k = 1, \ldots, 40$ is the index of time block (year).

6.2.2 Dependence Model

We ignore the dependence structure between years, model the dependence between months within each year through a C-Vine. We first select the copula functions for the C-Vine structure within each year by using the copula selection method we proposed in Section 4, which is implemented using the `VineCopula` package in `R` based on a dataset of 1260 years with each of the 42 stations in the training group contributing 30 years (the training period). All copula functions available in the `VineCopula` package are included in the candidate set for selection; the available copula functions, are described by Schepsmeier et al. (2020). Table 2 summarizes the selected bivariate copula functions, where the lth row corresponds to the lth level of tree in the C-Vine structure and variable l is the dominating variable in this level of tree. The lth tree and the l'th month in Table 2 gives the selected (conditional) bivariate copula functions between variables ε_{ikl} and $\varepsilon_{ikl'}$. The minimum ($\min(\hat{\tau})$) and maximum ($\max(\hat{\tau})$) values of the corresponding Kendall's Tau for each level of the tree are also provided in the last two columns in Table 2. We can see that the dependence between time points are moderate, especially in higher level of trees.

6.2.3 Model Fitting, Model Comparison and Prediction

Based on the selected copula functions, we perform composite likelihood estimation. The total number of parameters, which is around 150, is too large for common optimization algorithm to optimize simultaneously and obtain simultaneous estimators. The four vine-based methods provide comparable prediction results by simulations, thus we implement composite likelihood estimation under two-stage estimation procedures (VINE4) here. The estimation for marginal parameters are summarized in Table 3 and those for dependence parameters are summarized in Tables 15 and 16 in supplementary materials.

Table 3: The estimates of marginal parameters for each month l under simultaneous estimation and two-stage estimation of composite likelihood method (standard error in the parentheses).

	Two-Stage Estimation						
l	β_{0l}	β_{1l}	β_{2l}	β_{2l2}	β_{3l}	β_{4l}	σ_l
1	−135.740(62.240)	−1.978(0.041)	−0.210(0.037)	–	−0.009(0.002)	0.101(0.030)	3.204(0.064)
2	−34.827(27.651)	−1.739(0.042)	−0.256(0.039)	–	−0.008(0.003)	0.043(0.014)	3.164(0.067)
3	22.785(89.626)	−1.429(0.119)	−44.929(19.069)	16.994(5.543)	−0.005(0.003)	0.021(0.044)	2.207(0.116)
4	2.431(26.704)	−1.012(0.099)	−28.691(12.179)	25.054(4.627)	−0.003(0.002)	0.025(0.133)	1.944(0.090)
5	44.601(6.172)	−0.708(0.100)	−14.893(26.378)	25.789(6.031)	−0.002(0.007)	0.007(0.102)	1.939(0.034)
6	−100.571(21.824)	0.681(0.046)	−17.278(12.666)	22.259(3.084)	−0.002(0.003)	0.075(0.010)	1.578(0.034)
7	30.540(45.497)	−0.626(0.036)	−21.546(5.831)	19.626(2.988)	−0.004(<0.001)	0.010(0.022)	1.417(0.034)
8	1.261(23.072)	−0.685(0.032)	−27.126(5.479)	15.480(3.112)	−0.006(0.001)	0.025(0.011)	1.482(0.032)
9	−90.891(13.597)	−0.877(0.073)	−27.676(10.608)	8.679(3.121)	−0.007(0.001)	0.074(0.006)	1.335(0.063)
10	−54.479(10.110)	−0.918(0.020)	−0.117(0.012)	–	−0.008(<0.001)	0.048(0.005)	1.518(0.029)
11	−28.415(11.045)	−1.275(0.028)	−0.061(0.018)	–	−0.009(<0.001)	0.042(0.006)	2.085(0.047)
12	−68.781(19.581)	−1.764(0.045)	−0.112(0.028)	–	−0.008(<0.001)	0.068(0.010)	3.324(0.067)

In the estimation results, β_{1l} is negative for all 12 months, which suggests high-latitude areas tend to have lower temperature year around and this trend is more obvious in winter months (i.e., $|\beta_{1l}|$ is larger in months 1, 2, 3, 11 and 12). For winter months, i.e., months 1-2 and 10-12, the mean temperature has a linear negative relation with the longitude. For months 3-9 in spring and summer, the mean temperature has a quadratic relation with the longitude. β_{3l} is negative but close to zero, suggesting that as the elevation increases, the mean temperature will slightly decrease. β_{1l}, the annual temperature increase of the lth month in Celsius degree, is positive in all 12 months, which suggests a mildly increasing trend of temperature change over years. The findings perfectly align with our expectations.

We are interested in both subject extrapolation (predicting temperature for a new station based on geographical information and time) and time extrapolation (predicting temperature for a future time). In practice, the former allows us to predict temperatures for locations without a station and the latter allows us to forecasting future temperatures. For subject extrapolation, we predict temperatures for the 5 stations in the test group from January 1978 to December 2008, of which the results are provided in Section 4.3 in Supplementary Materials. For time extrapolation, we predict for 37 stations in the training group from January 2009 to December 2018. There are five stations closed after 2008 and data from January 2009 to December 2018 are not available. We are interested in short-term, mid-term and long-term prediction. For short-term prediction, the prediction for the lth month is made based on information from previous $l-1$ months in the same year and the prediction of the first months is using the marginal distribution; in other words, this is prediction for the next month. For mid-term prediction, the prediction for the lth month is made based on the temperature in the first season (months 1-3) in the same year, for $l = 4, \ldots, 12$; in other words, this is the prediction made for the rest of the year. For long-term prediction, we are predicting the change of the temperature in a decade.

We compare the prediction performance of VINE4 with MRM, LRM and AR using the evaluation metrics MAE and Percentage Outperformance as we did in the simulation studies:

- MRM: The monthly marginal regression model (MRM) (11) and (12) without considering the dependence structure.
- LRM: A linear regression model (LRM) includes `month` x_5 as a covariate to account for the

variation across months. The LRM model is selected by the AIC criterion and fitted to be

$$\begin{aligned} Y_{ikl} =& \beta_0 + \beta_1 \cdot \texttt{latitude} + \beta_2 \cdot \texttt{longitude} + \beta_3 \cdot \texttt{elevation} \\ &+ \beta_4 \cdot \texttt{year} + \sum_{j=1}^{2} \beta_{5j} \cdot \texttt{month}^j + \varepsilon_{ikl}, \end{aligned}$$

where $\varepsilon_{ikl} \sim N(0, \sigma^2)$.

- AR: An autoregressive (AR) time series model, which is selected and fitted to be

$$\begin{aligned} Y_{it} =& \beta_0 + \beta_1 \cdot \texttt{latitude} + \beta_2 \cdot \texttt{longitude} + \beta_3 \cdot \texttt{elevation} \\ &+ \beta_4 \cdot \texttt{year} + \beta_5 sin\left(\frac{\pi t}{6}\right) + \beta_6 cos\left(\frac{\pi t}{6}\right) + \varepsilon_{it}, \end{aligned}$$

  where $\varepsilon_{it} \sim AR(2)$ for $t = 1, \ldots, 360$.
- SARIMA: A seasonal autoregressive integrated moving average (SARIMA) model, which is commonly used for seasonal time series data prediction:

$$Y_{it} = \beta_0 + \beta_1 \cdot \texttt{latitude} + \beta_2 \cdot \texttt{longitude} + \beta_3 \cdot \texttt{elevation} + \varepsilon_{it},$$

  where $\varepsilon_{it} \sim \text{SARIMA}(3, 1, 3)(1, 0, 1, 12)$ for $t = 1, \ldots, 360$.

6.2.4 Prediction Results

We evaluate the prediction performance of our proposed method for short-term, mid-term and long-term prediction. Figure 3 contains two subfigures, which corresponds to the prediction performance for short-term (on the left) and mid-term (on the right) prediction, respectively. The mid-term prediction was made for months 4-12, but the short-term prediction was made for all 12 months, little previous information is available for months 1-3 and it tends to have large prediction errors in the first three months. Therefore, the short-term prediction has larger median MAEs across all methods.

From the boxplots of both short-term and mid-term predictions, the VINE4 has a smaller or comparable median MAEs compared to the other methods, and the MAEs of VINE4 are the least variant. Since the dependence between months within each year is moderate, the advantage of the VINE4 method versus the marginal model (MRM) is limited, which agrees with our findings in Section 5.2. We provide the station-by-station MAEs for short-term time extrapolation in Table 19 and that for mid-term time extrapolation in Table 20 in the supplementary materials.

For long-term prediction, we take an average of β_{4l} for $l = 1, \ldots, 12$ in Table 3 and obtain the average of the temperature annual increase of the 12 months, which is 0.045 in Celsius degree. Based on the training group of 42 stations and the training period of 3 decides from January 1978 to December 2008, we predict the temperature increase in the next decade (January 2009 to December 2018) in the area is 0.45 Celsius degree (with 95% confidence interval [-0.019, 0.109]). The actual temperature increase from January 2009 to December 2018 is 0.65 Celsius degree.

7 Discussion

In this paper, we propose a R-Vine based regression model for analyzing longitudinal data with long time span. We introduce composite likelihood methods which outperforms the likelihood-based methods in terms of robustness and computational efficiency. We conduct extensive simulation studies to evaluate the performance of the proposed methods. The numerical studies

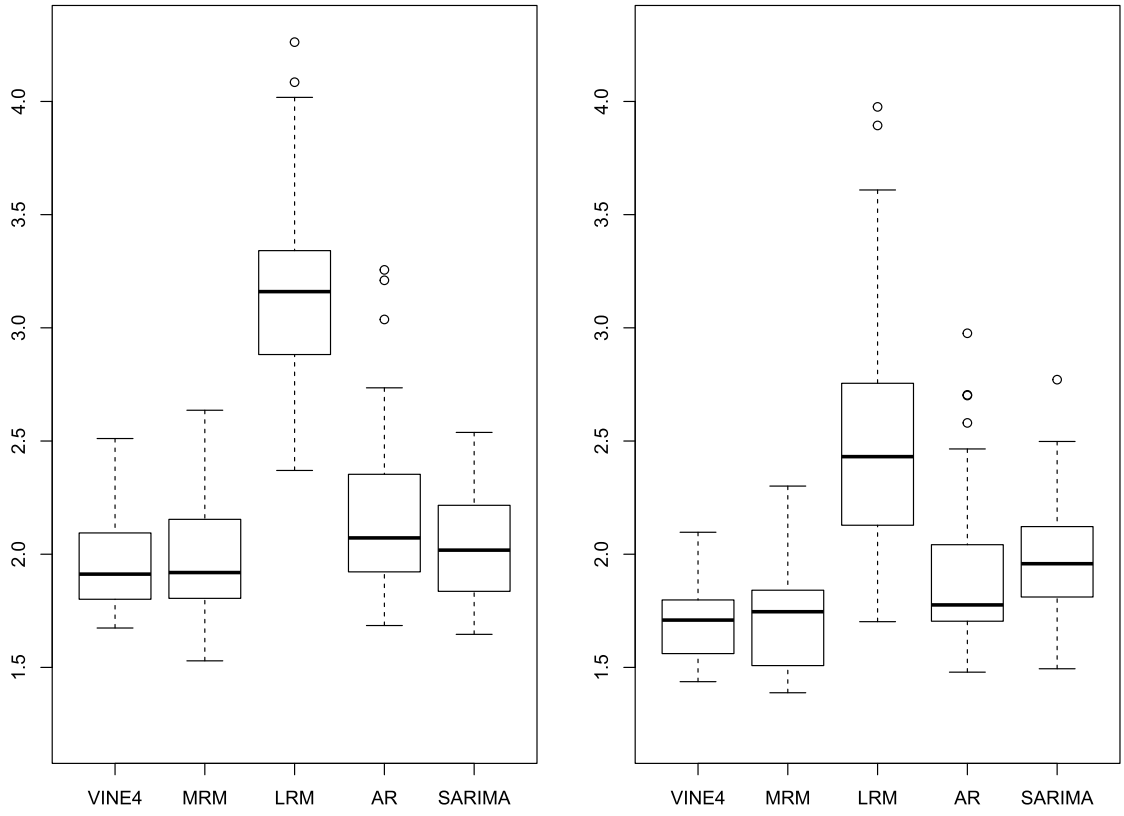

Figure 3: Boxplot of MAEs for the short-term (on the left) and mid-term (on the right) time extrapolation.

suggest that the (conditional) bivariate copulas can still be accurately selected and the parameters of interest can be consistently estimated with moderate efficiency loss when simultaneous procedure is used. Moreover, the model provides more precise prediction results than the conventional models in both the simulation studies and the real data analysis. Time extrapolation is what we usually care about in prediction problems, and both time and subject extrapolations are valuable for imputing missing response values.

References

Aas K, Czado C, Frigessi A, Bakken H (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2): 182–198.

Bedford T, Cooke RM (2002). Vines–a new graphical model for dependent random variables. *Annals of Statististics*, 30(4): 1031–1068.

Brechmann EC, Czado C, Aas K (2012). Truncated regular vines in high dimensions with application to financial data. *Canadian Journal of Statistics*, 40(1): 68–85.

Diggle P, Heagerty P, Heagerty PJ, Liang KY, Zeger S (2002). *Analysis of Longitudinal Data*. Oxford University Press.

Dissmann J, Brechmann EC, Czado C, Kurowicka D (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics and Data Analysis*, 59(1): 52–69.

Domma F, Giordano S, Perri P (2009). Statistical modelling of temporal dependence in financial data via a copula function. *Communications in Statistics – Simulation and Computation*, 38: 703–728.

Fang M, Tan KS, Wirjanto TS (2019). Sustainable portfolio management under climate change. *Journal of Sustainable Finance & Investment*, 9(1): 45–67.

Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G (2009). *Longitudinal Data Analysis*. CRC Press.

Frees E, Wang P (2006). Copula credibility for aggregate loss models. *Insurance: Mathematics and Economics*, 38: 360–373.

Hedeker D, Gibbons RD (2006). *Longitudinal Data Analysis*. John Wiley & Sons.

Hugo G (2013). *Migration and Climate Change*. Edward Elgar Publishing Limited.

Joe H (1997). *Multivariate Models and Multivariate Dependence Concepts*. Chapman and Hall/CRC.

Killiches M, Czado C (2018). AD-vine copula-based model for repeated measurements extending linear mixed models with homogeneous correlation structure. *Biometrics*, 74(3): 997–1005.

Lambert P, Vandenhende F (2002). A copula-based model for multivariate non-normal longitudinal data: analysis of a dose titration safety study on a new antidepressant. *Statistics in Medicine*, 21: 3197–3217.

Lim B, Spanger-Siegfried E, Burton I, Malone E, Huq S (2004). *Adaptation Policy Frameworks for Climate Change: Developing Strategies, Policies and Measures*. Cambridge University Press.

Lindsay B (1988). Composite likelihood methods. *Contemporary Mathematics*, 80: 220–239.

Lindsay BG, Yi GY, Sun J (2011). Issues and strategies in the selection of composite likelihoods. *Statistica Sinica*, 21: 71–105.

Madsen L, Fang Y (2011). Joint regression analysis for discrete longitudinal data. *Biometrics*, 67: 1171–1176.

Nelsen RB (2007). *An Introduction to Copulas*. Springer Science & Business Media.

OBrien K (2010). Responding to climate change: The need for an integral approach. In: *Integral Theory in Action: Applied, Theoretical, and Constructive Perspectives on the AQAL Model* (S Esbjörn-Hargens, ed.), 65–78. SUNY Press.

Parmesan C, Yohe G (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421(6918): 37–42.

Ruscone MN, Osmetti SA (2016). Modelling the dependence in multivariate longitudinal data by pair copula decomposition. In: *Soft Methods for Data Science* (MB FerraroPaolo Giordani, B Vantaggi, M Gagolewski, M Ángeles Gil, P Grzegorzewski, O Hryniewicz, eds.), 373–380. Springer.

Schepsmeier U, Stoeber J, Brechmann EC, Graeler B, Nagler T, Erhardt T, et al. (2020). *VineCopula: Statistical Inference of Vine Copulas*. R package version 2.4.1.

Shen C, Weissfeld L (2006). A copula model for repeated measurements with non-ignorable non-monotone missing outcome. *Statistics in Medicine*, 25: 2427–2440.

Shi P, Yang L (2018). Pair copula constructions for insurance experience rating. *Journal of the*

American Statistical Association, 113(521): 122–133.

Smith M, Min A, Almeida C, Czado C (2010). Modeling longitudinal data using a pair-copula decomposition of serial dependence. *Journal of the American Statistical Association*, 105: 1467–1479.

Smith MS (2015). Copula modelling of dependence in multivariate time series. *International Journal of Forecasting*, 31: 815–833.

Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, et al. (2013). Climate change 2013: The physical science basis.

Varin C (2008). On composite marginal likelihoods. *AStA Advances in Statistical Analysis*, 92: 1–28.

Varin C, Reid N, Firth D (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21: 5–42.

Verbeke G, Fieuws S, Molenberghs G, Davidian M (2014). The analysis of multivariate longitudinal data: A review. *Statistical Methods in Medical Research*, 23(1): 42–59.

Verbeke G, Molenberghs G (2009). *Linear Mixed Models for Longitudinal Data*. Springer Science & Business Media.

Vincent LA, Wang XL, Milewska EJ, Wan H, Yang F, Swail V (2012). A second generation of homogenized canadian monthly surface air temperature for climate trend analysis. *Journal of Geophysical Research: Atmospheres*, 117: D18110.

Walls TA, Schafer JL (Eds.) (2006). *Models for Intensive Longitudinal Data*. Oxford University Press.

Wheeler T, von Braun J (2013). Climate change impacts on global food security. *Science*, 341(6145): 508–513.

Yi GY (2017). Composite likelihood/pseudolikelihood. In: *Wiley StatsRef: Statistics Reference Online*. Wiley Online Library. https://doi.org/10.1002/9781118445112.stat07855.