


Article

Vine-Copula-Based Quantile Regression for Cascade Reservoirs Management

Wafaa El Hannoun ^{1,*} , Salah-Eddine El Adlouni ² and Abdelhak Zoglat ¹

¹ Laboratory of Mathematics, Statistics and Applications, Faculty of Sciences, Mohammed V University in Rabat, Rabat 1014, Morocco; a.zoglat@um5r.ac.ma

² Department of Mathematics and Statistics, Université de Moncton, Moncton, NB E1A 3E9, Canada; salah-eddine.el.adlouni@umoncton.ca

* Correspondence: wafaa.elhannoun@um5r.ac.ma

Abstract: This paper features an application of Regular Vine (R-vine) copulas, a recently developed statistical tool to assess composite risk. Copula-based dependence modelling is a popular tool in conditional risk assessment, but is usually applied to pairs of variables. By contrast, Vine copulas provide greater flexibility and permit the modelling of complex dependency patterns using a wide variety of bivariate copulas which may be arranged and analysed in a tree structure to explore multiple dependencies. This study emphasises the use of R-vine copulas in an analysis of the co-dependencies of five reservoirs in the cascade of the Saint-John River basin in Eastern Canada. The developed R-vine copulas lead to the joint and conditional return periods of maximum volumes, for hydrologic design and cascade reservoir management in the basin. The main attraction of this approach to risk modelling is the flexibility in the choice of distributions used to model heavy-tailed marginals and co-dependencies.

Keywords: cascade reservoirs; quantile regression; Vine-Copulas; watershed management



Citation: El Hannoun, W.; El Adlouni, S.-E.; Zoglat, A. Vine-Copula-Based Quantile Regression for Cascade Reservoirs Management. *Water* **2021**, *13*, 964. <https://doi.org/10.3390/w13070964>

Academic Editor: Krzysztof Kochanek

Received: 16 February 2021
Accepted: 29 March 2021
Published: 31 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A successful reservoir system management requires pertinent information to make appropriate decisions about the release or diversion of water at a particular time t or over different periods of the year. Such choices are crucial and enormously complex because of the multiple uses of reservoirs such as flood control, hydroelectric production, irrigation, drinking and industrial water supply, navigation and water sports. For instance, fixing water upper extremes is very advantageous in flooding problems management, and coping with the destruction of water structures. In the same way, the utility of lower extremes management is establishing guidelines for drought management, water shortage management, and other disasters on flora and fauna. In practice, this type of management is dictated by well-defined rules and operation policies, often in the form of curves stipulating which action should be taken, depending on the current status of the system. The decision calls for optimization techniques examining all possible alternatives [1]. There is a multitude of methods that can be used for flood control management, for example implicit stochastic optimization (ISO) [2], explicit stochastic optimization (ESO) and parameterization-simulation-optimization (PSO) [3]. Each of these methods has its advantages and limitations, as discussed in [4]. More recently, the multiple reservoir management method, called the Aggregation-decomposition (AGDP) method, has been introduced to solve dimensionality problems in large-scale reservoir systems. It has been the subject of several studies, in particular those of [4–6] where different operating rules are developed.

In this framework, we discuss the best hydrological modelling approach for reproducing observed streamflow and handling extremes, low and high quantiles. Our method also deals with information transfer from one station to another, which improves the knowledge of geographical mechanisms behind an event. The derivation of the contribution of each

upstream station in the downstream station quantiles can help predict runoff in ungauged stations and determine the responsible upstream stations of extreme downstream events.

To this aim, we introduce a new approach based on R-vine copulas. As their name suggests, they are based on the combination of theories of R-vines and copulas. The latter have been extensively used in hydrology. For instance, for the estimation of extreme values, Ref. [7] considered flooding as a joint behavior of peak and volume. Asymmetric copulas were used in [8] to model the structure of spatial interdependence between precipitation rate and its occurrence. Ref. [9] used copulas to model the correlation structure of runoff time series pairs and transfer properties from one basin to another. Ref. [10] used copula-based multivariate analysis with a joint return period to define particular T-year rainfall patterns. However, the multivariate case is not easy to handle with copulas. R-vine copulas propose the division of the multivariate structure to pair copulas, which are easier to manipulate. This approach has many remarkable properties, as detailed in [11,12]. Two particular types of R-vines (see Figure 1) are the canonical (C-) vine which has a star structure and the drawable (D-) vine, which has a path structure. The order of the nodes of the first tree entirely determines the construction of these two R-vine types, making them easy to use. R-vines are increasingly being used, principally in finance and climatology. For example, Ref. [13] use vine-copulas to analyse the interdependencies between bank and insurers in the financial market. Ref. [14] use C-vines to analyse stock exchange. Ref. [15] demonstrate the superiority of vine-copulas over conventional copulas when modelling the dependence structure of a credit portfolio. Ref. [16] used vine-copulas in trading strategies and asset allocation. Ref. [17] used R-vine models to investigate the existence of different global dependence regimes. In climatology, Ref. [18] considered a copula approach to model the dependence of multivariate factors of rainfall. Ref. [19] proved the existence of a relationship between the vine-copula parameters and the station distances in a daily mean temperature study. Ref. [20] used vine copulas to incorporate all relevant dependencies between storm variables. Ref. [21] develop a method using vine copulas to estimate prediction uncertainty in an environmental framework. Ref. [22] obtained a stochastic simulation of stream-flow scenarios using vine copulas. Ref. [23] use vine copulas to model the multivariate dependence among the glacier discharge and other related meteorological variables. More recently, Ref. [24] used vine-copulas to forecast the release of a hydro-power reservoir conditioned by its initial volume. Vine-copulas have also been used in diverse other fields, for example in the analysis of aviation safety [25] and to capture spatial and temporal dependence between hydroelectric power plants [26]. We can find in [27] a detailed survey of vine-copulas applications.

Our approach, called “the reduced-network vine” combines all the information available on all sections of a watershed. We prove throughout this article that, the notion of a “basin”, can be captured through high quantile modelling. This approach will help to know the exact quantity of the quantiles of the volume of water. It will help quantify the minimum and maximum volumes expected in a dam and how to manage water requirements or face expected floods. It will also help decide how much excess water is to transfer and how much water should be released.

We capitalize on the argumentation of [28] for regular vines and the one of [29] for D-vines about the performance of quantile regression to other regression models. In both cases, vine models make less restrictive assumptions on the choice of the covariates influencing the response and can solve analytically for conditional quantiles. Unlike the classical approach (Kendall’s Tau (τ) R-vine), our method finds the quantile regression for the station downstream due to the new constraint of having the station downstream estimated. Moreover, using the reduced-network R-vine copes with the imposed order condition in a D-vine, making the restrictive model assumptions easier to obtain and allowing the formulation of the conditional form.

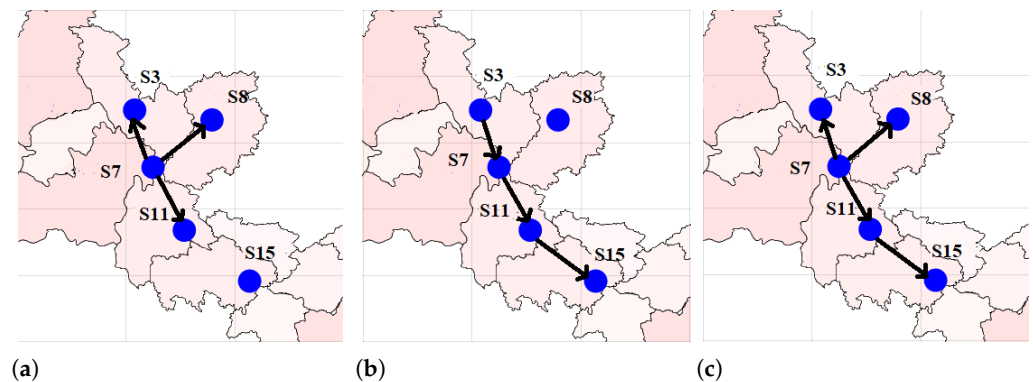


Figure 1. Three different vine structures obtained from the Saint-John basin structure. The blue circles represent the vine nodes which are the analyzed control stations listed in Table 1. The thick black arrows represent the vine edges. The black lines represent the sub-catchment delimitations. (a) A C-vine with 4 nodes. (b) A D-vine with 4 nodes. (c) An R-vine with 5 nodes.

Table 1. Control stations selected to illustrate the Vine-copula approach.

Station	Code	Name	River	Drainage Area (km ²)
S3	FX01AF002	Grand Falls	Saint John	21,900
S7	FX01AG003	Tinker	Aroostook	6060
S8	FX01AH002	Riley Brook	Tobique	2230
S11	FX01AJ010	Coldstream	Becaguimec stream	350
S15	FX01AK010	Mactaquac Generating Station	Saint John	39,900

Our contribution aims to estimate the quantiles of hydrological variables in stations located downstream of a basin. We show how the quantile regression obtained from an R-vine copula model based on the basin's structure allows us to find the desired conditional quantiles. Besides, our method makes it possible to find the station(s) responsible for extreme hydrological events downstream, and therefore, leads to a better-cascade dam management. We illustrate our approach by analysing the co-dependencies of five reservoirs in the cascade of the Saint-John River basin in Eastern Canada. The developed R-vine copulas lead to the joint and conditional return periods of water's maximum volume and the hydrologic design and cascade reservoir management in the basin.

In Section 2, we recall the main definitions of copulas and R-vine-copulas and how to obtain conditional copula expressions. We also detail how to estimate the conditional quantiles. In Section 3, we present our data and the results of the new R-vine quantile regression method for the studied variables with a discussion. In Sections 4 and 5, we discuss our main contributions and some perspectives for further studies. Supplementary material to our work is in Appendices A and B.

2. Materials and Methods

2.1. Copulas

A bivariate copula C is a distribution $C : [0, 1]^2 \rightarrow [0, 1]$ with uniform margins. The theorem of Sklar [30] shows that for a bivariate distribution F with continuous margins F_1 and F_2 , there exists a unique copula $C(\cdot, \cdot)$, such that:

$$F(x_1, x_2) = C_{12}(F_1(x_1), F_2(x_2)), \forall (x_1, x_2) \in \mathbb{R}^2 \quad (1)$$

The following functions, related to copulas, can be found in [31,32]:

$$\begin{aligned}
 \text{Density} & c_{12}(u_1, u_2) = \frac{\partial^2 C_{12}(u_1, u_2)}{\partial u_1 \partial u_2} \\
 \text{Bivariate density} & f(x_1, x_2) = c_{12}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1) \cdot f_2(x_2) \\
 \text{Conditional density} & f(x_2/x_1) = c_{12}(F_1(x_1), F_2(x_2)) \cdot f_2(x_2) \\
 \text{Conditional density of the copula} & h(u/v) = \frac{\partial C(u, v)}{\partial v}
 \end{aligned} \tag{2}$$

2.2. R-Vines

We also recall the notion of a vine as introduced in [11].

Definition 1. *V is a vine on n elements if*

1. $V = (T_1, \dots, T_m)$.
2. T_1 is a tree with nodes $N_1 = \{1, \dots, n\}$ and a set of edges denoted E_1 .
3. For $i = 2, \dots, m$, T_i is a tree with nodes $N_i \subset N_1 \cup E_1 \cup E_2 \cup \dots \cup E_{i-1}$ and edge set E_i .

As described in [11], V is a regular vine when $m = n$, T_i is a connected tree with edge set E_i and node set $N_i = E_{i-1}$, with cardinality $\#N_i = n - (i - 1)$ for $i = 1, \dots, n$. In addition, the proximity condition must hold, i.e., for $i = 2, \dots, n - 1$, if $a = \{a_1, a_2\}$ and $b = \{b_1, b_2\}$ are two nodes in N_i connected by an edge then $\#a \cap b = 1$.

2.3. R-Vine Copulas

We define the specification of the R-vine copula corresponding to an R-vine as in [33].

Definition 2. Let $F = (F_1, \dots, F_n)$ be a vector of continuous invertible distribution functions, V an n -dimensional R-vine and $B = \{B_e / i = 1, \dots, n - 1; e \in E_i\}$ where B_e is a bivariate copula and E_i is the edge set of tree T_i of the R-vine V . The triplet (F, V, B) is called an R-vine copula specification.

Bedford et al. [11,34] showed that the density of an R-vine-copula, specified by the assignment of an appropriate bivariate copula to the edges of the R-vine, is equal to the product of the conditional and unconditional copula assigned to its edges. More precisely, they proved the following theorem.

Theorem 1. Let (F, V, B) be an R-vine specification on n elements. The distribution F is unique and its density is given by

$$f_{1, \dots, n}(x) = \prod_{k=1}^n f_k(x_k) \prod_{i=1}^{n-1} \prod_{e \in E_i} c_{C_{e,a}, C_{e,b}/D_e}(F_{C_{e,a}/D_e}(x_{C_{e,a}}/x_{D_e}), F_{C_{e,b}/D_e}(x_{C_{e,b}}/x_{D_e})), \tag{3}$$

where $x = (x_1, \dots, x_n)$, $e = a, b$, $x_{D_e} = x_i / i \in D_e$ and f_i denotes the density of F_i for $i = 1, \dots, n$.

To obtain conditional copulas, Joe [35] showed that:

$$\begin{aligned}
 F_{C_{e,a}/D_e}(x_{C_{e,a}}/x_{D_e}) &= \frac{\partial C_{C_a/D_a}(F_{C_{a,a_1}/D_a}(x_{C_{a,a_1}}/x_{D_a}), F_{C_{a,a_2}/D_a}(x_{C_{a,a_2}}/x_{D_a}))}{\partial F_{C_{a,a_2}/D_a}(x_{C_{a,a_2}}/x_{D_a})} \\
 &:= h(F_{C_{a,a_1}/D_a}(x_{C_{a,a_1}}/x_{D_a}), F_{C_{a,a_2}/D_a}(x_{C_{a,a_2}}/x_{D_a})),
 \end{aligned} \tag{4}$$

where each term of the partial derivative of the copula must be obtained recursively.

To specify the necessary copulas for a pair-copula construction, we use the graphical structure of R-vines [36]. Adapting an R-vine copula specification to a given data set requires, in general, the following separate tasks, as detailed in [37].

1. Selection of vine R (structure), i.e., selection of unconditioned and conditioned pairs to be used.
2. Choice of a family of bivariate copula for each pair selected in the previous step.
3. Estimation of the corresponding parameter(s) for each copula.

Other methods for selecting regular vine-copulas are investigated in [38].

We have tried to fit several combinations of copulas (Joe, Tawn, and two-parameter mix copulas) to our data. However, these copulas were not appropriate for our data and resulted in poor estimators. We have, therefore, limited our choice to copulas that best fit our data : Gumbel, Clayton and Frank copulas. These copulas are members of the Archimedean family. Because of their flexible structure, Archimedean copulas are the most commonly used to model dependence in hydrology [39]. Gumbel copula has a range between independence and perfect positive dependence, Clayton copula has more mass in the negative tail and Frank copula is a flexible copula which takes no restrictions in its parameter [40]. The distributions of these copulas and their properties can be found in [31]. The choice of bivariate copulas (conditional and unconditional) of each tree is made individually each time. In this work, we use the maximum likelihood method to obtain copula parameter estimators (MLE) [32], and retain the copulas minimizing the Akaike information criterion (AIC).

2.4. Quantile Regression

As demonstrated in [41], the R-vine-copula-based quantile regression showed its superiority to linear regression in terms of making inferences. The goal is to predict the quantile of a response variable Y given the outcome of certain predictor variables X_1, \dots, X_d , ($d \geq 1$), where $Y \sim F_Y$ and $X_j \sim F_j, j = 1, \dots, d$ [29]. Hence, all the effort is made on the joint modelling of Y and X , and more particularly on the quantile function for α in $[0, 1]$ [29,42]:

$$q_\alpha(x_1, \dots, x_d) := F_{Y/X_1, \dots, X_d}^{-1}(\alpha/x_1, \dots, x_d), \quad (5)$$

using probability integral transforms [43,44]. More precisely, let $V := F_Y(Y)$ and $U_j := F_j(X_j)$ we have that:

$$\begin{aligned} F_{Y/X_1, \dots, X_d}(y/x_1, \dots, x_d) &= P(Y \leq y/X_1 = x_1, \dots, X_d = x_d) \\ &= P(V \leq v/U_1 = u_1, \dots, U_d = u_d) \\ &= C_{V/U_1, \dots, U_d}(v/u_1, \dots, u_d). \end{aligned} \quad (6)$$

Applying the inverse function F_Y^{-1} , we have that

$$F_{Y/X_1, \dots, X_d}^{-1}(\alpha/x_1, \dots, x_d) = F_Y^{-1}(C_{V/U_1, \dots, U_d}^{-1}(\alpha/u_1, \dots, u_d)). \quad (7)$$

An estimate of the conditional quantile can, therefore, be obtained by estimating the marginal distributions $F_Y, F_j (j = 1, \dots, d)$ as well as the copula $C_{V/U_1, \dots, U_d}$:

$$\hat{q}_\alpha(x_1, \dots, x_d) := \hat{F}_Y^{-1}(\hat{C}_{V/U_1, \dots, U_d}^{-1}(\alpha/\hat{u}_1, \dots, \hat{u}_d)). \quad (8)$$

More details about R-vine-copula-based quantile regression are to be found in [29].

3. Results

3.1. Data

The Saint-John River watershed is an inter-provincial watershed connecting Quebec and New Brunswick. It is also an international watershed connecting Canada and the United States. It covers an area of approximately 55,200 km², of which just over half is in New Brunswick [45]. The Saint-John River is often known by an increase in water levels above the flood threshold in several collectivities, the most remarkable of which is that of spring (April-May) of the year 2008 [45].

Reservoirs in New Brunswick generally serve several purposes, of which flood control, both at tributary and mainstream level. Therefore, the joint operation rule should be able to coordinate not only the individual reservoirs but also the different uses of the water. From this perspective, the R-vine-copula-based quantile regression is considered suitable for a large basin such as that of the Saint-John River (Figure 2).

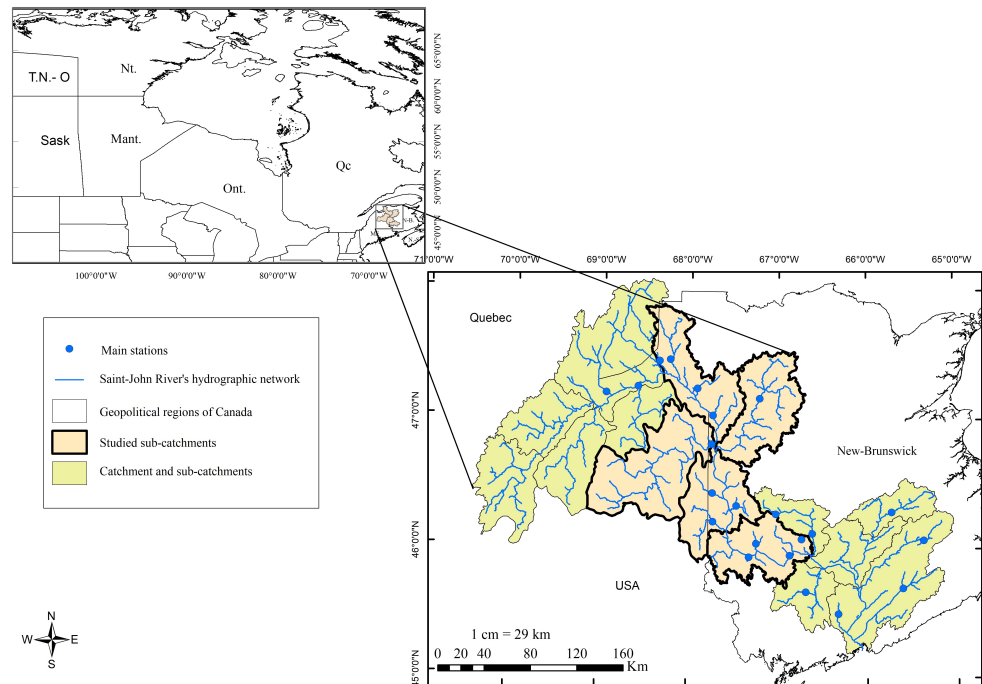


Figure 2. The Saint-John basin and the studied sub-basins.

To cope with flooding episodes in the Saint-Jean basin, the implementation of efficient management rules is essential for the decision-maker. It is within this framework that we try to derive management rules for a reservoir system by limiting flood losses at the level of five selected flood control sections, based on our database. In Figure 3, the structure of reservoirs determined by these control stations chosen to cover the entire basin can be seen. We also give the station codes and names in Table 1.

Among the characteristics of the hydrographs identified for all the stations and all the observation periods, we considered the peak flow and the flood volume [46–48]. We choose these two indices to illustrate the use of the R-vine-copula-based quantile regression approach on five sections of flood control. We then have a matrix of 10 components for each of the 42 years of the observation period (from 1975 to 2016). For the reader's convenience, the time series of the volume variables are represented in Appendix A.

We used the Anderson–Darling test as a criterion to choose the most fitted distribution to our data [49]. Gamma distribution (Pearson type III) usually represents the maximum annual flows, the volume of annual runoff. Weibull distribution is a particular case of the Generalized extreme value distribution (GEV) recommended for representing maxima [50]. These distributions have fat tails and are the most convenient to represent the extremes. Gumbel distribution is another special case of the GEV distribution when its Kappa shape parameter is zero. We represent the fitted distributions and the results of the Box-Ljung independence test in Table 2.

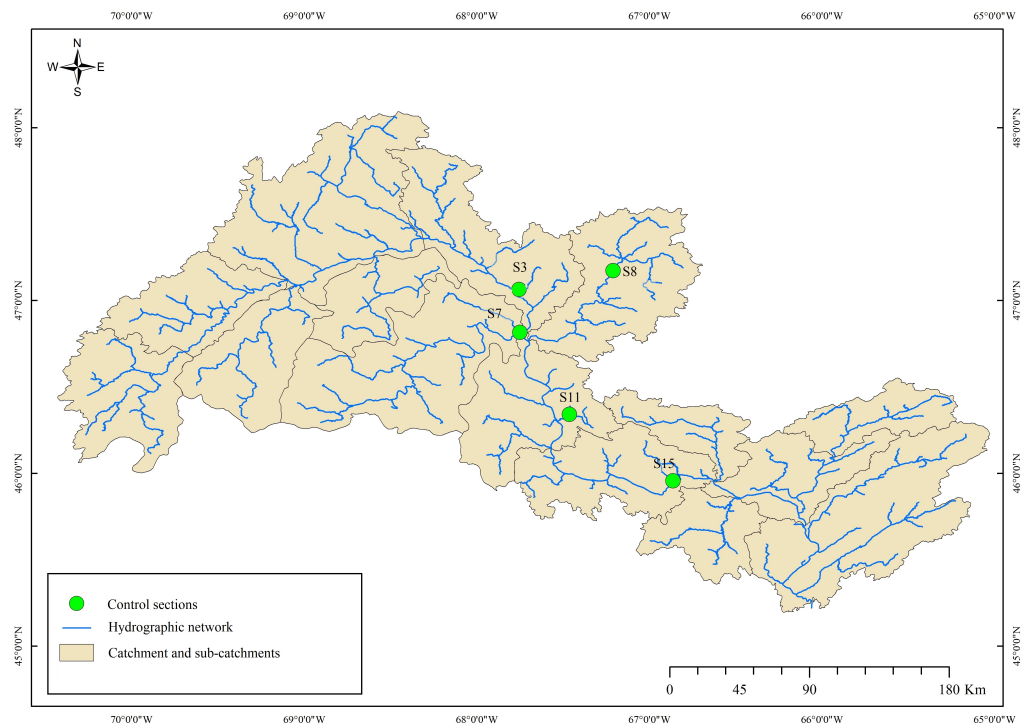


Figure 3. Location of the control sections used to illustrate the Vine-copula approach.

Table 2. Marginal distributions and results of the independence test.

Variable	Station	Distribution	Parameter 1	Parameter 2	Box-Ljung Test
Volume	3	Gumbel	0.2054731	0.8775550	0.5819
	7	Weibull	3.956469	1.100047	0.2502
	8	Gamma	12.26178	12.26174	0.6783
	11	Weibull	3.498557	1.111809	0.864
	15	Gamma	18.47347	18.47331	0.778
Peak flow	3	Gumbel	0.2391086	0.8574496	0.6619
	7	Gumbel	0.2508672	0.8574092	0.2957
	8	Gumbel	0.2432496	0.8625942	0.3618
	11	Gumbel	0.2801832	0.8363660	0.365
	15	Gumbel	0.2201248	0.8738290	0.2701

3.2. The New R-Vine

The (unique) reduced-network R-vine corresponding to the Saint-John basin is given in Figure 4. In Table 3, the fitted copulas corresponding to each edge of the R-vine with their parameters are provided. The associated fitted normalized contour plots of the selected copulas are presented in Figure 5 to visually investigate their influence.

To compare the proposed R-vine structure with a classical model, we construct the vine corresponding to our data using the maximum spanning tree method. The obtained Kendall’s τ values summarized in Figure 6 suggest a D-vine structure represented in Figure 7. We compare the efficiency of our model by illustrating the basin dependency structure with that of the D-vine and an all-Gauss D-vine as a reference structure. The results of Table 4 show that the Tau D-Vine model has the highest MLE, followed by the new R-Vine then all Gauss D-Vine. For the results of the AIC and the Bayesian information criterion (BIC), we have the same classification. Moreover, based on Vuong’s tests results [51], we cannot reject the hypothesis that the new R-vine and the Tau D-Vine are equivalent. For the peak flow variable, the maximum spanning tree method suggests an R-vine. The results of the efficiency comparison and Vuong’s tests in Table 4 lead to the same conclusion as for the volume variable. But even with the results preferring the

classically obtained vine-copulas to the new R-vine, we keep this last to do the regression because it gives consistent results with the geographical position of the stations.

Table 3. Bivariate copulas and conditional bivariate copulas of the reduced-network R-vine. Sur Gumbel and I stand for Survival Gumbel and Independence copulas, respectively.

Tree	Edge	Copula	Volume Parameter	K-Tau	Copula	Peak Flow Parameter	K-Tau
1	4,5	Sur Gumbel	1.45	0.31	Sur Gumbel	1.42	0.30
	2,4	Sur Gumbel	1.65	0.39	Sur Gumbel	1.58	0.37
	2,3	Frank	5.38	0.48	Gumbel	1.50	0.34
	1,2	Frank	11.21	0.70	Gumbel	2.12	0.53
2	2,5;4	Gumbel	2.13	0.53	Gumbel	3.34	0.70
	3,4;2	I	-	0.00	I	-	0.00
	1,3;2	I	-	0.00	I	-	0.00
3	3,5;2,4	Clayton	0.67	0.25	I	-	0.00
	1,4;3,2	I	-	0.00	I	-	0.00
4	1,5;3,2,4	Clayton	1.18	0.37	Gumbel	1.53	0.35

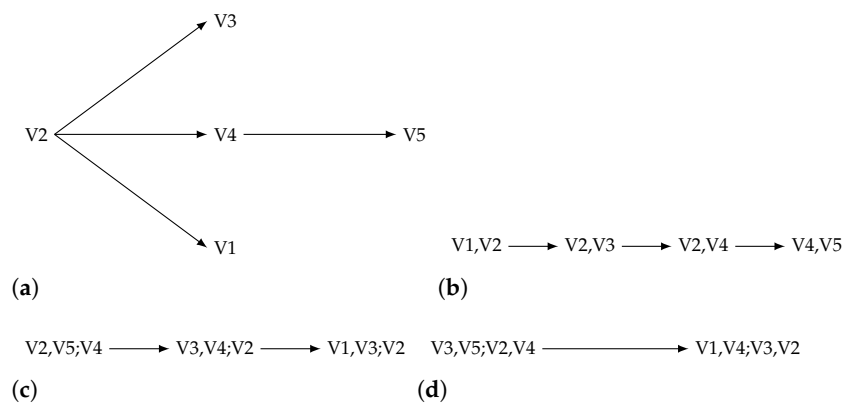


Figure 4. The reduced-network R-vine. V1, V2, V3, V4 and V5 stand for stations 3, 7, 8, 11 and 15, respectively. (a) Tree 1. (b) Tree 2. (c) Tree 3. (d) Tree 4.

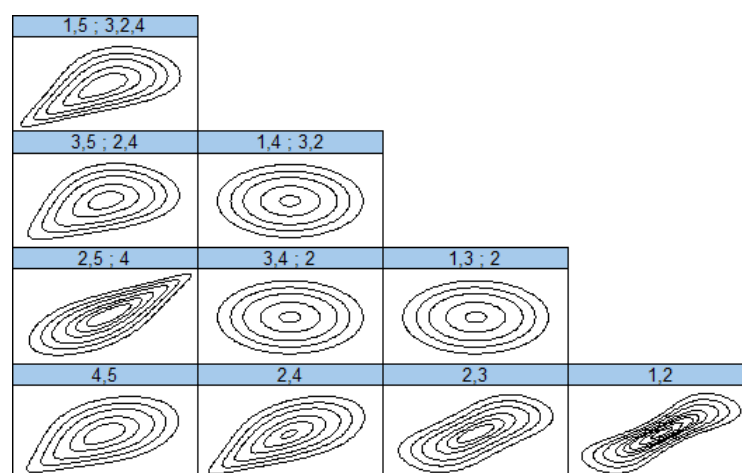


Figure 5. Fitted R-vine normalized contour plots of the specified pair copulas for the volume variable.

more distant from station 15) or station 8 (up-tributary and the farthest from station 15). On the other hand, the scenarios for station 11 have not been represented because they have no noticeable effect on the extreme quantiles of station 15 (even if it is down-mainstream and closest to station 15).

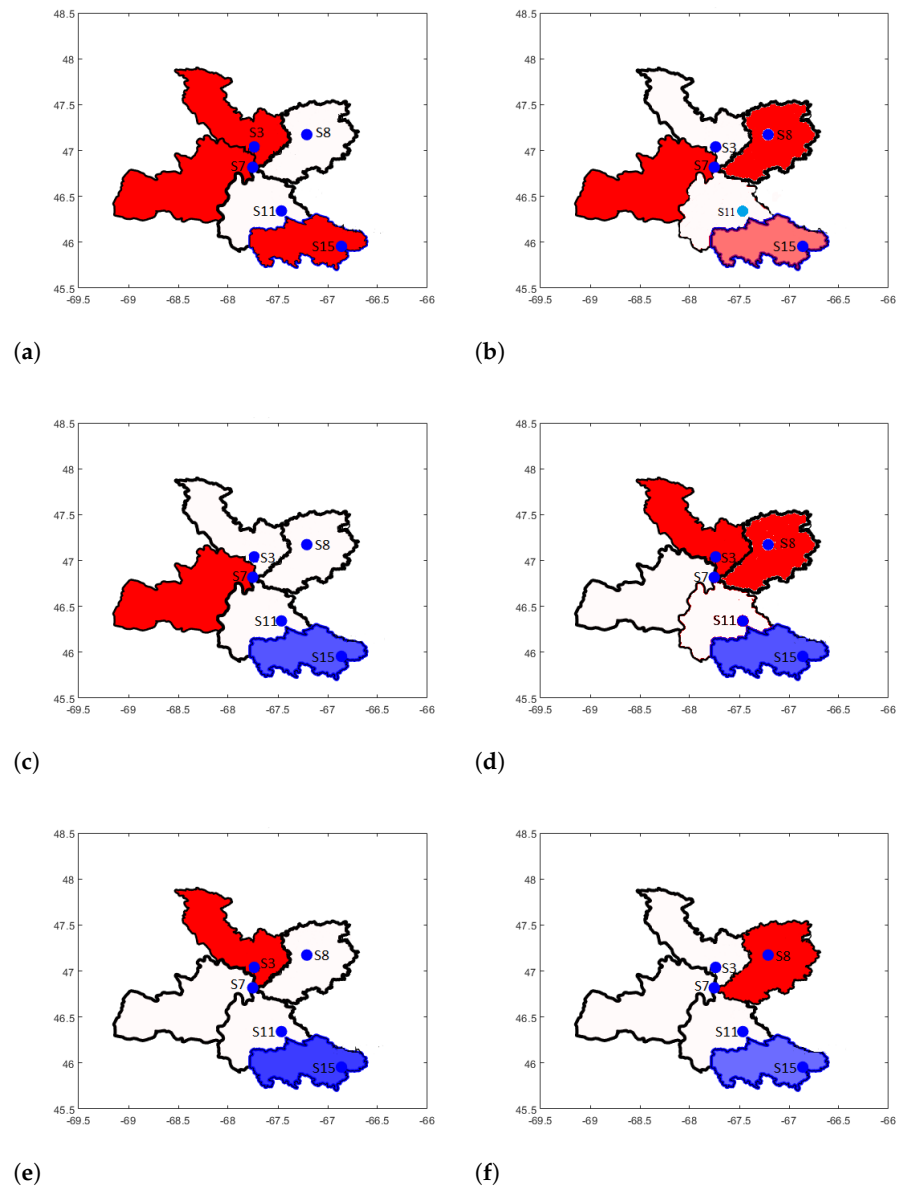


Figure 8. Illustration of the obtained scenarios with respect to downstream conditional quantiles. The red color means values are set to Q_{100} as in Table 5. White means values are set to Q_2 . For station 15, blue means the conditional quantile S15 Q_{100} is less than the marginal index, it is pink when they are almost equal and red when S15 Q_{100} is much greater than the marginal index. (a–f): scenarios 5, 8, 2, 6, 1, and 3 respectively.

We resume the obtained quantiles of the volume in station 15 in each of the above scenarios in Table 5. We also give a comparison with the observed and marginal quantiles using indexes.

Table 5. Classification of possible scenarios according to the severity of the extreme quantile of station 15. M. index: Marginal index. O. index : Observed index.

Scenarios	S 3	S 7	S 8	S 11	S15 Q_{100} (1000 m ³)	M. Index	O. Index
Scenario 5	Q_{100}	Q_{100}	Q_2	Q_2	300.91	1.77	1.81
Scenario 8	Q_2	Q_{100}	Q_{100}	Q_2	166.32	0.98	1.00
Scenario 7	Q_{100}	Q_2	Q_2	Q_{100}	153.73	0.90	0.92
Scenario 4	Q_2	Q_2	Q_2	Q_{100}	148.81	0.87	0.90
Scenario 2	Q_2	Q_{100}	Q_2	Q_2	145.54	0.86	0.88
Scenario 6	Q_{100}	Q_2	Q_{100}	Q_2	145.22	0.85	0.87
Scenario 1	Q_{100}	Q_2	Q_2	Q_2	141.78	0.83	0.85
Scenario 3	Q_2	Q_2	Q_{100}	Q_2	139.56	0.82	0.84

To complete the results, we set the values of the first four stations with the quantiles of return periods 2, 20 and 100 years. We then calculate the conditional quantiles of station 15 based on these values for different probabilities. Figure 9 represents the conditional distributions of the extreme events at station 15 as a function of potential scenarios in the upstream sub-basins. It is then possible to deduce the conditional distributions of the quantiles, at station S15, depending on the levels recorded in the other four stations. This is an estimate based on the multivariate model and, therefore, takes into account the uncertainties and interactions between the series of extremes observed at each of the measuring stations.

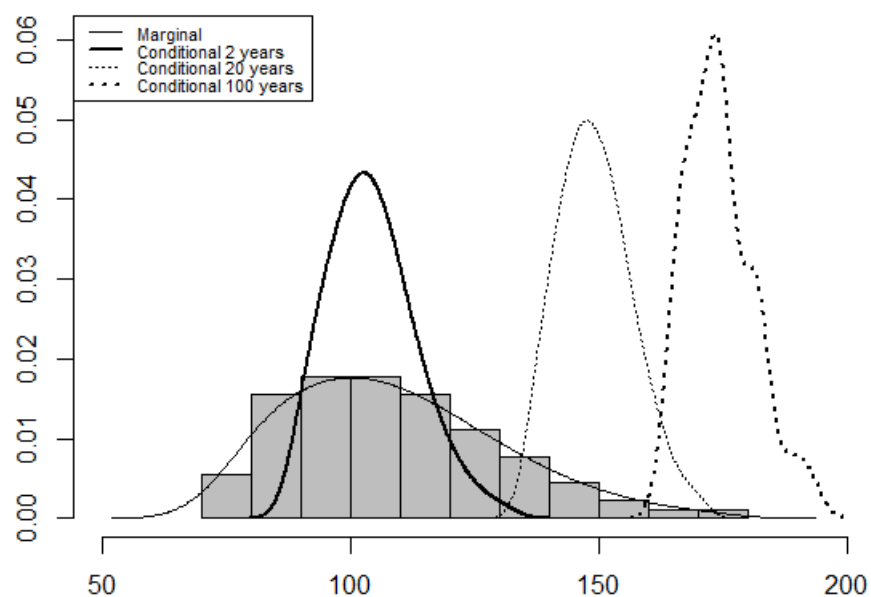


Figure 9. Conditional quantiles of the volume (1000 m³) in station 15 according to the variations in probabilities with fixed covariates.

4. Discussion

This study established multivariate models for flood-risk assessment of cascade reservoirs. The management approaches for multi-reservoir flood control mainly concern minimizing the maximum water level in the dam itself or discharges at downstream flood control points. For a large watershed, such as the Saint-John River, this management is even more challenging. It should take into account the interactions between the sub-basins. Moreover, the time lag and the flooding interdependence along the river are two significant parameters. They should be taken into consideration during the management process. The proposed multivariate vine copula model offers some flexibility to combine all risk sources with different dependence structures. The flooding capacity of existing reservoirs

cannot hold back flooding downstream when extreme flooding has occurred in the upper Saint-John River basin. Therefore, more reservoirs are needed to control flooding in some downstream areas. The sub-basins that require more control are those connected to stations S3 and S7. This result might suggest the necessity to control all the stations on the main river and chief tributaries. The model also makes it possible to estimate the contribution of each of the sub-basins to extreme floods.

The Saint-John River has seasonal regulation reservoirs for medium and low floods. In the case of extreme upstream floods, the valves of downstream reservoirs are opened prematurely, causing high levels downstream in early spring. By better estimating the conditional risk of floods downstream of the basin, managers can make informed decisions to reduce vulnerability to flooding.

One of the extensions to our study would be the use of Extreme value copulas such as Joe, Tawn or 2-parameter Archimedean copulas (BB class). Due to the known ability of this class of copulas to model tail behaviours, this extension would be adapted to other hydroclimatic regions. We also think that an extension of this analysis could involve all the basin's sub-catchments. In this case, each sub-catchment network could be modelled separately by a vine structure, then all the vine models obtained can be regrouped to one multivariate model.

5. Conclusions

In this article, we have introduced a new approach to combine all the information available on the different sections of a river basin. Flooding in a large basin, as in the case of the Saint-John River, is a highly complex phenomenon and depends on several temporal and spatial parameters. The R-vine-copula-based quantile regression approach enables dimensionality to be reduced while retaining essential information on the contribution of each section to the entire basin's flood. A similar study can be done based on climate forecasts to produce decision-making tools on the management of available reservoirs and help to mitigate the effects of floods and protect vulnerable areas to flooding. The three main achievements of the work can be summarized as follows: (1) Conception of the multivariate distribution for multi-reservoirs volumes' extremes; (2) Development of the conditional quantile regression expressions at the downstream station; (3) Simulation of scenarios for cascade reservoirs, considering both sub-basins local characteristics and interactions.

The use of an R-vine structure suggested by the geographic system of the watershed allows precise modelling of the dependency structure between stations for a given variable. The proposed model can reproduce the periodic behaviour of the volume and peak flow series. The choice of the reduced-network R-vine allows for the alleviation of the conditions imposed during the construction of other vine types (calculation of the correlation coefficients and respect of the covariates order for the D-vine).

The generated scenarios show that the extreme conditions in the two stations S3 and S7 are decisive for the downstream sub-basin. These indicate the need for an improvement in the capacity and management of flood volumes at the level of these two sub-basins. Indeed, station S15 is located near Fredericton city, the capital of New Brunswick province, which experiences frequent flooding.

The predicted quantiles are of crucial importance for decision-makers, allowing better management of existing dams as well as the development of new hydraulic structures. Besides the studied case, vine copulas could be very useful for many other hydro-meteorological variables with a spatial and/or temporal component. Such models also allow for information transfer from gauged to ungauged areas.

Author Contributions: Conceptualization, A.Z., S.-E.E.A. and W.E.H.; methodology, A.Z., S.-E.E.A. and W.E.H.; software, W.E.H.; validation, S.-E.E.A. and A.Z.; formal analysis, W.E.H., A.Z. and S.-E.E.A.; investigation, S.-E.E.A., W.E.H. and A.Z.; resources, W.E.H.; writing—original draft preparation, W.E.H.; writing—review and editing, W.E.H., S.-E.E.A. and A.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Natural Sciences and Engineering Research Council of Canada grant number RGPIN-2019-05746.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The observed daily flow data are available from the Historical Data provided by Environment Canada https://climat.meteo.gc.ca/historical_data/search_historic_data_f.html (accessed on 20 June 2020).

Acknowledgments: The authors would like to thank the Associate Editor and the three reviewers for all recommendations and comments that improve the quality of the paper. The authors would also like to thank Anas Boukili Makhoukhi, research assistant at Université de Moncton for providing them with figures related to Saint-John basin (Figures 2 and 3).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

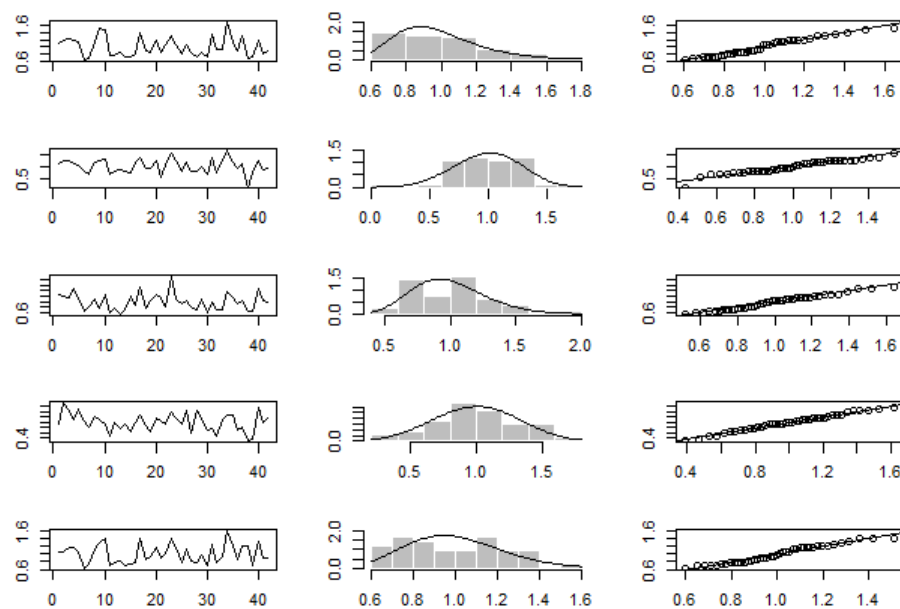


Figure A1. Time series, histograms with densities of theoretical laws and QQ-plots of the volume variable of the five studied stations. From top to bottom: station 3, 7, 8, 11 and 15.

Appendix B

As in [32], the relative distribution density of the 5 nodes R-vine is as follows:

$$f_{12345} = \overbrace{f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdot f_5} \cdot \overbrace{c_{12} \cdot c_{23} \cdot c_{24} \cdot c_{45}} \cdot \overbrace{c_{13/2} \cdot c_{34/2} \cdot c_{25/4}} \cdot \overbrace{c_{14/23} \cdot c_{35/24}} \cdot \overbrace{c_{15/234}}, \quad (A1)$$

where the grouped terms represent, respectively, the densities of the univariate distributions, the densities of the unconditional copulas of tree 1, the densities of the 1-conditional copulas of tree 2, the densities of the 2-conditional copulas of tree 3 and the 3-conditional copula density of tree 4.

The conditional distribution in the downstream station of our study case is given by:

$$F(x_5/x_1, x_2, x_3, x_4) = \frac{\partial C_{15,234}(F(x_5/x_2, x_3, x_4), F(x_1/x_2, x_3, x_4))}{\partial F(x_1/x_2, x_3, x_4)}, \quad (A2)$$

where the expressions of the 3-conditional distributions are:

$$F(x_1/x_2, x_3, x_4) = \frac{\partial C_{14,23}(F(x_1/x_2, x_3), F(x_4/x_2, x_3))}{\partial F(x_4/x_2, x_3)} \tag{A3}$$

$$F(x_5/x_2, x_3, x_4) = \frac{\partial C_{35,24}(F(x_5/x_2, x_4), F(x_3/x_2, x_4))}{\partial F(x_3/x_2, x_4)}. \tag{A4}$$

The expression of the 2-conditional distributions are

$$F(x_1/x_2, x_3) = \frac{\partial C_{13,2}(F(x_1/x_2), F(x_3/x_2))}{\partial F(x_3/x_2)}, \quad F(x_4/x_2, x_3) = \frac{\partial C_{43,2}(F(x_4/x_2), F(x_3/x_2))}{\partial F(x_3/x_2)}, \tag{A5}$$

$$F(x_3/x_2, x_4) = \frac{\partial C_{34,2}(F(x_3/x_2), F(x_4/x_2))}{\partial F(x_4/x_2)}, \quad F(x_5/x_2, x_4) = \frac{\partial C_{52,4}(F(x_5/x_4), F(x_2/x_4))}{\partial F(x_2/x_4)}, \tag{A6}$$

and the expressions of the 1-conditional distributions are:

$$F(x_1/x_2) = \frac{\partial C_{1,2}(F(x_1), F(x_2))}{\partial F(x_2)}, \quad F(x_3/x_2) = \frac{\partial C_{2,3}(F(x_2), F(x_3))}{\partial F(x_2)}, \tag{A7}$$

$$F(x_4/x_2) = \frac{\partial C_{2,4}(F(x_2), F(x_4))}{\partial F(x_2)}, \quad F(x_2/x_4) = \frac{\partial C_{2,4}(F(x_2), F(x_4))}{\partial F(x_4)}, \tag{A8}$$

$$F(x_5/x_4) = \frac{\partial C_{4,5}(F(x_4), F(x_5))}{\partial F(x_4)}, \tag{A9}$$

where the $F(x_i)$ correspond to the univariate distributions. Then, following [29], we can write the conditional distribution of X_5 given (U_1, U_2, U_3, U_4) :

$$\begin{aligned} C_{X_5/U_1, U_2, U_3, U_4}(x_5/u_1, u_2, u_3, u_4) &= h_{X_5/U_1, U_2, U_3, U_4}(C_{X_5/U_2, U_3, U_4}(x_5/u_2, u_3, u_4), C_{X_1/U_2, U_3, U_4}(x_1/u_2, u_3, u_4)) \\ &= h_{X_5/U_1, U_2, U_3, U_4}(h_{X_5/U_2, U_3, U_4}(C_{X_5/U_2, U_4}(x_5/u_2, u_4), C_{U_3/U_2, U_4}(u_3/u_2, u_4)), \\ &\quad h_{U_1/U_2, U_3, U_4}(C_{U_1/U_2, U_3}(u_1/u_2, u_3), C_{U_4/U_2, U_3}(u_4/u_2, u_3))) \\ &= h_{X_5/U_1, U_2, U_3, U_4}(h_{X_5/U_2, U_3, U_4}(h_{X_5/U_2, U_4}(C_{X_5/U_4}(x_5/u_4), C_{X_2/U_4}(x_2/u_4)), \\ &\quad h_{U_3/U_2, U_4}(C_{U_3/U_2}(u_3/u_2), C_{U_4/U_2}(u_4/u_2))), h_{U_1/U_2, U_3, U_4}(h_{U_1/U_2, U_3} \\ &\quad (C_{U_1/U_2}(u_1/u_2), C_{U_3/U_2}(u_3/u_2)), h_{U_4/U_2, U_3}(C_{U_4/U_2}(u_4/u_2), \\ &\quad C_{U_3/U_2}(u_3/u_2)))) \\ &= h_{X_5/U_1, U_2, U_3, U_4}(h_{X_5/U_2, U_3, U_4}(h_{X_5/U_2, U_4}(h_{X_5/U_4}(x_5/u_4), h_{X_2/U_4}(x_2/u_4)), \\ &\quad h_{U_3/U_2, U_4}(h_{U_3/U_2}(u_3/u_2), h_{U_4/U_2}(u_4/u_2))), h_{U_1/U_2, U_3, U_4}(h_{U_1/U_2, U_3} \\ &\quad (h_{U_1/U_2}(u_1/u_2), h_{U_3/U_2}(u_3/u_2)), h_{U_4/U_2, U_3}(h_{U_4/U_2}(u_4/u_2), \\ &\quad h_{U_3/U_2}(u_3/u_2)))). \end{aligned} \tag{A10}$$

By inverting, we get:

$$\begin{aligned} C_{X_5/U_1, U_2, U_3, U_4}^{-1}(\alpha/u_1, u_2, u_3, u_4) &= h_{X_5/U_4}^{-1} \left\{ h_{X_5/U_2, U_4}^{-1} \left(h_{X_5/U_2, U_3, U_4}^{-1} \left(h_{X_5/U_1, U_2, U_3, U_4}^{-1}(\alpha, h_{U_1/U_2, U_3, U_4} \right. \right. \right. \\ &\quad \left. \left. \left(h_{U_1/U_2, U_3} \left(h_{U_1/U_2}(u_1/u_2), h_{U_3/U_2}(u_3/u_2) \right), h_{U_4/U_2, U_3} \left(h_{U_4/U_2} \right. \right. \right. \right. \\ &\quad \left. \left. \left. (u_4/u_2), h_{U_3/U_2}(u_3/u_2) \right) \right) \right), h_{U_3/U_2, U_4} \left(h_{U_3/U_2}(u_3/u_2), h_{U_4/U_2} \right. \right. \\ &\quad \left. \left. (u_4/u_2) \right) \right), h_{U_2/U_4}(u_2/u_4), u_4 \right\}, \end{aligned} \tag{A11}$$

by introducing the inverted estimated marginal as in Equation (8), we obtain an estimate of the desired conditional quantile.

References

1. Oliveira, R.; Loucks, D.P. Operating rules for multireservoir systems. *Water Resour. Res.* **1997**, *33*, 839–852. [CrossRef]
2. Young, G.K. Finding reservoir operating rules. *J. Hydraul. Div.* **1967**, *93*, 297–322. [CrossRef]
3. Celeste, A.B.; Billib, M. Evaluation of stochastic reservoir operation optimization models. *Adv. Water Resour.* **2009**, *32*, 1429–1443. [CrossRef]

4. Zhang, J.; Li, Z.; Wang, X.; Lei, X.; Liu, P.; Feng, M.; Khu, S.T.; Wang, H. A novel method for deriving reservoir operating rules based on flood classification-aggregation-decomposition. *J. Hydrol.* **2019**, *568*, 722–734. [[CrossRef](#)]
5. Li, L.; Liu, P.; Rheinheimer, D.E.; Deng, C.; Zhou, Y. Identifying explicit formulation of operating rules for multi-reservoir systems using genetic programming. *Water Resour. Manag.* **2014**, *28*, 1545–1565. [[CrossRef](#)]
6. Liu, P.; Guo, S.; Xu, X.; Chen, J. Derivation of aggregation-based joint operating rule curves for cascade hydropower reservoirs. *Water Resour. Manag.* **2011**, *25*, 3177–3200. [[CrossRef](#)]
7. De Michele, C.; Salvadori, G. A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. *J. Geophys. Res. Atmos.* **2003**, *108*. [[CrossRef](#)]
8. Bárdossy, A.; Pegram, G. Copula based multisite model for daily precipitation simulation. *Hydrol. Earth Syst. Sci.* **2009**, *13*, 2299–2314. [[CrossRef](#)]
9. Samaniego, L.; Bárdossy, A.; Kumar, R. Streamflow prediction in ungauged catchments using copula-based dissimilarity measures. *Water Resour. Res.* **2010**, *46*. [[CrossRef](#)]
10. De Luca, D.L.; Biondi, D. Bivariate return period for design hyetograph and relationship with t-year design flood peak. *Water* **2017**, *9*, 673. [[CrossRef](#)]
11. Bedford, T.; Cooke, R.M. Vines: A new graphical model for dependent random variables. *Ann. Stat.* **2002**, *30*, 1031–1068. [[CrossRef](#)]
12. Joe, H.; Kurowicka, D. *Dependence Modeling: Vine Copula Handbook*; World Scientific: Vancouver, BC, Canada, 2011.
13. Brechmann, E.C.; Hendrich, K.; Czado, C. Conditional copula simulation for systemic risk stress testing. *Insur. Math. Econ.* **2013**, *53*, 722–732. [[CrossRef](#)]
14. Krouthén, J. Extreme Joint Dependencies with Copulas. A New Approach for the Structure of C-Vines. Master's Thesis, Department of Mathematics, Uppsala University, Uppsala, Sweden, 2015.
15. Geidosch, M.; Fischer, M. Application of vine copulas to credit portfolio risk modeling. *J. Risk Financ. Manag.* **2016**, *9*, 4. [[CrossRef](#)]
16. Bevacqua, E.; Maraun, D.; Hobæk Haff, I.; Widmann, M.; Vrac, M. Multivariate statistical modelling of compound events via pair-copula constructions: Analysis of floods in Ravenna (Italy). *Hydrol. Earth Syst. Sci.* **2017**, *21*, 2701–2723. [[CrossRef](#)]
17. Fink, H.; Klimova, Y.; Czado, C.; Stöber, J. Regime switching vine copula models for global equity and volatility indices. *Econometrics* **2017**, *5*, 3. [[CrossRef](#)]
18. Gyasi-Agyei, Y. Copula-based daily rainfall disaggregation model. *Water Resour. Res.* **2011**, *47*. [[CrossRef](#)]
19. Erhardt, T.M.; Czado, C.; Schepsmeier, U. R-vine models for spatial time series with an application to daily mean temperature. *Biometrics* **2015**, *71*, 323–332. [[CrossRef](#)] [[PubMed](#)]
20. Vernieuwe, H.; Vandenberghe, S.; De Baets, B.; Verhoest, N. A continuous rainfall model based on vine copulas. *Hydrol. Earth Syst. Sci.* **2015**, *19*, 2685–2699. [[CrossRef](#)]
21. Musafir, G.N.; Thompson, M.H. Non-linear optimal multivariate spatial design using spatial vine copulas. *Stoch. Environ. Res. Risk Assess.* **2017**, *31*, 551–570. [[CrossRef](#)]
22. Pereira, G.; Veiga, A. PAR (p)-vine copula based model for stochastic streamflow scenario generation. *Stoch. Environ. Res. Risk Assess.* **2018**, *32*, 833–842. [[CrossRef](#)]
23. Gómez, M.; Ausín, M.C.; Domínguez, M.C. Vine copula models for predicting water flow discharge at King George Island, Antarctica. *Stoch. Environ. Res. Risk Assess.* **2018**, *32*, 2787–2807. [[CrossRef](#)]
24. Avila, L.; Miriam, R.; Kaviski, E. Probabilistic long-term reservoir operation employing copulas and implicit stochastic optimization. *Stoch. Environ. Res. Risk Assess.* **2020**, *34*, 931–947. [[CrossRef](#)]
25. Nápoles, O.M. Bayesian Belief Nets and Vines in Aviation Safety and other Applications. Ph.D. Thesis, Delft University of Technology, Delft, The Netherlands, 2009.
26. Pereira, G.; Veiga, A.; Erhardt, T.; Czado, C. Spatial R-vine copula for streamflow scenario simulation. In Proceedings of the 2016 Power Systems Computation Conference (PSCC), Genoa, Italy, 20–24 June 2016; pp. 1–7.
27. Hernandez, J.A. Vine Copula Modelling of Dependence and Portfolio Optimization with Application to Mining and Energy Stock Return Series from the Australian Market. Ph.D. Thesis, Edith Cowan University, Perth, Australia, 2015.
28. Cooke, R.M.; Joe, H.; Chang, B. Vine regression. In *Resources for the Future Discussion Paper*; RFF: Washington, DC, USA, 2015; pp. 15–52.
29. Kraus, D.; Czado, C. D-vine copula based quantile regression. *Comput. Stat. Data Anal.* **2017**, *110*, 1–18. [[CrossRef](#)]
30. Sklar, M. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **1959**, *8*, 229–231.
31. Nelsen, R.B. *An Introduction to Copulas*; Springer Science & Business Media: Portland, OR, USA, 2007.
32. Aas, K.; Czado, C.; Frigessi, A.; Bakken, H. Pair-copula constructions of multiple dependence. *Insur. Math. Econ.* **2009**, *44*, 182–198. [[CrossRef](#)]
33. Dißmann, J.F. Statistical Inference for Regular Vines and Application. Diploma Thesis, Zentrum Mathematik, Technische Universität München, München, Germany, 2010.
34. Bedford, T.; Cooke, R.M. Probability density decomposition for conditionally dependent random variables modeled by vines. *Ann. Math. Artif. Intell.* **2001**, *32*, 245–268. [[CrossRef](#)]
35. Joe, H. *Families of m-Variate Distributions with Given Margins and m (m-1)/2 Bivariate Dependence Parameters*; Lecture Notes-Monograph Series; Institute of Mathematical Statistics: Hayward, ON, Canada, 1996; pp. 120–141.

36. Czado, C. *Analyzing Dependent Data with Vine Copulas*; Lecture Notes in Statistics; Springer: Garching, Germany, 2019.
37. Dissmann, J.; Brechmann, E.C.; Czado, C.; Kurowicka, D. Selecting and estimating regular vine copulae and application to financial returns. *Comput. Stat. Data Anal.* **2013**, *59*, 52–69. [[CrossRef](#)]
38. Czado, C.; Jeske, S.; Hofmann, M. Selection strategies for regular vine copulae. *J. Société Française Stat.* **2013**, *154*, 174–191.
39. Singh, V.P. *Entropy Theory and Its Application in Environmental and Water Engineering*; John Wiley & Sons: Carrllton, TX, USA, 2013.
40. Alina, B. Copula Modeling for World's Biggest Competitors. Master's Thesis, Amsterdam School of Economics, Universiteit Van Amsterdam, Amsterdam, The Netherlands, 2018.
41. Chang, B.; Joe, H. Prediction based on conditional distributions of vine copulas. *Comput. Stat. Data Anal.* **2019**, *139*, 45–63. [[CrossRef](#)]
42. El Adlouni, S. Quantile regression C-vine copula model for spatial extremes. *Nat. Hazards* **2018**, *94*, 299–317. [[CrossRef](#)]
43. Rosenblatt, M. Remarks on a multivariate transformation. *Ann. Math. Stat.* **1952**, *23*, 470–472. [[CrossRef](#)]
44. Schepsmeier, U. Efficient information based goodness-of-fit tests for vine copula models with fixed margins: A comprehensive review. *J. Multivar. Anal.* **2015**, *138*, 34–52. [[CrossRef](#)]
45. Newton, B.; Burrell, B.C. The April–May 2008 flood event in the Saint John River Basin: Causes, assessment and damages. *Can. Water Resour. J./Rev. Can. Des Ressources Hydr.* **2016**, *41*, 118–128. [[CrossRef](#)]
46. El-Jabi, N.; Rousselle, J. *Hydrologie Fondamentale*; Éditions de l'École Polytechnique de Montréal: Montréal, QC, Canada, 1987.
47. Mediero, L.; Jiménez-Álvarez, A.; Garrote, L. Design flood hydrographs from the relationship between flood peak and volume. *Hydrol. Earth Syst. Sci.* **2010**, *14*, 2495. [[CrossRef](#)]
48. Pramanik, N.; Panda, R.K.; Sen, D. Development of design flood hydrographs using probability density functions. *Hydrol. Process. Int. J.* **2010**, *24*, 415–428. [[CrossRef](#)]
49. Langat, P.K.; Kumar, L.; Koech, R. Identification of the most suitable probability distribution models for maximum, minimum, and mean streamflow. *Water* **2019**, *11*, 734. [[CrossRef](#)]
50. Perreault, L.; Bobée, B. *Loi Généralisée des Valeurs Extrêmes: Propriétés Mathématiques et Statistiques: Estimation des Paramètres et des Quantiles X_T de période de Retour T*; Number R350; INRS-Eau: Québec, QC, Canada, 1992.
51. Vuong, Q.H. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* **1989**, *57*, 307–333. [[CrossRef](#)]