UC3M Working Papers
Statistics and Econometrics
16-12
ISSN 2387-0303
Octubre 2016

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-48

VINE COPULA MODELS FOR PREDICTING WATER FLOW DISCHARGE AT KING GEORGE ISLAND, ANTARCTICA

Gómez, M.[a], Ausín, M.C.[a], Domínguez, M. C.[b]

Abstract

In order to understand the future behavior of the glaciers, their mass balance should be studied. The loss of water produced by melting, known as glacier discharge, is one of the components of this mass balance. In this paper, a vine copula structure is proposed to model the multivariate and nonlinear dependence among the glacier discharge and other related meteorological variables such as temperature, humidity, solar radiation and precipitation. The multivariate distribution of these variables is divided in four cases according to the presence or not of positive discharge and/or positive precipitation. Then, each different case is modelled with a vine copula. The conditional probability of zero discharge for given meteorological conditions is obtained from the proposed joint distribution. Moreover, the structure of the vine copula allows us to derive the conditional distribution for the glacier discharge for the given meteorological conditions. Three different prediction methods for the future values of the discharge are used and compared.

The proposed methodology is applied to a large database collected since 2002 by the GLACKMA association from a measurement station located in the King George Island in the Antarctica. Seasonal effects are included by using different parameters for each season.

We have found that the proposed vine copula model outperforms a previous work where we only used the temperature to predict the glacier discharge using a time-varying bivariate copula.

Keywords:* *Glacier discharge, vine copula, prediction, meteorological, finite mixtures

[a] Department of Statistics, Universidad Carlos III de Madrid.
[b] Department of Applied Mathematics, Universidad de Salamanca.

Vine copula models for predicting water flow discharge in Antarctic glaciers

Mario Gómez[1], M. Concepción Ausín[2], M. Carmen Domínguez[3]

[1−2]Department of Statistics, Universidad Carlos III de Madrid,

[3]Department of Applied Mathematics University of Salamanca, Spain.

Abstract

In order to understand the future behavior of the glaciers, their mass balance should be studied. The loss of water produced by melting, known as glacier discharge, is one of the components of this mass balance. In this paper, a vine copula structure is proposed to model the multivariate and nonlinear dependence among the glacier discharge and other related meteorological variables such as temperature, humidity, solar radiation and precipitation. The multivariate distribution of these variables is divided in four cases according to the presence or not of positive discharge and/or positive precipitation. Then, each different case is modelled with a vine copula. The conditional probability of zero discharge for given meteorological conditions is obtained from the proposed joint distribution. Moreover, the structure of the vine copula allows us to derive the conditional distribution for the glacier discharge for the given meteorological conditions. Three different prediction methods for the future values of the discharge are used and compared. The proposed methodology is applied to a large database collected since 2002 by the GLACKMA association from a measurement station located in the King George Island in the Antarctica. Seasonal effects are included by using different parameters for each season. We have found that the proposed vine copula model outperforms a previous work where we only used the temperature to predict the glacier discharge using a time-varying bivariate copula.

Keywords: Glacier discharge, vine copula, prediction, meteorological, finite mixtures

1 Introduction

The study of the mass balance in glaciers is crucial for the correct quantification of water resources (Hamlet and Lettenmaier, 1999; Marsh 1999). Mass balance is the difference between accumulation (mainly in form of fallen snow) and ablation (produced by sublimation, calving and melting). Glacier discharge is defined as the rate of flow of meltwater through a vertical section perpendicular to the direction of the flow (Cogley et al. 2011). It is produced by surface runoff or by flowing inside the glacier and exit through the front or the base.

The Antarctic Peninsula is considered as one of the Recent Rapid Regional Climate Warming, which refer to those areas where the regional changes have been deeper than the global warming, as noted by the Intergovernmental Panel on Climate Change (IPCC) (Turner et al. 2005; Vaughan et al. 2003). Periods of melting of the glaciers in this area have been increasing year by year (Domínguez and Eraso 2007). As consequence, there has been a retreat of the glaciers and changes in their heights (Rückamp et al 2011). Also, trends in surface melting have been found (Barrand et al 2013).

The study of the relationship between glacier behavior and climate is a fundamental issue in glaciology. These relationships can be analyzed with the energy balance equations which evaluate the most important energy fluxes between the atmosphere and the glacier surface. These equations are computed from physically based calculations (see e.g. Braun 2001; Sicart et al. 2008) and involve complex equations and measurements. The melt rate is the sum of all individual fluxes. Alternatively, temperature index models use only the air temperature to empirically model this relationship. A complete review of these methods can be found in Hock (2003). A leading form of temperature-index models is the so called degree-day model which is based on an assumed relationship between ablation and air temperature. There are some studies that incorporate more variables to this model, such as the direct solar radiation (Hock 1999) or the albedo and the shortwave radiation (Pellicciotti et al. 2005). The main problem in this type of models is that they implicitly assume a linear relationship between temperature and discharge.

In this paper we propose the use of multivariate copulas to model the non-linear relationship among the discharge, temperature and several other meteorological variables. Copulas are statistical instruments that allow us modelling the relationship among the variables independently of the marginal distributions choice (Genest and Favre 2007). See Nelsen (2006) for an extensive review about copulas. In climate and hydrology, many research papers model the relationship between a pair of variables using bidimensional copulas (see e.g. De Michele and Salvadori 2003; Favre et al. 2004; De Michele et. al 2005; Scholzel and Friederichs 2008; Cong and Brady 2012; Hao and Singh 2012; Carnicero et al. 2013; Sahardi et al. 2016; Gómez et al. 2017). However, the number of works studying the relationship among more than two variables is much smaller. Standard multivariate copulas are available, such as the multivariate Gaussian or the t-Student copulas. However, they have shown to be rather inflexible as, for example, they assume that the dependence is symmetric in both tails, which is not realistic in this context. Alternatively, it is possible to model multivariate distributions using vine copulas. A vine copula is a flexible structure that decomposes the multivariate copula in a set of bivariate copulas. Vine copulas have been successfully used in a few number of papers in hydrology. For example, Gyasi-Agyei and Melching (2012) model the internal dependence structure between net storm event depth, maximum wet periods depth, and the total wet periods duration. Gyasi-Agyei (2013) models the dependence between total depth, total duration of wet periods, and the maximum proportional depth of a wet period in a rainfall disaggregation model. Xiong et al. (2015) study the dependence between annual maxima daily discharge, annual maxima 3-day flood volume and annual maxima 15-day flood volume to understand the change-point detection of multivariate hydrological series. Note that these papers deal with three hydrological variables, while in our work we study the relation between five variables which, in addition, may take discrete values.

This work has two main goals. First, we wish to predict the conditional probability of having no glacier discharge given the observations of temperature, humidity, radiation and precipitation. Also, we want to predict the future values of the discharge, with the conditional distribution of the discharge obtained through the vine copula. This paper extends our previous bivariate copula model (Gómez et al. 2017), based on a time-varying relationship between discharge and temperature, by the inclusion of three new meteorological variables which, as the case of the precipitation, may take zero values. We consider vine copulas to define the structure of our new model. We also propose a new way of dealing with zero values in the glacier discharge since these were considered as missing observations in Gómez et al. (2017), while it is assumed now that the glacier discharge may be equal to zero with positive probability. Another difference is the way seasonality has been taken into account. Due to the increase in the number of variables, for simplicity, we have divided each hydrologic year into four periods with different behaviour in the discharge regime instead of assuming a time-varying seasonality.

The remainder of the paper is organized as follows. First, the study area and the considered database are described in Section 2. This is followed by the proposal of a vine copula model and an estimation method in Section 3. The proposed methodology is applied, in Section 4, to the GLACKMA database. Finally, Section 5 concludes with some discussion and extensions.

2 Study area and data

King George is the largest of the South Shetland Islands, which is an archipelago placed near of the coast of the Antarctic Peninsula in the Southern Ocean. See Braun (2001) for a complete description of the island. Fig. 1 shows the study area, located in the south west side of the King George island, where GLACKMA has placed one of their eight Pilot Experimental Catchment Areas, named as CPE-KG-62°S, so as to study the discharge of the Collins Glacier, see www.glackma.es. Glacier discharge, measured in $m^3/sec \cdot Km^2$, is accurately estimated as an exponential function of the level of a river which receives the melted water from the catchment area. Also, we have selected a collection of meteorological variables such as the air temperature ($^\circ C$), the percentage of humidity (%), the solar radiation ($Watt/m^2$) and the precipitation (mm). These meteorological data have been provided by the Bellinghausen Russian base (via GLACKMA), sited near of the catchment area. See Domínguez and Eraso (2007) and Gómez et al. (2017) for a description of the catchment area and further details about the variables.

The available data are from 10/01/2002 to 09/30/2012, composed by daily measurements of the mean daily temperature, humidity, radiation and discharge and the daily cumulative precipitation. A preliminary study of these data shows that discharge and precipitation have a large number of zero-values. In particular, the value of the discharge was zero in 62% of the observed days and the value of the precipitation was zero

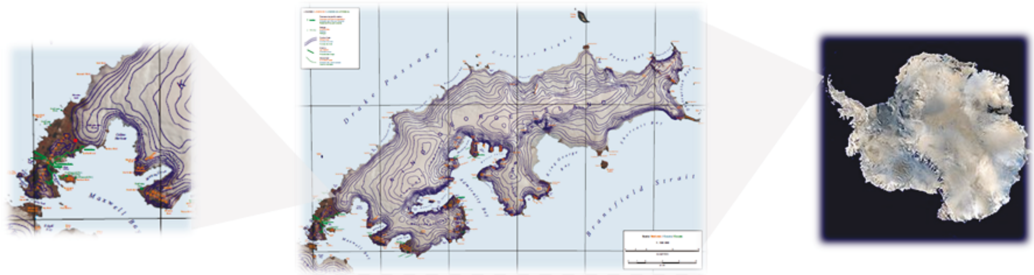

Figure 1: Right panel shows the location of King George Island (Antarctica). Central panel shows the island, mostly covered by Collins glacier. Left panel zooms in the location of the CPE-KG-62°S station, indicated with a red point. (Braun et al. 2002).

in 31% of the observed days. This fact has a definite impact in the design of the vine copula model. Fig. 2 shows the scatter plot of each pair of variables and the histogram of each individual variable for those days when the discharge was larger than zero. Apparently, there are strong relationships between the variables although these relations are not linear. Then, we suggest the use of copulas to model these non-linear relationships. The lower panel shows the value of the tau-rank correlations between each pair of variables. The values of these sample rank correlation, whose size is proportional to its absolute value, will help us to decide the order of the variables in the vine structure.

3 Methodology

In this section, we introduce a method to predict future values of the glacier discharge given the observed values of the meteorological variables. First, we propose a copula model to describe the multivariate joint distribution of the five variables where two of them have a large number of zero values. Then, we obtain the conditional probability of having no discharge. Finally, we derive the conditional distribution of the discharge given the other meteorological variables.

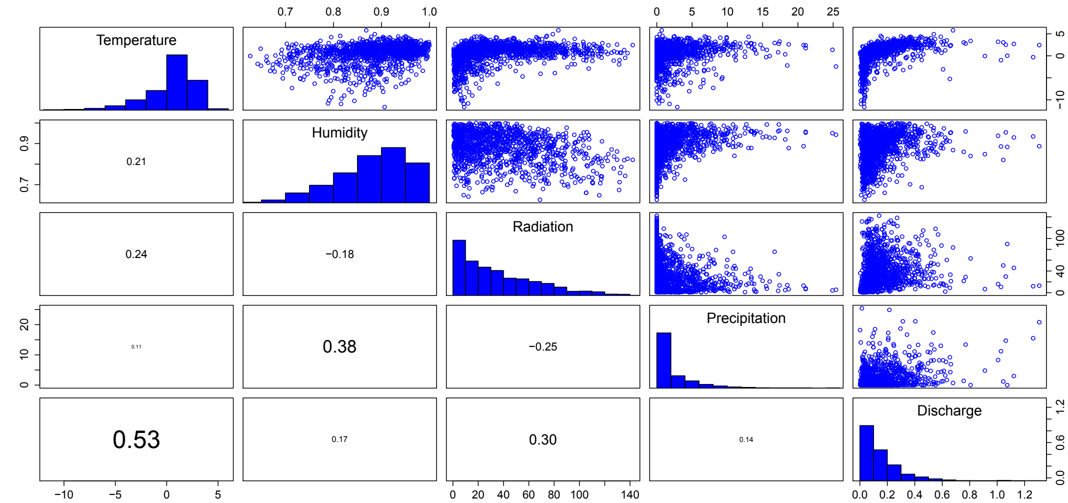

Figure 2: Scatter plot, histograms and Kendall's τ of the five variables when the values of the discharge are greater than zero. Size of the values of the τ is proportional to its absolute value.

3.1 Multivariate copula model

Let T, H, R, P and D be random variables, where T is the temperature, H the humidity, R the radiation, P the precipitation and D the discharge. As commented in the previous section, in practice, it is usually observed that both, the precipitation and the discharge, show a large number of zero values. This fact has a quite important impact in the construction of our proposed model. Erhardt and Czado (2012) and Brechmann et al. (2014) propose a model for a multivariate distribution in which the mixed variables are decomposed in zero inflated and continuous positive components. Following this idea, we define the joint distribution as a mixture of four different joint distributions, depending on the joint probability of presence of zero or positive values of the discharge and the precipitation. Thus, the joint density function of the multivariate variable (T, H, R, P, D) is decomposed as,

$$f(t,h,r,p,d) = \begin{cases} f(t,h,r \mid D=0, P=0), & \text{with } \Pr(D=0,P=0) \quad (1a) \\ f(t,h,r,d \mid D>0, P=0), & \text{with } \Pr(D>0,P=0) \quad (1b) \\ f(t,h,r,p \mid D=0, P>0), & \text{with } \Pr(D=0,P>0) \quad (1c) \\ f(t,h,r,p,d \mid D>0, P>0), & \text{with } \Pr(D>0,P>0) \quad (1d) \end{cases}$$

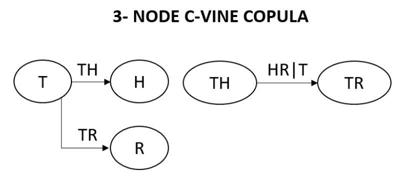

Figure 3: Structure of a c-vine copula with 3 nodes.

where the joint probability of having or not discharge and precipitation is obtained empirically. Then, we define each of these four joint density functions in terms of copulas using the theorem of Sklar (1959). For example, (1a) can be expressed as,

$$f(t,h,r \mid D=0,P=0) = c_{THR}\left(u_t^{00},u_h^{00},u_r^{00}\right) \cdot f_T^{00}(t) \cdot f_H^{00}(h) \cdot f_R^{00}(r), \tag{2}$$

where c_{THR} is the multivariate copula density describing the dependence structure in the variable (T,H,R) and

$$u_t^{00} = F_T\left(t \mid D=0,P=0\right), \quad f_T^{00}(t) = f_T\left(t \mid D=0,P=0\right), \tag{3}$$

and analogously for u_h^{00}, u_r^{00}, f_h^{00} and f_r^{00}, where the superscripts denote that the variables are conditioned on zero discharge and zero precipitation.

Our proposal is to use a vine copula for the multivariate copula in (2). A vine copula (Aas et al. 2009) is a decomposition of a multivariate copula as a product of bivariate copulas. Different vine decompositions can be obtained for the same multivariate copula. Bedford and Cooke (2001) introduced a graphical structure called regular vine structure to organize the different pairs of copulas. More specifically, a vine copula for n variables is a structure composed of $n-1$ trees, where the edges of one tree are the nodes of the next tree. Different bivariate copulas can be selected for each edge, introducing more flexibility. In particular, we consider a family of vine copulas, known as canonical vines (c-vine), where each tree has a node that is connected to all other. For example, the expression (2) can be decomposed, using the c-vine structure shown in Fig. 3, as,

$$c_{THR}(u_t^{00},u_h^{00},u_r^{00}) = c_{HR|T}(u_{h|t}^{00},u_{r|t}^{00}) \cdot c_{TH}(u_t^{00},u_h^{00}) \cdot c_{TR}(u_t^{00},u_r^{00}),$$

where $c_{HR|T}$, c_{TH} and c_{TR} are the density functions of the bivariate copulas in each edge and

$$\begin{aligned} u^{00}_{h|t} &= F\left(u^{00}_h \mid u^{00}_t\right) = \frac{\partial C_{TH}(u^{00}_t, u^{00}_h)}{\partial u^{00}_t} = F_H\left(h \mid T=t, P=0, D=0\right), \\ u^{00}_{r|t} &= F\left(u^{00}_r \mid u^{00}_t\right) = \frac{\partial C_{TR}(u^{00}_t, u^{00}_r)}{\partial u^{00}_t} = F_R\left(r \mid T=t, P=0, D=0\right), \end{aligned}$$

which are the conditional distribution functions of the uniform variable introduced in (3). Similarly, the expression (1b) can be expressed in terms of copulas as,

$$f\left(t,h,r,d \mid D>0, P=0\right) = c_{THRD}\left(u^{10}_t, u^{10}_h, u^{10}_r, u^{10}_d\right) \cdot f^{10}_T(t) \cdot f^{10}_H(h) \cdot f^{10}_R(r) \cdot f^{10}_D(d), \tag{4}$$

where c_{THRD} is the multivariate copula density describing the dependence of the variable (T, H, R, D) and

$$u^{10}_t = F_T\left(t \mid D>0, P=0\right), \quad f^{10}_T(t) = f_T\left(t \mid D>0, P=0\right),$$

and analogously for u^{10}_h, u^{10}_r, u^{10}_d and f^{10}_H, f^{10}_R, f^{10}_D, where the superscripts denotes that the variables are conditioned on positive discharge and zero precipitation. The multivariate copula in (4) can be decomposed as a product of bivariate copulas, using the vine structure shown in Fig. 4, as,

$$\begin{aligned} c_{THRD}\left(u^{10}_t, u^{10}_h, u^{10}_r, u^{10}_d\right) =& c^{10}_{RD|TH}\left(u^{10}_{hr|t}, u^{10}_{hd|t}\right) \cdot c^{10}_{HR|T}\left(u^{10}_{h|t}, u^{10}_{r|t}\right) \cdot c^{10}_{HD|T}\left(u^{10}_{h|t}, u^{10}_{d|t}\right) \\ & \cdot c^{10}_{TH}\left(u^{10}_t, u^{10}_h\right) \cdot c^{10}_{TR}\left(u^{10}_t, u^{10}_r\right) \cdot c^{10}_{TD}\left(u^{10}_t, u^{10}_d\right), \end{aligned} \tag{5}$$

where $c_{RD|TH}$, $c_{HR|T}$, $c_{HD|T}$, c_{TH}, c_{TR} and c_{TD} are the density functions of the bivariate copulas in each edge, and

$$u^{10}_{hr|t} = F\left(u^{10}_{r|t} \mid u^{10}_{h|t}\right) = \frac{\partial C_{HR|T}(u^{10}_{h|t}, u^{10}_{r|t})}{\partial u^{10}_{h|t}} = F_R\left(r \mid T=t, H=h, D>0, P=0\right),$$

and analogously for $u^{10}_{h|t}$, $u^{10}_{r|t}$ and $u^{10}_{d|t}$.

Similar expressions can be obtained for (1c) and (1d) which are shown in Appendix B.

3.2 Marginal distributions

Now, we define a marginal distribution model for each one of the five meteorological variables T, H, R, P and D. We decompose each variable in four cases according to the presence or not of precipitation and

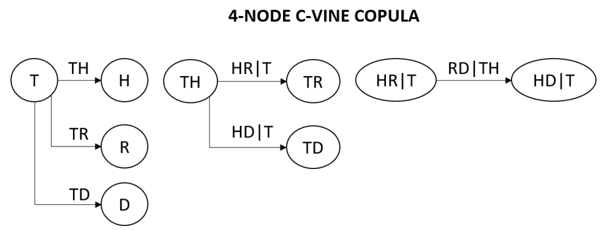

Figure 4: Structure of a c-vine copula with 4 nodes.

discharge, as in (1). For example, the density function of the temperature can be expressed as:

$$f_T(t) = \begin{cases} f_T^{00}(t) = f_T(t \mid D = 0, P = 0), \text{ with } \Pr(D = 0, P = 0) \\ f_T^{10}(t) = f_T(t \mid D > 0, P = 0), \text{ with } \Pr(D > 0, P = 0) \\ f_T^{01}(t) = f_T(t \mid D = 0, P > 0), \text{ with } \Pr(D = 0, P > 0) \\ f_T^{11}(t) = f_T(t \mid D > 0, P > 0), \text{ with } \Pr(D > 0, P > 0). \end{cases}$$

Then, for each of these four cases, we assume a parametric model based on finite mixture models. In particular, for the temperature, we consider finite mixture of Gaussian distributions. For example, the first density function can be written as,

$$f_T^{00}(t) = \sum_{i=1}^{K} \omega_i^{(T)} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(t-\mu_i)^2}{2\sigma_i^2}\right),$$

where $\omega_i^{(T)}$ are the weights of each component ($\sum_{i=1}^{K} \omega_i^{(T)} = 1$). Similar Gaussian mixtures for f_T^{10}, f_T^{01} and f_T^{11}. Then, we may obtain u_t^{00}, u_t^{10}, u_t^{01} and u_t^{11} using the cumulative distribution function of a Gaussian mixture. Therefore, note that we have a set of parameters to estimate each of the four Gaussian mixtures.

The same procedure is followed for the humidity and the radiation. Each of these variables is divided in four cases according to the presence or not of discharge and precipitation. Then, a finite mixture of Beta densities is selected for each of the four cases in the humidity and a finite mixture of Gamma densities for each case in the radiation. Their respective densities are,

$$f_H^{00}(h) = \sum_{i=1}^{K} \omega_i^{(H)} \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i)\Gamma(\beta_i)} h^{\alpha_i - 1}(1-h)^{\beta_i - 1},$$

$$f_R^{00}(r) = \sum_{i=1}^{K} \omega_i^{(R)} \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} r^{\alpha_i - 1} \exp(-\beta_i r),$$

and similar Beta mixtures for f_H^{10}, f_H^{01} and f_H^{11} and Gamma mixtures for f_R^{10}, f_R^{01} and f_R^{11}. Then, we may obtain u_h^{00}, u_h^{10}, u_h^{01}, u_h^{11}, u_r^{00}, u_r^{10}, u_r^{01} and u_r^{11} using the cumulative distribution function of a Beta mixture and a Gamma mixture respectively. Again, note that we have a set of parameters to estimate each of the four Beta mixtures and a set of parameters for each of the four Gamma mixtures.

For the precipitation, given that it is greater than zero, two cases are differentiated corresponding with the presence or absence of discharge. Then, a finite mixture of Gamma densities is selected for each case, where the density function can be expressed as,

$$f_P^{01}(p) = \sum_{i=1}^{K} \omega_i^{(P)} \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} p^{\alpha_i - 1} \exp(-\beta_i p),$$

and a similar Gamma mixture for f_P^{11}. Again, we may obtain u_p^{01} and u_p^{11}, using the cumulative distribution function of a Gamma mixture. Here, we only have a set of parameters for each of the two Gamma mixtures.

Finally, for the discharge and given that it is greater than zero, two cases are distinguished corresponding with the presence or absence of precipitation. Then, finite mixtures of Gamma densities are considered for each case whose density can be written as,

$$f_D^{10}(d) = \sum_{i=1}^{K} \omega_i^{(D)} \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} d^{\alpha_i - 1} \exp(-\beta_i d),$$

and similar Gamma mixture for f_D^{11}. Then, we may obtain u_d^{10} and u_d^{11}, using the cumulative distribution function of a Gamma mixture. Here, we have a set of parameters for each of the two gamma mixtures.

Finally, note that the number of terms in each mixture, K, could be different in each group and in every variable and it will be selected using model selection criteria as will be explained in Subsection 3.4.

3.3 Conditional probability

Once we have defined the multivariate model given by the copulas and the marginal distributions, we may obtain many quantities of interest. For example, we may obtain the conditional probability of zero discharge for one particular day whose meteorological variables have been observed. Using the Bayes theorem, this probability is given by,

$$P(D = 0 \mid T = t, H = h, R = r, P = p) = \frac{f(t, h, r, p \mid D = 0) \cdot \Pr(D = 0)}{f(t, h, r, p)}, \tag{6}$$

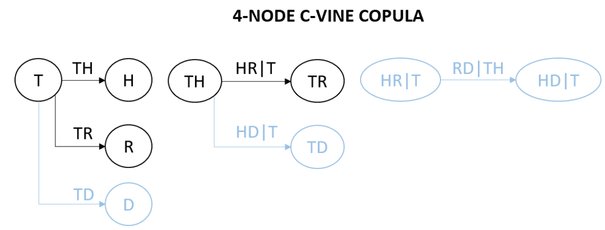

Figure 5: Structure of a c-vine copula with 3 nodes inherited from a c-vine copula with 4 nodes.

For the case when the precipitation is zero, the numerator in (6) can be expressed as,

$$f(t,h,r,P=0 \mid D=0)\cdot \Pr(D=0) = f(t,h,r \mid D=0,P=0)\cdot \Pr(D=0,P=0), \tag{7}$$

where the first term is obtained from (1a). And the denominator of (6) can be expressed as,

$$f(t,h,r,P=0) = f(t,h,r \mid D=0,P=0)\cdot \Pr(D=0,P=0) + f(t,h,r \mid D>0,P=0)\cdot \Pr(D>0,P=0),$$

where the first term is obtained from (7) and the second term can be obtained from (5) since it can be expressed in terms of a vine copula as,

$$f(t,h,r \mid D>0,P=0) = c^{10}_{HR|T}\left(u^{10}_{h|t},u^{10}_{r|t}\right)\cdot c^{10}_{TH}\left(u^{10}_{t},u^{10}_{h}\right)\cdot c^{10}_{TR}\left(u^{10}_{t},u^{10}_{r}\right)\cdot f^{10}_{T}(t)\cdot f^{10}_{H}(h)\cdot f^{10}_{R}(r),$$

and the terms of this expression already appear in the equation (5). Fig. 5 shows how the vine copula for describing the dependence of the variable (T,H,R,D) can be used to obtain the marginal dependence of $(T,H,R \mid D>0)$.

Similarly, for the case when the precipitation is positive, the numerator of (6) can be expressed as,

$$f(t,h,r,p \mid D=0)\cdot \Pr(D=0) = f(t,h,r,p \mid D=0,P>0)\cdot \Pr(D=0,P>0), \tag{8}$$

where the first term is obtained from (1c). And the denominator in (6) can be expressed as

$$f(t,h,r,p) = f(t,h,r,p \mid D=0,P>0)\cdot \Pr(D=0,P>0) + f(t,h,r \mid D>0,P>0)\cdot \Pr(D>0,P>0),$$

where the first term is obtained from (8) and the second term can be obtained from (17) since it can be expressed in terms of a vine copula as,

$$\begin{aligned} f(t,h,r,p) =& c^{11}_{RP|TH}\left(u^{11}_{hr|t},u^{11}_{hp|t}\right)\cdot c^{11}_{HR|T}\left(u^{11}_{h|t},u^{11}_{r|t}\right)\cdot c^{11}_{HP|T}\left(u^{11}_{h|t},u^{11}_{p|t}\right) \\ &\cdot c^{11}_{TH}\left(u^{11}_{t},u^{11}_{h}\right)\cdot c^{11}_{TR}\left(u^{11}_{t},u^{11}_{r}\right)\cdot c^{11}_{TP}\left(u^{11}_{t},u^{11}_{p}\right)\cdot f^{11}_{T}(t)\cdot f^{11}_{H}(h)\cdot f^{11}_{R}(r)\cdot f^{11}_{P}(p). \end{aligned} \tag{9}$$

As in the previous case, we can use the vine structure selected for describing the dependence of the variable (T, H, R, P, D) to obtain the vine copula for describing the marginal dependence of the variable $(T, H, R, P \mid D > 0)$. Thus, all the required terms in (9) can be found in (17).

Furthermore, using the defined multivariate distribution in (1), we may obtain the conditional distribution function of the discharge for given values of the meteorological variables, using the conditional probability in (6) as,

$$F_D(d \mid T = t, H = h, R = r, P = p) = \begin{cases} 0, & \text{with } \Pr(D = 0 \mid t, h, r, p) \\ \\ F(d \mid t, h, r, p, D > 0), & \text{with } 1 - \Pr(D = 0 \mid t, h, r, p) \end{cases} . \tag{10}$$

For the case when $p = 0$, the second part can be expressed in terms of the c-vine copulas as

$$F_D(d \mid t, h, r, P = 0, D > 0) = \frac{\partial C^{10}_{RD|TH}\left(u^{10}_{hr|t}, u^{10}_{hd|t}\right)}{\partial u^{10}_{hr|t}} = F\left(u^{10}_{hd|t} \mid u^{10}_{hr|t}\right). \tag{11}$$

For the case when $p > 0$, the second part of (10) can be expressed as,

$$F_D(d \mid t, h, r, p, D > 0) = \frac{\partial C^{11}_{PD|THR}\left(u^{11}_{rp|th}, u^{11}_{rd|th}\right)}{\partial u^{11}_{rp|th}} = F\left(u^{11}_{rd|th} \mid u^{11}_{rp|th}\right), \tag{12}$$

where,

$$u^{11}_{rp|th} = F_{RP|TH}(u^{11}_{hp|t} \mid u^{11}_{hr|t}) = \frac{\partial C_{RP|TH}(u^{11}_{hr|t}, u^{11}_{hp|t})}{\partial u^{11}_{hr|t}},$$

and similarly for $u^{11}_{rd|th}$.

3.4 Parameter estimation and model selection

Now, given the set of data on discharge and other meteorological variables, we want to estimate the parameters for our proposed model (1). First, we divide the sample in four groups according to the presence or not of discharge and/or precipitation and estimate the probabilities of each group, which correspond to the joint probabilities of having or not zero discharge and/or precipitation, using empirical frequencies. The parameters can be different in each group, but the estimation procedure is the same. First, we select the number of mixture components of the marginal functions for each variable using the Bayesian Information Criterion (BIC), which generally penalizes number of parameters in the model. Then, these mixture parameters are estimated by the maximum likelihood method, separately for each variable. Next, the values of u_x^{jk}, for $j = 1, 0$; $k = 1, 0$; and $x = t, h, r, p, d$, are obtained as explained in Subsection 3.2.

The next step is to fit the vine copula model to the data set. Note that here we assume the so called *simplifying assumption*, which imposes that each pair of copulas of conditional distributions does not depend on the values of the variables which are conditioned on. Although, this assumption has been criticized (Acar et al. 2012 and Spanhel and Kurz 2015), Killiches et al. (2016) propose the use of this assumption especially when the number of parameters is large. Moreover, Haff et al. (2010) came to the conclusion that vine copulas built with this assumption are "a rather good solution, even when the simplifying assumption is far from being fulfilled by the actual model".

In order to select a c-vine copula structure, we firstly need to set an order for the variables. As suggested by Aas et al. (2009), we try to set the variables with strongest dependencies in the first nodes of the tree. Therefore, we order the variables regarding on the values of the Kendall's tau. Thus, we have considered the temperature as the root variable, followed by the humidity, the radiation, the precipitation and, finally, the discharge. Appropriate pair-copula families are selected and estimated sequentially, using the BIC to determine the best copula family. For simplicity in the notation, we refer by u_x to the different possibilities u_x^{jk}, for $j = 1, 0$; $k = 1, 0$; and $x = t, h, r, p, d$, for example u_t denotes the u for the different cases of the temperature. The value of the parameters is estimated by maximum likelihood as follows,

1. Fit bivariate copulas for u_t and u_x, for $x = h, r, p, d$, for all the edges in the first tree.
2. Generate the series $u_{x|t} = F_{x|t}(u_i \mid u_t)$, for $x = h, r, p, d$, using the fitted copula from the previous step.
3. Fit bivariate copulas for $u_{h|t}$ and $u_{x|t}$, for $x = r, p, d$ for all the edges in the second tree.
4. Using the same procedure, generate series from the edges and fit copulas between the nodes for the remaining trees.

The copula for each node can be selected between an elliptical copula (Gaussian or t-copula) or an one-parameter Archimedean copula (Gumbel, Frank, Joe or Clayton). Before selecting the copula, an independence test, based on Kendall's tau, is performed on every pair of series using a significance level of

0.05%. All these estimations have been made using the functions available in the R package *VineCopula* (Schepsmeier et al. 2017; R Core Team 2013).

Three different estimators are considered and compared to obtain predictions of the future values of the discharge. First of all, we consider the median of the conditional distribution of the discharge given the observed meteorological values (10). This can be calculated as the value, $\hat{d}$, such that,

$$0.5 = P(D = 0 \mid t, h, r, p) + (1 - P(D = 0 \mid t, h, r, p)) \cdot F_D(\hat{d} \mid t, h, r, p; D > 0),$$

where the distribution function, F_D, is given in (11) if the observed precipitation is zero, or in (12) if it is different from zero.

Also, we consider the mean of the conditional distribution (10), given the observed values of temperature, humidity, radiation and precipitation in that day. This can be approximated using a Monte Carlo simulation by taking the sample mean of a set of simulated values from (10). This is detailed in Appendix A.

Finally, we propose a prediction method based on the conditional probability of zero discharge (6) and the conditional distribution function of the discharge given the values of the meteorological variables (10). Firstly, the conditional probability of the discharge is estimated. If this probability is larger than 0.5, we consider that the predictive discharge for that day is 0. If the estimated probability of zero discharge is smaller than 0.5, we estimate the mean of the conditional distribution when the discharge is positive in (11) and (12) using a Monte Carlo simulation as before.

In order to examine the performance of our proposed c-vine model we need two different evaluations, one for the probability of having zero discharge and other for the predicted amount of discharge with the different prediction methods. For the first one we use the Brier Score (Brier 1950), that measures the distance between the probability and the observation of an event,

$$BS = \frac{1}{n} \sum_{i=1}^{n} (p_i - o_i)^2, \tag{13}$$

where p_i is the probability that the event will happen and o_i takes value 1 if the event happens and 0 otherwise. For the predictive discharge we will use the Mean Squared Error (MSE) and the Mean Absolute

Error (MAE),

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{d}_i - d_i)^2, \tag{14}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \mid \hat{d}_i - d_i \mid, \tag{15}$$

where $\hat{d}_i$ is the estimated value and d_i is the true observed value.

4 Application

In this section, our proposed vine copula model is applied to the data provided by GLACKMA from their catchment area in glacier Collins in King George Island. First, the database is divided in groups according to four different hydrological periods to capture seasonality. Second, the model parameters, both the marginal distribution parameters and vine copula parameters, are estimated. Third, the conditional probability of having no discharge and the complete predictive discharge distribution is obtained using the estimated vine copula model. Finally, the obtained results are compared with those obtained with the bivariate copula model of our previous work (Gómez et al. 2017).

4.1 Parameter estimation

Recall that the GLACKMA database consists in five time series of data collected during eleven years. Here, the first ten years are used for parameter identification and data from 01/10/2011 to 30/09/2012 are used for model verification.

First of all, we want to capture the seasonal behaviour of the discharge. In our previous work (Gómez et al. 2017), we try to capture it using partial sums of Fourier terms, but this procedure increases rapidly the number of parameters when we add more meteorological variables. On the other side, Braun (2001) has found three major ablation phases plus a non-ablation phase for each year in glacier behaviour. This suggests us to divide the data in four different periods in order to capture the changes in the relationship between the variables. Table 1 shows the different periods selected for this study. As a justification of this division, Fig. 6 shows the boxplots of the average daily glacier, grouped by weeks, in the different periods. Apparently,

Period	Dates	Description
1	26th November - 30th December	Discharge start period. Since the last weeks of spring to early summer. Days can be positive or zero discharge.
2	31st December - 7th April	Main discharge period. Most of the summer. Almost every day have positive discharge.
3	8th April - 15th June	Discharge end period. Since the end of summer and most of autumn. Days can be zero or positive discharge.
4	16th June - 25th November	Zero discharge period. Late autumn, all the austral winter and early spring. There is always zero discharge.

Table 1: Distribution of the periods of discharge in King George Island.

there are different behaviors in the discharge regime. Note that the fourth period has zero discharge in the observed values. Thus, the model will always predict zero discharge in this period, that is, the equation (6) will always be zero independently of the values of the other variables because the empirical probability of having no discharge is equal to one.

Firstly, we determine the number of components and the mixture parameters of the marginal models for the first three periods. As an example, Fig. 7 shows the adjustment of the mixtures to the observations of the five variables in the second period for the case with positive discharge. The number of mixture components is shown at the bottom of each plot. An apparently good adjustment between the mixture models and the empirical distributions is observed for all variables and periods. The mixture marginal distribution parameter values are listed in Table 2. The first column indicates the period, the second refers to the group (j, k), where $j = 0, 1$, respectively, for zero or positive discharge and $k = 0, 1$, respectively, for zero or positive precipitation, and the third shows the number of observations available and used to fit the mixtures.

The following step is to select the copula family and its parameters for each edge in the different vine copula structures. Fig. 8 shows the structure of the c-vine copulas with the value of the parameter for every edge; each row in each edge correspond to one of the first three periods. Note that some edges have the independence copula, denoted by the letter **I**, this means that no dependence is found in that edge.

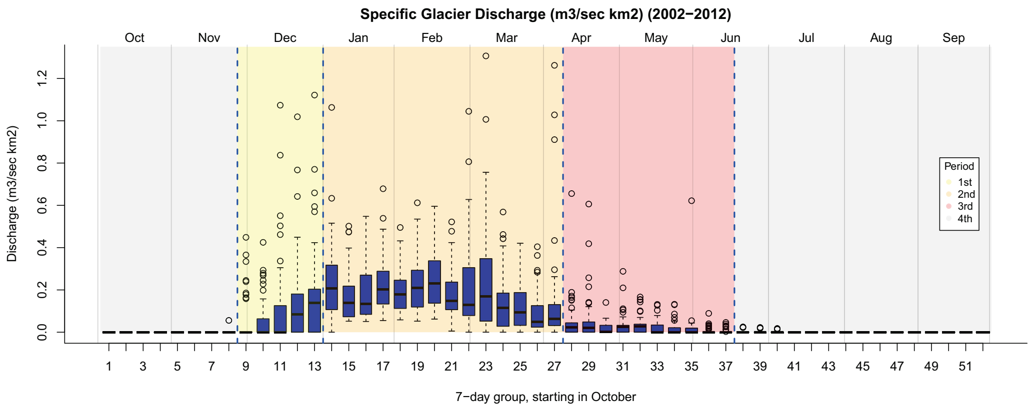

Figure 6: Boxplots of the glacier discharge in each week from 2002 to 2012. Different periods are separated by vertical lines and different color shadows.

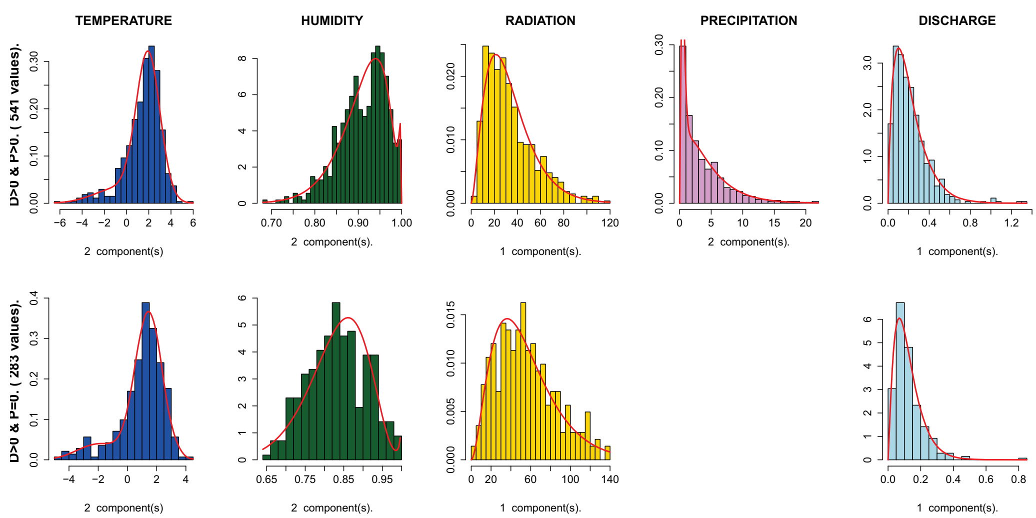

Figure 7: Histogram of the observed values of the variables, compared with the density function of the adjusted mixture. All histograms are for 2nd period and groups of data with positive discharge, with positive precipitation for the first row and without it for the second one. At the bottom of each plot appears the number of mixture components.

Per	Gr.	N			Temperature		
	DP		ω_1	μ_1	σ_1	μ_2	σ_2
1	00	40		-0.809 (0.191)	1.206 (0.135)		
	01	94		-0.163 (0.111)	1.074 (0.078)		
	10	68	0.220 (0.247)	-0.862 (1.513)	1.133 (0.651)	0.904 (0.170)	0.651 (0.129)
	11	83		0.915 (0.142)	1.295 (0.101)		
2	00	21	0.621 (0.106)	-2.370 (0.234)	0.829 (0.176)	0.665 (0.164)	0.460 (0.115)
	01	16		-1.298 (0.486)	1.944 (0.344)		
	10	283	0.145 (0.056)	-1.954 (0.790)	1.460 (0.430)	1.444 (0.101)	0.937 (0.073)
	11	541	0.177 (0.070)	-0.693 (0.814)	2.189 (0.266)	1.941 (0.082)	1.071 (0.077)
3	00	102		-5.375 (0.378)	3.817 (0.267)		
	01	225	0.665 (0.067)	-6.636 (0.499)	3.523 (0.262)	-0.992 (0.271)	1.230 (0.205)
	10	69		-2.363 (0.288)	2.392 (0.204)		
	11	234	0.517 (0.053)	-3.633 (0.377)	2.759 (0.208)	0.359 (0.107)	0.833 (0.086)
Per	Gr.	N			Humidity		
	DP		w_1	α_1	β_1	α_2	β_2
1	00	40		23.527 (5.314)	4.582 (0.99)		
	01	94		31.247 (4.716)	2.421 (0.332)		
	10	68		15.254 (2.648)	3.472 (0.570)		
	11	83		24.086 (3.847)	2.589 (0.379)		
2	00	21		40.476 (12.524)	10.439 (3.173)		
	01	16		16.284 (5.840)	3.326 (1.123)		
	10	283	0.029 (0.019)	6.016 (6.587)	0.240 (0.129)	19.277 (2.065)	3.957 (0.442)
	11	541	0.967 (0.012)	24.565 (1.859)	2.504 (0.192)	305.374 (195.601)	1.829 (0.723)
3	00	102		19.921 (2.829)	3.662 (0.492)		
	01	225		22.324 (2.15)	3.027 (0.271)		
	10	69		14.021 (2.431)	2.783 (0.449)		
	11	234		18.336 (1.761)	2.022 (0.174)		
Per	Gr.	N			Radiation		
	DP		ρ_1	α_1	β_1	α_2	β_2
1	00	40	0.455 (0.09)	69.832 (33.129)	0.605 (0.279)	86.14 (34.047)	1.146 (0.466)
	01	94		12.000 (1.726)	0.184 (0.027)		
	10	68		13.508 (2.287)	0.148 (0.026)		
	11	83		6.229 (0.942)	0.104 (0.016)		
2	00	21		6.980 (2.104)	0.236 (0.074)		
	01	16		2.543 (0.846)	0.103 (0.038)		
	10	283		2.913 (0.232)	0.053 (0.005)		
	11	541		2.721 (0.156)	0.081 (0.005)		
3	00	102	0.591 (0.071)	6.542 (1.492)	2.707 (0.729)	4.308 (1.864)	0.381 (0.142)
	01	225	0.536 (0.150)	4.347 (1.553)	1.945 (0.879)	1.858 (0.456)	0.284 (0.052)
	10	69		1.608 (0.251)	0.184 (0.034)		
	11	234		1.838 (0.157)	0.339 (0.033)		
Per	Gr.	N			Precipitation		
	DP		ρ_1	α_1	β_1	α_2	β_2
1	01	40	0.607 (0.188)	2.06 (0.597)	1.753 (0.896)	1.381 (0.513)	0.277 (0.093)
	11	94		1.242 (0.173)	0.583 (0.099)		
2	01	21		0.731 (0.221)	0.213 (0.090)		
	11	16	0.235 (0.057)	3.487 (0.960)	5.825 (2.337)	1.482 (0.192)	0.326 (0.032)
3	01	102	0.219 (0.051)	7.817 (2.848)	17.953 (7.749)	1.490 (0.186)	0.548 (0.067)
	11	225	0.245 (0.079)	3.503 (1.205)	5.018 (2.611)	1.406 (0.196)	0.354 (0.045)
Per	Gr.	N			Discharge		
	DP		ρ_1	α_1	β_1	α_2	β_2
1	10	40		1.788 (0.283)	11.055 (2.015)		
	11	94		1.640 (0.233)	8.372 (1.390)		
2	10	21		2.247 (0.177)	18.094 (1.593)		
	11	16		1.871 (0.105)	8.525 (0.549)		
3	10	102	0.936 (0.035)	4.029 (0.787)	127.662 (28.612)	18.01 (16.619)	139.819 (121.816)
	11	225	0.355 (0.054)	0.986 (0.143)	9.555 (1.899)	6.744 (1.214)	221.226 (43.461)

Table 2: Mixture parameters. The first column shows the period and the second indicates if the group has positive discharge (1 in the first digit) and positive precipitation (1 in the second digit). Third column informs about the number of observed values. The number between parenthesis is the error on the parameter estimation.

Order	BIC	Vuong statistic	p-value
THRPD	-19.438	0	1
TDPRH	-19.238	0.190	0.849
HTRPD	-19.194	0.079	0.937
HDRPT	-19.641	-0.042	0.967
RPDTH	-13.809	1.315	0.188
RTDPH	-19.021	0.131	0.896
PTRHD	-13.343	1.191	0.234
PDTRH	-12.591	1.640	0.101
DPRHT	-17.347	0.345	0.730
DTHRP	-20.409	-0.183	0.855

Table 3: BIC value of different order combinations for the 5-cvine copula in the first period. Vuong test of comparison with the selected order (THRPD) and the correspondent p-value.

Group	Period 1		Period 2		Period 3	
	White	p-value	White	p-value	White	p-value
00	02.13	0.15	09.24	0.32	08.51	0.60
01	21.53	0.14	19.09	0.60	17.80	0.42
10	19.19	0.79	16.00	0.39	21.76	0.39
11	60.83	0.83	72.90	0.79	73.62	0.06

Table 4: White statistic and p-value of the goodness-of-fit test over the twelve c-vine copulas for the selected order.

Different order in the variables within the c-vine structure has been considered and compared using the BIC criteria and very close values have been found. Also, we have used the Vuong test (Vuong 1989) to look for differences between different orders but no significative difference have been found. Table 3 shows some of the results obtained for the c-vine copula for days with positive discharge and precipitation in the first period (5 nodes). The results for other groups and periods are similar. Then, we have selected the more convenient order to facilitate the evaluation of the probability of discharge and predictive discharge, that is T-H-R-P-D. Also, a goodness-of-fit test has been performed for each of the twelve copulas obtained with the proposed order, which is based on the information matrix equality of White, as detailed in Schepsmeier (2016). Table 4 shows the White statistic and the correspondent p-value for each c-vine copula. Observe that both the model and the parameters seem to be appropriate.

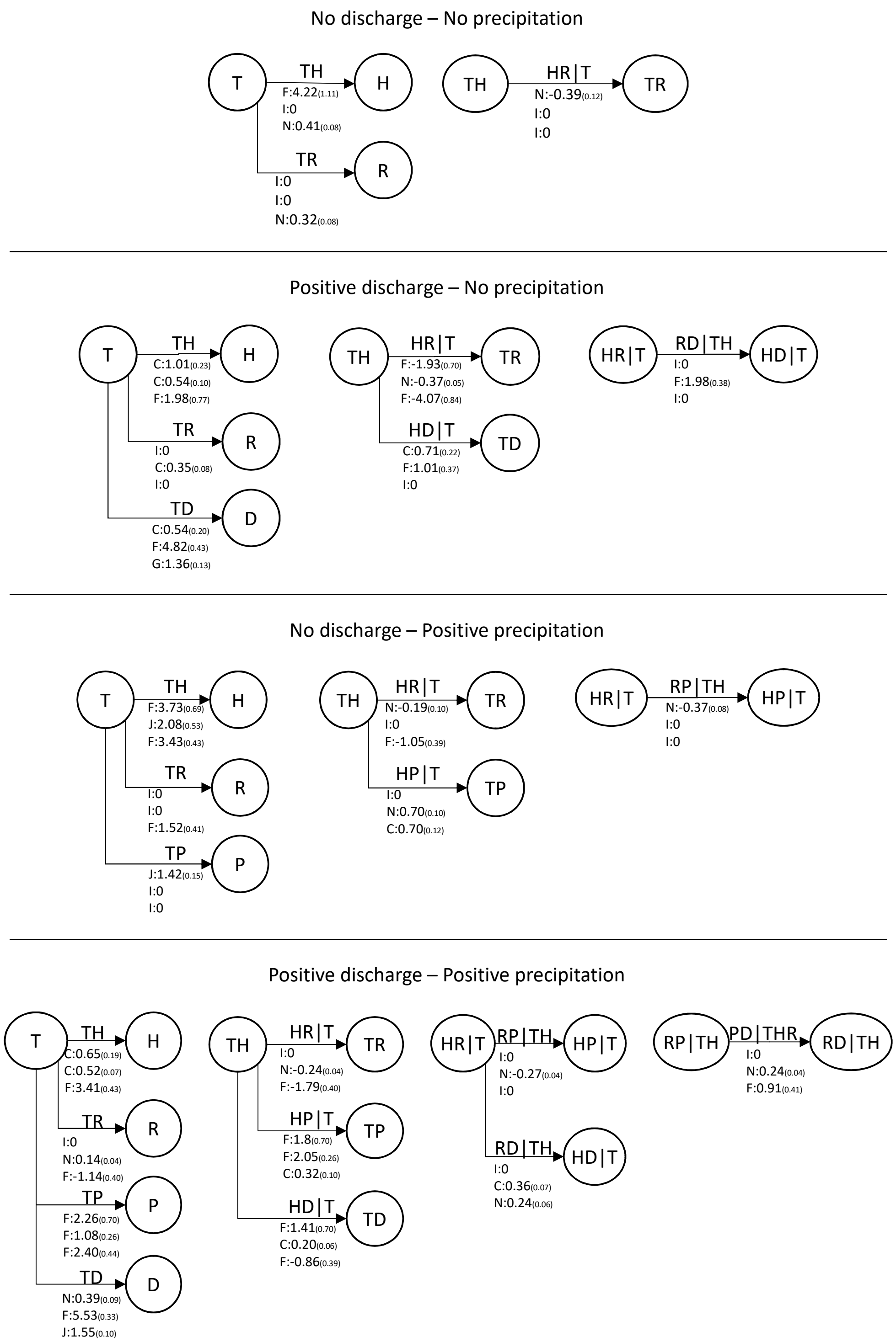

Figure 8: c-vine copula parameters for all periods and groups. The values in each node are for each period. The copulas are I=Independent, N=Gaussian, C=Clayton, G=Gumbel, F=Frank, J=Joe. The parameter estimation error is between parenthesis.

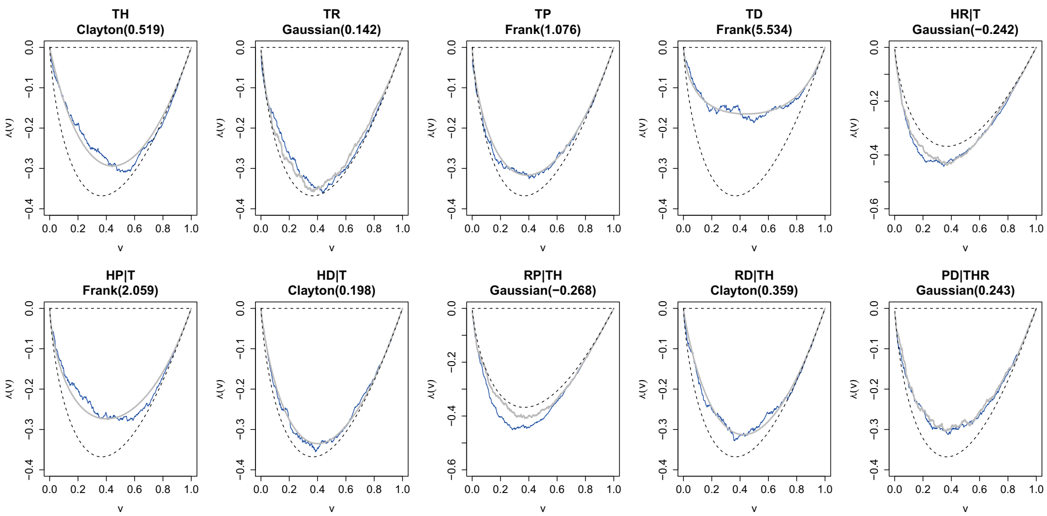

Figure 9: Empirical λ-function for the 10 nodes of the 2nd period and group with discharge and precipitation. The blue line is the empirical and the grey one is the theoretical. The dashed lines are bounds corresponding to independence and comonotonicity ($\lambda = 0$) copulas, respectively.

In order to examine the goodness of fit of the estimated bivariate copulas in each c-vine structure, we make use of the $\lambda-$function (Genest and Rivest 1993). As an illustration, Fig. 9 shows the comparison between the empirical $\lambda-$function for each edge and the theoretical $\lambda-$function for the corresponding copula, for the 10 edges of the c-vine copula, for the second period and the group for data with positive values of discharge and precipitation. The dashed lines are bounds corresponding to the independence and comonotonicity copulas, respectively. For the sake of brevity, only one tree has been shown. Apparently, there is good fit between the selected and the empirical copula in all edges for all selected vine structures.

4.2 Conditional probability of discharge

Once we have obtained all the model parameters, we are interested in estimating the probability of having zero discharge conditioned to the observed values for the temperature, humidity, radiation and precipitation in each day. As commented above, we will predict positive discharge, with (6), for a particular day if the estimated conditional probability of zero discharge is smaller than 0.5. Table 5 compares predictions with the observed values of the discharge. For the in-sample data (2002-2011). We obtain the 92.7% of days with zero discharge are correctly predicted with the c-vine model, whereas 88.6% of days with positive discharge

		Predicted					
		2002-2011			2011-2012		
		$D = 0$	$D > 0$	Total	$D = 0$	$D > 0$	Total
Observed	$D = 0$	1883	149	2032	206	21	227
	$D > 0$	147	1148	1295	12	127	139

Table 5: Comparison between observed discharge and predictions with the vine copula model. On the left, for days with the data used to fit the model (2002-2011). On the right, for the days of the last year (2011-2012), used to validate the model.

	2002-2011		2011-2012	
	Logistic model	Vine model	Logistic model	Vine model
Global	**0.0643**	**0.0607**	**0.0815**	**0.0761**
Period 12	0.1401	0.1314	0.2474	0.2900
Period 2	0.0359	0.0320	0.0254	0.0240
Period 3	0.1999	0.1904	0.1861	0.1463

Table 6: Comparison between the Brier Score obtained with a logistic regression and with the vine copula model. On the left, for days with the data used to fit the model (2002-2011). On the right, for the days of the last year (2011-2012), used to validate the model.

are correctly predicted. The performance of the copula model is even better for the out-of-sample data from the last hydrological year, used to validate the model. Our model has a 90.9% and 90.7% of correctly predicted days for days with zero and positive discharge respectively. We have compared these probabilities with the ones obtained with a logistic regression, which has been developed in the same conditions as the vine model, that is, one model for each period. Table 6 shows the Brier Score (13) for both models, we can see that the vine copula model outperforms the logistic regression, globally and in each period and for the in-sample and the out of sample data. Note that, the smaller the Brier Score the better the predictions.

Additionally, we are also interested in studying the behaviour of this conditional probability of discharge in terms of the conditioned meteorological variables. As an illustration, Fig. 10 shows the estimated probability of positive discharge as a function of the temperature for different values of the humidity, in the presence or absence of precipitation and a fixed value for the radiation. Note that the positive precipitation increases the probability of zero discharge, especially when the temperatures are below zero. In both plots we can also

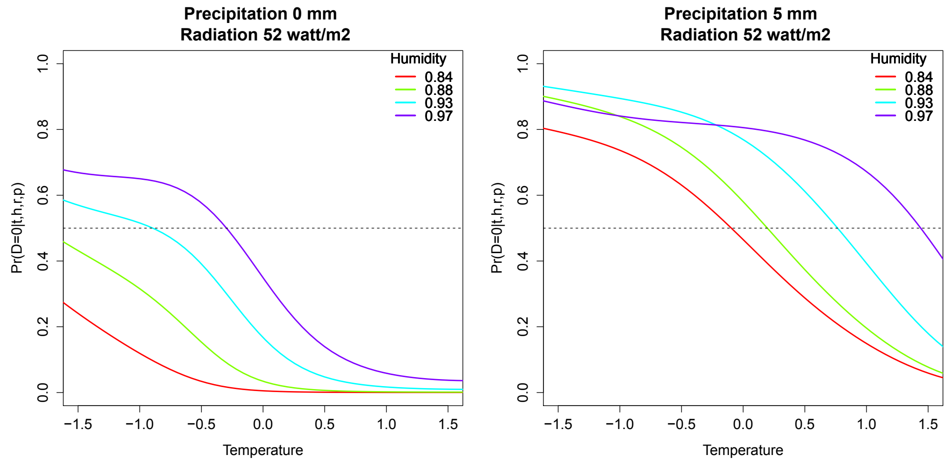

Figure 10: Evolution of the probability of having zero discharge during the first period conditioned to different values of the meteorological variables.

see that higher temperatures cause a decay in the probability of having zero discharge and that an increase of the percentage of humidity increases the probability of having zero discharge.

4.3 Predictive discharge

Finally, as described in Subsection 3.3, the predictive discharge distribution from (10) have been obtained, for all days and three point estimations for the discharge using the three methods explained in Subsection 3.4. These predictions have been compared with those obtained with our previous bivariate copula model (Gómez et al. 2017). Table 7 shows the MSE (14) and the MAE (15) obtained for both models. We may observe that in all cases the errors with the vine copula model are smaller than those obtained with the the bivariate copula model. Then, clearly, the vine copula model gives more accurate predictions of the discharge. Finally, the proposed model is validated with all described methods with the observed values of the discharge for the year (2011-12). The two last columns of Table 7 show these measures of the MSE and MAE. It can be observed that the errors of the proposed model are smaller than the ones produced by the bivariate copula model. Therefore, it can be concluded that the use of more meteorological variables in the proposed vine copula model provides more accurate predictions that using simply the temperature as in our previous bivariate copula model.

As an example, the left panel of Fig. 11 shows the observed values of the discharge for the year 2005-06

		2002-2011		2011-2012	
Model	Method	SME	MAE	SME	MAE
Vine copula model	Median	0.00621	0.02798	0.01212	0.04682
	Mean	0.00605	0.03175	0.01084	0.04871
	Proposed method	0.00608	0.03031	0.01061	0.04489
Bivariate copula model		0.00718	0.03317	0.03753	0.05362

Table 7: Errors of the predicted discharge when vine copula model and bivariate copula model are used. The first two columns have been obtained with the data used to fit the model. The other two have been obtained with the data of the last year, used to validate.

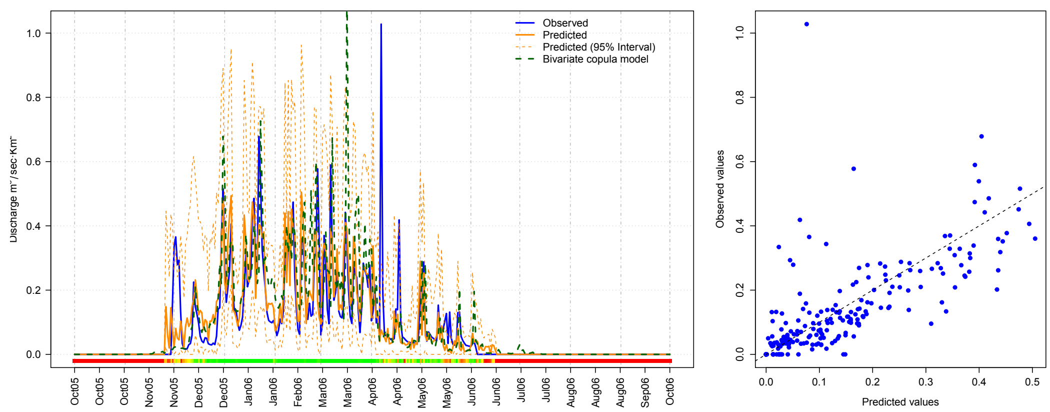

Figure 11: Left panel contains the time series of the observed values of the discharge, predictions with both models and 95% credible intervals for the vine copula model in the year 2005-06. The bottom of the plot shows the conditional probability of discharge of each day in a scale from red (probability zero) to green (probability one). Right panel contains the comparison between the predicted and the observed values with the proposed method in the vine copula model.

compared with the predictive discharge obtained with the proposed vine copula model, together with the corresponding 95% credible intervals, and the predictions obtained with the bivariate copula model. Also, the bottom of the plot illustrates the conditional probability of discharge zero, from red for probability 1 to green for probability 0. The left panel of Fig. 11 shows the scatter plot between the predicted and the observed values.

5 Conclusion and future work

In this paper, we have proposed a vine copula model for modelling the relationship between the glacier discharge and other meteorological variables, such as, temperature, humidity, solar radiation and precipitation. The probability of zero discharge for each future day is determined given the observed values of the meteorological variables. Also, the predictive value of the discharge is obtained from its conditional distribution given the observations of the meteorological variables. This model has been applied to the data collected by GLACKMA from the glacier Collins between 2002 and 2012. The database has been divided into four period according to the different hydrological seasons and the parameters have been adjusted to obtain the joint distribution of the five variables in each one of these periods. The model has a good performance in all the periods.

Observe that in this work only c-vine structures has been explored. However, other vine copula structures such as d-vine copulas or, in general, r-vine copulas could be analyzed. In addition, alternative bivariate copula models could be introduced.

The monitor station in King George island have been registering data which have not been already collected by the GLACKMA association. Our intention is to validate our proposed model with these new data whenever they are available. Moreover, the proposed model could be used in other glaciers whose discharge data is being collected by this association from their Pilot Experimental Watersheds, which are installed in different glaciers in both hemispheres at different latitudes and altitudes. Furthermore, the model could be used to predict the discharge in glaciers where measuring the real discharge is complex and only the meteorological data is available.

A Appendix I: Algorithm

In this appendix, we explain the algorithm to obtain the predictive values of the discharge with the conditional probability given in (10). Algorithm 1 details the estimation procedure to obtain the predictive mean of the glacier discharge given the temperature, humidity, radiation and precipitation.

Algorithm 1 Predictive discharge (using the mean)

Require: $t, h, r, p, \theta, F_T, F_H, F_R, F_P, F_D$

procedure
    if p=0 **then**
        Compute $u_t = F_T(t)$, $u_h = F_H(h)$ and $u_r = F_R(r)$
        Compute $u_{th} = C_{TH}(u_h \mid u_t; \theta_{TH})$ and $u_{tr} = C_{TR}(u_r \mid u_t; \theta_{TR})$
        Compute $u_{thr} = C_{HR|T}(u_{tr} \mid u_{th}; \theta_{THR})$
        Simulate $u_{thrd} \sim U(0,1)$
        Obtain the value u_{thd} that verify $u_{thrd} = C_{RD|TH}(u_{thd} \mid u_{thr}; \theta_{THRD})$
        Obtain the value u_{td} that verify $u_{thd} = C_{HD|T}(u_{td} \mid u_{th}; \theta_{THD})$
        Obtain the value u_d that verify $u_{td} = C_{TD}(u_d \mid u_t; \theta_{TD})$
        Obtain $\widehat{d} = F_D^{-1}(u_d)$
    else
        Compute $u_t = F_T(t)$, $u_h = F_H(h)$, $u_r = F_R(r)$ and $u_p = F_P(p)$
        Compute $u_{th} = C_{TH}(u_h \mid u_t; \theta_{TH})$, $u_{tr} = C_{TR}(u_r \mid u_t; \theta_{TR})$ and $u_{tp} = C_{TP}(u_p \mid u_t; \theta_{TP})$
        Compute $u_{thr} = C_{HR|T}(u_{tr} \mid u_{th}; \theta_{THR})$ and $u_{thp} = C_{HP|T}(u_{tp} \mid u_{th}; \theta_{THP})$
        Compute $u_{thrp} = C_{RP|TH}(u_{thp} \mid u_{thr}; \theta_{THP})$
        Simulate $u_{thrpd} \sim U(0,1)$
        Obtain the value u_{thrd} that verify $u_{thrdp} = C_{PD|THR}(u_{thrd} \mid u_{thrp}; \theta_{THRPD})$
        Obtain the value u_{thd} that verify $u_{thrd} = C_{RD|TH}(u_{thd} \mid u_{thr}; \theta_{THRD})$
        Obtain the value u_{td} that verify $u_{thd} = C_{HD|T}(u_{td} \mid u_{th}; \theta_{THD})$
        Obtain the value u_d that verify $u_{td} = C_{TD}(u_d \mid u_t; \theta_{TD})$
        Obtain $\widehat{d} = F_D^{-1}(u_d)$
    end if
end procedure

For the case that we want to estimate the predictive median of the discharge, we may replace in Algorithm 1 the, instructions (6) and (16) by "Compute $u_{thrpd} = 1 - \frac{0.5}{\Pr(D=0|t,h,r,p)}$" and "Compute $u_{thrd} = 1 - \frac{0.5}{\Pr(D=0|t,h,r)}$" respectively.

Finally, for the last prediction method, the conditional probability, $\Pr(D = 0 \mid t, h, r, p)$ is estimated at the beginning of the algorithm and then, it is predicted that $\widehat{d} = 0$ if the estimated probability of zero discharge is greater than 0.5 or obtained with the algorithm if it is smaller.

B Appendix II: Density functions as copulas

The joint density functions in (1c) and (1d) can be expressed in terms of a vine copulas as,

$$
\begin{aligned}
f(t,h,r,p \mid D=0,P>0) =& c^{01}_{RP|TH}\left(u^{01}_{HR|T}, u^{01}_{HP|T}\right) \cdot c^{01}_{HR|T}\left(u^{01}_{H|T}, u^{01}_{R|T}\right) \cdot c^{01}_{HP|T}\left(u^{01}_{H|T}, u^{01}_{P|T}\right) \\
&\cdot c^{01}_{TH}\left(u^{01}_{t}, u^{01}_{h}\right) \cdot c^{01}_{TR}\left(u^{01}_{t}, u^{01}_{r}\right) \cdot c^{01}_{TP}\left(u^{01}_{t}, u^{01}_{p}\right) \\
&\cdot f^{01}_{T}(t) \cdot f^{01}_{H}(h) \cdot f^{01}_{R}(r) \cdot f^{01}_{P}(p),
\end{aligned} \tag{16}
$$

$$
\begin{aligned}
f(t,h,r,p,d \mid D>0,P>0) =& c^{11}_{PD|THR}(u^{11}_{rp|th}, u^{11}_{rd|th}) \cdot c^{11}_{RP|TH}\left(u^{11}_{hr|t}, u^{11}_{hp|t}\right) \cdot c^{11}_{RD|TH}\left(u^{11}_{hr|t}, u^{11}_{hd|t}\right) \\
&\cdot c^{11}_{HR|T}\left(u^{11}_{h|t}, u^{11}_{r|t}\right) \cdot c^{11}_{HP|T}\left(u^{11}_{h|t}, u^{11}_{p|t}\right) \cdot c^{11}_{HD|T}\left(u^{11}_{h|t}, u^{11}_{d|t}\right) \\
&\cdot c^{11}_{TH}\left(u^{11}_{t}, u^{11}_{h}\right) \cdot c^{11}_{TR}\left(u^{11}_{t}, u^{11}_{r}\right) \cdot c^{11}_{TP}\left(u^{11}_{t}, u^{11}_{p}\right) \cdot c^{11}_{TD}\left(u^{11}_{t}, u^{11}_{d}\right) \\
&\cdot f^{11}_{T}(t) \cdot f^{11}_{H}(h) \cdot f^{11}_{R}(r) \cdot f^{11}_{P}(p) \cdot f^{11}_{D}(d),
\end{aligned} \tag{17}
$$

where the superscripts denote the condition of zero or non-zero of discharge and precipitation values: $(01 : D = 0, P > 0)$ and $(11 : D > 0, P > 0)$, and (omitting the superscripts for more clarity)

$u_{rp|th} = F_{RP|TH}(u_{hp|t} \mid u_{hr|t})$, $u_{rd|th} = F_{RD|TH}(u_{hd|t} \mid u_{hr|t})$

$u_{hr|t} = F_{HR|T}(u_{r|t} \mid u_{h|t})$, $u_{hp|t} = F_{HP|T}(u_{p|t} \mid u_{h|t})$, $u_{hd|t} = F_{HD|T}(u_{d|t} \mid u_{h|t})$

$u_{h|t} = F_{H|T}(u_h \mid u_t)$, $u_{r|t} = F_{R|T}(u_r \mid u_t)$, $u_{p|t} = F_{P|T}(u_p \mid u_t)$, $u_{d|t} = F_{D|T}(u_d \mid u_t)$

$u_t = F_T(t)$, $u_h = F_H(h)$, $u_r = F_R(r)$, $u_p = F_P(p)$, $u_d = F_D(d)$

$F_{H|T}(u_h \mid u_t) = \frac{\partial C_{TH}(u_t,u_h)}{\partial u_t}$ and similar expressions for the other variables

$F_{HR|T}(u_{r|t} \mid u_{h|t}) = \frac{\partial C_{HR|T}(u_{h|t}, u_{r|t})}{\partial u_{h|t}}$ and similar expressions for the other variables

$F_{RP|TH}(u_{hp|t} \mid u_{hr|t}) = \frac{\partial C_{RP|TH}(u_{hr|t}, u_{hp|t})}{\partial u_{hr|t}}$ and similar expressions for the other variables

Figures in Table 8 show the structures of the different c-vine copulas used in this paper.

Acknowledgements

We are very grateful to the GLACKMA association. The first and second authors acknowledge financial support by MEC project ECO2015-66593-P from the Spanish Government. The third author would like to thank the Russian, Argentinean, German, Uruguayan and Chilean Antarctic Programs for their continuous logistic support over the years. The crews of Bellingshausen, Artigas, and Carlini station as well as the Dallmann Laboratory provided a warm and pleasant environment during fieldwork. GLACKMA's contribution was also partially financed by the European Science Foundation, ESF project IMCOAST (EUI2009-04068) and the Ministerio de Educación y Ciencia (CGL2007-65522-C02-01/ANT).

References

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

[2] Acar, E. F., Genest, C., NešLehová, J. (2012). Beyond simplified pair-copula constructions. *Journal of Multivariate Analysis*, 110, 74-90.

[3] Barrand, N.E., Vaughan, D.G., Steiner, N., Tedesco, M., Kuipers Munneke, P., Broeke, M.R., Hosking, J.S. (2013). Trends in Antarctic Peninsula surface melting conditions from observations and regional climate modeling. *Journal of Geophysical Research: Earth Surface*, 118(1), 315-330.

[4] Bedford, T. and R. M. Cooke (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence* 32, 245-268.

[5] Braun, M. (2001). Ablation on the ice cap of King George Island (Antarctica). (Doctoral thesis, University of Freiburg, Germany).

[6] Braun M et al. (2002) Satellite image map of King George Island, Antarctica. Department of Physical Geography, Albert-Ludwigs-Universit Freiburg. https://doi.pangaea.de/10.1594/PANGAEA.114703

[7] Brechmann, E., Czado, C., Paterlini, S. (2014). Flexible dependence modeling of operational risk losses and its impact on total capital requirements. *Journal of Banking & Finance*, 40, 271-285.

[8] Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. Monthly weather review, 78(1), 1-3.

[9] Carnicero, J.A, Ausín, M.C., Wiper, M.P., (2013) Non-parametric copulas for circular-linear and circular-circular data: an application to wind directions. *Stoch Environ Res Risk Assess*, 27(8), 1991-2002.

[10] Cogley, J.G., Hock, R., Rasmussen, L.A., Arendt, A.A., Bauder, A., Braithwaite, R.J., Jansson, P., Kaser,G., Mller, M., Nicholson, L., Zemp, M., (2011). *Glossary of Glacier Mass Balance and Related Terms*, IHP-VII Technical Documents in Hydrology No. 86, IACS Contribution No. 2, UNESCO-IHP, Paris.

[11] Cong, R., Brady, M. (2012). The interdependence between rainfall and temperature: Copula analyses. *The Scientific World Journal*, vol. 2012.

[12] De Michele, C., Salvadori, G. (2003). A generalized Pareto intensity duration model of storm rainfall exploiting 2-copulas. *Journal of Geophysical Research: Atmospheres*, 108(D2).

[13] De Michele, C., Salvadori, G., Canossi, M., Petaccia, A., Rosso, R. (2005). Bivariate statistical approach to check adequacy of dam spillway. *Journal of Hydrologic Engineering*, 10(1), 50-57.

[14] Domínguez, M. C., Eraso, A., (2007). Substantial changes happened during the last years in the icecap of King George, Insular Antactica. In: Tyk, A., Stefaniak, K. (eds.): *Karst and Cryokarst, Studies of the Faculty of Earth Sciences*, University of Silesia 45, 87-110.

[15] Erhardt, V., Czado, C. (2012). Modeling dependent yearly claim totals including zero claims in private health insurance. *Scandinavian Actuarial Journal*, 2012(2), 106-129.

[16] Favre, A. C., El Adlouni, S., Perreault, L., Thimonge, N., Bobe, B. (2004). Multivariate hydrological frequency analysis using copulas. *Water Resources Research*, 40(1). http://dx.doi.org/10.1029/2003WR002456

[17] Genest C, Rivest L. P. (1993). Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association* 88, 1034-1043.

[18] Genest, C., Favre, A. C., (2007). Everything you always wanted to know about copula but were afraid to ask. *Journal of Hydrology Engineering*, 4, 347-368.

[19] Gomez, M., Ausin, M.C., Dominguez, M.C. (2017). Seasonal copula models for the analysis of glacier discharge at King George Island, Antarctica. *Stoch Environ Res Risk Assess*, 31, 1107-1121. http://dx.doi.org/10.1007/s00477-016-1217-7.

[20] Gyasi-Agyei, Y., Melching, C.S. (2012). Modelling the dependence and internal structure of storm events for continuous rainfall simulation. *Journal of Hydrology*, 464, 249-261.

[21] Gyasi-Agyei, Y. (2013). Evaluation of the effects of temperature changes on fine timescale rainfall. *Water Resources Research*, 49(7), 4379-4398. http://dx.doi.org/10.1002/wrcr.20369

[22] Haff, I. H., Aas, K., Frigessi, A. (2010). On the simplified pair-copula construction – Simply useful or too simplistic?. *Journal of Multivariate Analysis*, 101(5), 1296-1310.

[23] Hamlet, A. F., Lettenmaier, D. P. (1999). Effects of climate change on hydrology and water resources in the Columbia River Basin. *Journal of the American Water Resources Association*, 35(6), 1597-1623.

[24] Hao, Z., Singh, V. P. (2012). Entropy-copula method for single-site monthly streamflow simulation. *Water Resources Research*, 48(6), 1944-7973. http://dx.doi.org/10.1029/2011WR011419

[25] Hock, R. (1999). A distributed temperature-index ice-and snowmelt model including potential direct solar radiation. *Journal of Glaciology*, 45(149), 101-111.

[26] Hock, R. (2003). Temperature index melt modelling in mountain areas. *Journal of Hydrology*, 282(1), 104-115.

[27] Killiches, M., Kraus, D., Czado, C. (2017). Examination and visualisation of the simplifying assumption for vine copulas in three dimensions. Australian & New Zealand Journal of Statistics, 59(1), 95-117.

[28] Marsh, P. (1999). Snowcover formation and melt: recent advances and future prospects. *Hydrological Processes*, 13, 2117-2134.

[29] Nelsen, R. B. (2006), *An introduction to Copulas*, Second Edition. Springer Science & Business Media, New York.

[30] Pellicciotti, F., Brock, B., Strasser, U., Burlando, P., Funk, M., Corripio, J. (2005). An enhanced temperature-index glacier melt model including the shortwave radiation balance: development and testing for Haut Glacier d'Arolla, Switzerland. *Journal of Glaciology*, 51(175), 573-587.

[31] R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

[32] Rückamp, M., Braun, M., Suckro, S., Blindow, N., (2011). Observed glacial changes on the King George Island ice cap, Antarctica, in the last decade. *Global and Planetary Change*, Volume 79, Pages 99-109

[33] Sarhadi, A., Burn, D. H., Ausín, M.C., Wiper, M.P. (2016). Time varying nonstationary multivariate risk analysis using a dynamic Bayesian copula. *Water Resources Research*, 52, 2327-2349, http://doi:10.1002/2015WR018525.

[34] Scholzel, C., Friederichs, P., (2008). Multivariate non-normally distributed random variables in climate research - introduction to the copula approach. *Nonlinear Processes in Geophysics*, 15, 761-772

[35] Schepsmeier, U. (2016). A goodness-of-fit test for regular vine copula models. Econometric Reviews, 1-22.

[36] Sicart, J.E., Hock, R., Six, D. (2008). Glacier melt, air temperature, and energy balance in different climates: The Bolivian Tropics, the French Alps, and northern Sweden. *Journal of Geophysical Research: Atmospheres*, 113(D24).

[37] Sklar, A., (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229-231.

[38] Spanhel, F., Kurz, M. S. (2015). Simplified vine copula models: Approximations based on the simplifying assumption. arXiv preprint arXiv:1510.06971.

[39] Turner, J., Colwell, S.R., Marshall, G.J., Lachlan-Cope, T.A., Carleton, A.M., Jones, P.D., Lagun, V., Reid, P.A., Iagovkina, S. (2005). Antarctic climate change during the last 50 years. *International journal of Climatology*. 25(3):279-294

[40] Vaughan, D. G., Marshall, G.J., Connolley, W. M., Parkinson, C., Mulvaney, R., Hodgson, D. A., ... Turner, J. (2003). Recent rapid regional climate warming on the Antarctic Peninsula. *Climatic change*, 60(3), 243-274.

[41] Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica: Journal of the Econometric Society, 307-333.

[42] Xiong, L., Jiang, C., Xu, C.Y., Yu, K.X., Guo, S. (2015), A framework of change-point detection for multivariate hydrological series, *Water Resources Research*, 51, 8198-8217. http://dx.doi.org/10.1002/2015WR017677

[43] Schepsmeier, U., Stoeber, J., Brechmann, EC., Graeler, B., Nagler, T., Erhardt, T. (2017). VineCopula: Statistical Inference of Vine Copulas. R package version 2.1.2. https://CRAN.R-project.org/package=VineCopula