Supporting Information for

Streamflow estimation at partially gaged sites using multiple dependence conditions via vine copulas

Kuk-Hyun Ahn[1]

1 Assistant Professor, Department of Civil and Environmental Engineering, Kongju National University, Cheon-an, South Korea; Corresponding author; e-mail: ahnkukhyun@gmail.com

Contents of this file

Introduction

This supporting information provides additional descriptions to support the conclusions of the primary article. The theoretical description for $\mathcal{M}_{Bivar}$ and $\mathcal{M}_{\mathrm{Kraus}}$ are first presented. Next, the theoretical description for upper and lower tail dependences is demonstrated.

Text S1. Description of $\mathcal{M}_{Bivar}$

The optimal bivariate copula is developed by selecting the minimum AIC values while considering the five bivariate copulas (Gaussian, Student-t, Frank, Gumbel, and Clayton copulas) described follows.

S1.1 Gaussian copula

The density of the Gaussian copula is given by

$$c\big(F_1(q_1), F_2(q_2)\big) = \frac{1}{\sqrt{1-\rho^2}}\ \exp\Big[-\frac{\rho^2(q_1^2+q_2^2)-2\rho^2 q_1 q_2}{2(1-\rho^2)}\Big] \qquad \text{Eq. (S1)}$$

where $F_1(q_1)$ and $F_2(q_2)$ are the marginal distribution functions of streamflow at two sites in the range [0, 1].

The h-function of the Gaussian copula is expressed as

$$h(F_1(q_1), F_2(q_2), \rho) = F\Big(\frac{F_1^{-1}(q_1)-\rho F_2^{-1}(q_2)}{\sqrt{1-\rho^2}}\Big) \qquad \text{Eq. (S2)}$$

S1.2 Student-t copula

The density of the Student-t copula is given by

$$\boldsymbol{c\big(F_1(q_1),F_2(q_2)\big)} = \frac{\Gamma(\frac{\varphi+2}{2})/\Gamma(\frac{\varphi}{2})}{\varphi\pi dt(q_1,\varphi)dt(q_2,\varphi)\sqrt{1-\rho^2}} \times \left(1+\frac{q_1^2+q_2^2-2\rho^2 q_1 q_2}{\varphi(1-\rho^2)}\right)^{-\frac{\varphi+1}{2}} \quad \text{Eq. (S3)}$$

where $\boldsymbol{\varphi}$ and $\boldsymbol{\rho}$ are the parameters of the copula, $\boldsymbol{dt(\cdot,\varphi)}$ is the probability density for the standard univariate Student-t distribution with $\boldsymbol{\varphi}$ degrees of freedom.

The h-function of the Student-t copula is formulated as

$$\boldsymbol{h(F_1(q_1),F_2(q_2),\rho,\varphi)} = t_{\varphi+1}\left(\frac{t_\varphi^{-1}(F_1(q_1))-\rho t_\varphi^{-1}(F_2(q_2))}{\sqrt{\frac{\left(\left(\varphi+t_\varphi^{-1}(F_2(q_2))\right)^2\right)(1-\rho^2)}{\varphi+1}}}\right) \quad \text{Eq. (S4)}$$

S1.3 Frank copula

The density of the Frank copula is given by

$$\boldsymbol{c\big(F_1(q_1),F_2(q_2)\big)} = \frac{\varphi(1-e^{-\varphi})e^{-\varphi(F_1(q_1)+F_2(q_2))}}{((1-e^{-\varphi})-(1-e^{-\varphi F_1(q_1)})(1-e^{-\varphi F_2(q_2)}))^2} \quad \text{Eq. (S5)}$$

where $\boldsymbol{\varphi}$ is the parameter of the copula

The h-function of the Frank copula is expressed as

$$\boldsymbol{h(F_1(q_1), F_2(q_2), \varphi) = \frac{\exp(-\varphi F_2(q_2))(\exp(-\varphi F_1(q_1))-1)}{(\exp(-\varphi)-1)+(\exp(-\varphi F_2(q_2))-1)(\exp(-\varphi F_1(q_1))-1)}} \quad \text{Eq. (S6)}$$

S1.4 Clayton copula

The density of the Clayton copula is given by

$$c\big(F_1(q_1), F_2(q_2)\big) = (1+\varphi)\big(F_1(q_1)\cdot F_2(q_2)\big)^{-1-\varphi}(F_1(q_1)^{-\varphi} + F_2(q_2)^{-\varphi} - 1)^{-1/\varphi-2}$$

Eq. (S7)

where $\boldsymbol{\varphi}$ is the parameter of the copula

The h-function of the Clayton copula is expressed as

$$h(F_1(q_1), F_2(q_2), \varphi) = F_2(q_2)^{-\varphi-1}(F_1(q_1)^{-\varphi} + F_2(q_2)^{-\varphi} - 1)^{-1-1/\varphi} \quad \text{Eq. (S8)}$$

S1.5 Gumbel copula

The density of the Gumbel copula is given by

$$\boldsymbol{c\big(F_1(q_1), F_2(q_2)\big) = C(F_1(q_1), F_2(q_2))\big(F_1(q_1)\cdot F_2(q_2)\big)^{-1}} \times$$

$$\boldsymbol{((-logF_1(q_1))^{\varphi} + (-logF_2(q_2))^{\varphi})^{-2+\frac{2}{\varphi}} \times (logF_1(q_1)logF_2(q_2))^{\varphi-1} \times (1 + (\varphi -}$$

$\mathbf{1})((-\boldsymbol{logF_1(q_1)})^{\varphi} + (-\boldsymbol{logF_2(q_2)})^{\varphi})^{\frac{1}{\varphi}})$

Eq. (S9)

where $C\big(F_1(q_1), F_2(q_2)\big) = \exp(-((-logF_1(q_1))^{\varphi} + (-logF_2(q_2))^{\varphi})^{\frac{1}{\varphi}})$

The h-function of the Gumbel copula is given by

$$h(F_1(q_1), F_2(q_2), \varphi) = C(F_1(q_1), F_2(q_2)) \cdot \frac{1}{F_2(q_2)} \cdot (-logF_2(q_2))^{\varphi-1} \times$$

$$((-logF_1(q_1))^{\varphi} + (-logF_2(q_2))^{\varphi})^{\frac{1}{\varphi}-1} \qquad \text{Eq. (S10)}$$

Text S2. Description of $\mathcal{M}_{\text{Kraus}}$

$\mathcal{M}_{\text{Kraus}}$ developed by Kraus and Czado (2017) are used to model the joint distribution of $q_1, \dots, q_k$ and calculate the conditional quantile function of q_k, given $q_1, \dots, q_{k-1}$ for $\phi \in (0, 1)$ as the inverse of the conditional distribution function:

$$q_k(\phi | q_1, \dots, q_{k-1}) \ := F_k^{-1}(C_{k|1,\dots,k-1}^{-1}\big(\phi | F_1(q_1), \dots, F_{k-1}(q_{k-1})\big)) \qquad \text{Eq. (S11)}$$

To easily estimate the conditional quantile function (i.e., $C_{k|1,\dots,k-1}^{-1}$), a D-vine copula is fitted to $(q_k, q_1, \dots, q_{k-1})$, where q_k is fixed as the first node in the first tree. To reduce the dimension of the covariates, a sequential vine construction is modeled by adding covariates while maximizing the conditional log-likelihood (cll):

$$cll\left(\widehat{\boldsymbol{\mathcal{F}}},\widehat{\boldsymbol{\theta}}\right) := \sum_{i=1}^{k} ln c_{F_i(q_i)|F_v(v)}(\widehat{\boldsymbol{\mathcal{F}}},\widehat{\boldsymbol{\theta}}) \qquad \text{Eq. (S12)}$$

where $\widehat{\boldsymbol{\mathcal{F}}}$ is the estimated pair-copula families and $\widehat{\boldsymbol{\theta}}$ is corresponding copula parameters given data.

The cll-based selection procedure provides an automatic forward covariate selection, leading to parsimonious models. Also, two penalized conditional likelihood functions (the AIC- and BIC-conditional log-likelihood) can also be considered to select the effective covariates in $\mathcal{M}_{\text{Kraus}}$.

Text S3. Upper and lower tail dependence

The dependence of streamflow between two sites is measured by common correlation coefficients such as Pearson, Spearman or Kendall. However, these coefficients focus on the dependence in the body of distribution (Bevacqua et al., 2017). Even though two streamflows are uncorrelated according to such common correlation coefficients, there can be a significant dependent in the tails of the distribution (i.e., a tail dependence) (Hobaek Haff et al., 2015).

Mathematically, given two streamflows q_1 and q_2, they are upper tail dependent if the following limit exists and is non-zero:

$$\lambda_{upper}(q_1, q_2) = \lim_{p \to 1} P\{q_1 > F_{q_1}^{-1}(p) \big| q_2 > F_{q_2}^{-1}(p)\} \quad \text{Eq. (S13)}$$

Similarly, the two streamflows are lower tail dependent if

$$\lambda_{lower}(q_1, q_2) = \lim_{p \to 0} P\{q_1 \leq F_{q_1}^{-1}(p) \big| q_2 \leq F_{q_2}^{-1}(p)\} \quad \text{Eq. (S14)}$$

exists and is non-zero.

Reference

Bevacqua, E., Maraun, D., Hobæk Haff, I., Widmann, M., & Vrac, M. (2017). Multivariate statistical modelling of compound events via pair-copula constructions: analysis of floods in Ravenna (Italy). *Hydrology and Earth System Sciences*, *21*(6), 2701–2723.

Hobaek Haff, I., Frigessi, A., & Maraun, D. (2015). How well do regional climate models simulate the spatial dependence of precipitation? An application of pair-copula constructions. *Journal of Geophysical Research: Atmospheres*, *120*(7), 2624–2646.

Kraus, D., & Czado, C. (2017). D-vine copula based quantile regression. *Computational Statistics & Data Analysis*, *110*, 1–18.