

## Stochastic modeling of hydroclimatic processes using vine copulas

Pouliasis, George ; Torres-Alves, Gina Alexandra; Morales-Napoles, Oswaldo

**DOI**

[10.3390/w13162156](https://doi.org/10.3390/w13162156)

**Publication date**

2021

**Document Version**

Final published version

**Published in**

Water (Switzerland)

**Citation (APA)**

Pouliasis, G., Torres-Alves, G. A., & Morales-Napoles, O. (2021). Stochastic modeling of hydroclimatic processes using vine copulas. *Water (Switzerland)*, 13(16), 1-19. Article 2156. <https://doi.org/10.3390/w13162156>

**Important note**

To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

Article

# Stochastic Modeling of Hydroclimatic Processes Using Vine Copulas

George Poulialis \*, Gina Alexandra Torres-Alves and Oswaldo Morales-Napoles 

Hydraulic Structures and Flood Risk, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands; G.A.torresalves@tudelft.nl (G.A.T.-A.); o.moralesnapoles@tudelft.nl (O.M.-N.)

\* Correspondence: g.poulialis@student.tudelft.nl

**Abstract:** The generation of synthetic time series is important in contemporary water sciences for their wide applicability and ability to model environmental uncertainty. Hydroclimatic variables often exhibit highly skewed distributions, intermittency (that is, alternating dry and wet intervals), and spatial and temporal dependencies that pose a particular challenge to their study. Vine copula models offer an appealing approach to generate synthetic time series because of their ability to preserve any marginal distribution while modeling a variety of probabilistic dependence structures. In this work, we focus on the stochastic modeling of hydroclimatic processes using vine copula models. We provide an approach to model intermittency by coupling Markov chains with vine copula models. Our approach preserves first-order auto- and cross-dependencies (correlation). Moreover, we present a novel framework that is able to model multiple processes simultaneously. This method is based on the coupling of temporal and spatial dependence models through repetitive sampling. The result is a parsimonious and flexible method that can adequately account for temporal and spatial dependencies. Our method is illustrated within the context of a recent reliability assessment of a historical hydraulic structure in central Mexico. Our results show that by ignoring important characteristics of probabilistic dependence that are well captured by our approach, the reliability of the structure could be severely underestimated.

**Keywords:** vine copula; copula; stochastic simulation; intermittent behavior; multivariate simulation; time series; hydroclimatic processes



**Citation:** Poulialis, G.; Torres-Alves, G.A.; Morales-Napoles, O. Stochastic Modeling of Hydroclimatic Processes Using Vine Copulas. *Water* **2021**, *13*, 2156. <https://doi.org/10.3390/w13162156>

Academic Editors: Panayiotis Dimitriadis, Ioannis Tsoukalas and Demetris Koutsoyiannis

Received: 13 July 2021  
Accepted: 30 July 2021  
Published: 5 August 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

In the field of hydrology, the study of time series and their synthetic generation is of great importance. They are used to drive models in a wide range of applications, from reservoir design [1–3] and planning [4–7], ecological flow estimation [8], to flood risk [9]. There are a fair amount of models available in the literature and despite potential differences between them, they share two common goals: i) adequate modeling of the marginal distribution and ii) adequate characterization of the auto-correlation structure.

Regarding the modeling of time series of hydroclimatic variables, several challenges can be identified. Firstly, these variables typically present highly skewed behavior, especially in fine time scales [10,11]. In common practice, skewed distributions are simulated using the Thomas–Fiering models [12]. These are linear models that approximate skewness by introducing non-Gaussian white noise in the generation scheme; however, the estimation of white noise distribution parameters requires moments with an order higher than two. The estimates of higher-order moments are very uncertain [13] and have a significant impact on the reproduction of the distribution [14]. Additionally, the linear combination of the i.i.d. non-Gaussian white noise terms does not necessarily result in the target distribution [15], but rather a distribution that has equal moments up the preserved order [16]. Furthermore, hydroclimatic processes exhibit a wide range of dependence structures. Typically, these are categorized as short-range dependence (SRD) and long-range dependence

(LRD) [17]. The main characteristic of SRD is the fast decay of the auto-correlation after a few lags, while LRD implies the persistence of high auto-correlation after many lags. It has been shown that LRD is present in many geophysical processes [17–21]. Most applications employ models that are able to reproduce only SRD with some notable exceptions [22,23].

Modeling auto-correlation with linear models (such as those typically used to account for SRD) can have some serious implications. Tsoukalas et al. [16] show that linear models result in unnatural bounded auto-correlation structures when processes other than Gaussian are simulated. This effect was termed *envelope behavior*. Envelope behavior is caused by the linear combination of white noise described by bounded distributions in the time series generation scheme. For example, when a Pearson-III distribution (which is bounded from below) is used to model the white noise terms in a first-order auto-regressive model (AR(1)), the resulting auto-correlation structure is bounded from below as well [16].

Tsoukalas et al. ([16]) show the importance of shifting the focus from the traditional narrow view of correlation, as expressed by the Pearson correlation coefficient, to the dependence structure in a joint distribution. This is a very important step not only in an effort to recognize the limitations of linear stochastic models, but also to understand better the dependence relationships between hydroclimatic variables. Contemporary research indicates that in many cases this relationship deviates from the Gaussian and the effects on the resulting uncertainty estimates are significant [24,25].

A promising alternative can be found in copulas. Copulas are joint distributions with uniform  $[0, 1]$  margins. They allow the modeling of the joint distribution through the decoupling of the dependence structure between variables and their marginal distributions. In many cases, this enables the entire description of the dependence structure between variables (or between lagged versions of the same variable) to one measure, such as Spearman's coefficient. Moreover, copula models are able to reproduce explicitly the marginal distributions.

In comparison to fitting a linear model, copula models can be relatively more complex and time-consuming. In order to fit a variable to a copula function, knowledge about the (physical) behavior of such a variable is required. Additionally, the complexity of these models grows significantly when dimensionality increases. Consequently, this hinders mathematical and computational handling as well as conceptual understanding. Nevertheless, copulas have been widely used and it is a rapidly growing field.

Copulas have gained popularity among the hydrological community [25–29]. Bivariate copulas have been used to generate synthetic time series at the River Nile [30] and the Colorado River [31]. Further, they have been applied to generate multi-site precipitation time series [32]. In the field of hydraulic engineering, for example, [33] applies copulas to produce wind speed and wave height time series for the scheduling of off-shore operations.

Bivariate copulas are limited on the order of auto-correlation they can preserve and the number of variables (for example in terms of geographical location) they can simultaneously simulate (some recent developments allow for approximating non-parametrically complex joint densities. See for example [34,35]). For this reason, extended copula models, i.e., vine copulas, can be used to represent a more complex joint distribution. This method, which was introduced by [36], uses bivariate copulas to decompose multivariate joint distributions. For example, ref. [37] generated cyclostationary stream flow time series considering higher-order auto-dependence. Jäger and Nápoles [38] demonstrated how to produce correlated significant wave heights and mean zero-crossing periods. Sarmiento et al. [39] successfully employed a regular vine (R-vine) model developed by [40] for the simulation of wind speed and direction time series. Vine copulas require the fitting of a large number of copula functions. For example, for the simulation of four variables (taking into consideration only the first-order auto-dependence and their cross-dependence), it would require the calculation of 22 copula parameters.

In this paper, we present a methodology that can be used to generate synthetic time series of hydroclimatic variables. Specifically, we present a case study that concerns an ancient dike that was located within an extinct lake in present-time Mexico City. In order

to recreate the hydrological conditions of the lake at the time where the dike was used, time series of precipitation and evaporation are needed. A first attempt to characterize the dependence between the two variables is the model developed in [41]. In this paper, the model presented in [41] is extended to reproduce the first-order auto-dependence structure in the evaporation process. Additionally, specific considerations to better reproduce cross-correlations between the evaporation and precipitation processes are introduced. Furthermore, a novel framework to simulate from an arbitrary number of geographical locations is presented. This approach introduces the coupling of the spatial and temporal models using repetitive sampling. In this way, the spatial model and the temporal model can be freely chosen according to each data set and the required parameters for the quantification of the joint distribution are reduced.

This paper starts by describing the general methodology to generate the time series. Next, the case study is presented including the description of the data. Then, the precipitation–evaporation model is applied to simulate daily realizations for one and four stations. Finally, a discussion of the results, conclusions, and recommendations for future work are presented.

## 2. Methods

### 2.1. First-Order Univariate Processes

At the core of the modeling methodology proposed here lies the notion of copulas. Copulas are multivariate distributions with uniform marginals in  $[0, 1]$ . They provide an elegant way to describe the dependence between variables by removing the effects of the marginal distributions. Sklar [42] showed that any multivariate joint distribution can be expressed as a function of its marginals. In a two-dimensional context his theorem states:

$$F_{X_1, X_2}(x_1, x_2) = C\{F_{X_1}(x_1), F_{X_2}(x_2)\} \tag{1}$$

where  $X_1$  and  $X_2$  are random variables with marginal distributions  $F_{X_1}(x_1)$  and  $F_{X_2}(x_2)$  and  $C$  the copula function. Equation (1) indicates that two components are required to define the joint distribution: the copula function and the marginal distributions.

Bivariate copulas have been used before to model time series [43]. The simplest case of temporal dependence regards only two time steps. Let  $\{X_t\}$  for  $t \in \mathbb{N}$  denote the time series of interest. In terms of a copula, the relationship between consecutive time steps can be written as  $F_{X_t, X_{t-1}}(x_t, x_{t-1}) = C\{F(x_t), F(x_{t-1})\}$  where,  $F(x_t) = F_{X_t}(x_t)$  is the marginal distribution of  $X_t$ . Moreover,  $P(X \leq x_t | X_{t-1} = x_{t-1}) = F(x_t | x_{t-1}) = C_{X_t | X_{t-1}}\{F(x_t) | F(x_{t-1}), \Theta\}$ , where  $C_{X_t | X_{t-1}}$  is the conditional copula with arguments  $\in [0, 1]^2$ . Notice that the parameter vector  $\Theta$  would model auto-correlation of order 1 for the time series of interest. Since  $\{X_t\}$  is assumed to be a stationary process  $F(x_t | x_{t-1})$  corresponds to a first-order Markov model, which is characterized with a single marginal distribution and an appropriate copula. Often, the conditional distribution is expressed as [44]:

$$C_{X_t | X_{t-1}}\{F(x_t) | F(x_{t-1})\} = \frac{\partial C\{F(x_t), F(x_{t-1})\}}{\partial F(x_{t-1})} \tag{2}$$

For simplicity, a different notation is introduced based on Aas et al. [45], according to which  $h_{X_t, X_{t-1}}(F(x_t), F(x_{t-1}), \Theta) = \frac{\partial C\{F(x_t), F(x_{t-1})\}}{\partial F(x_{t-1})}$ . Moreover, let  $h_{X_t, X_{t-1}}^{-1}(F(x_{t-1}), v, \Theta)$  be the inverse conditional distribution, where the  $F(x_{t-1})$  is the conditioning variable,  $v$  is an independent uniform  $[0, 1]$  random variable and  $\Theta$  the parameter vector of the bivariate copula  $C_{X_t, X_{t-1}}$ . Solving for  $x_t$  yields the definition of the first-order univariate process.

$$x_t = F^{-1}\left[h_{X_t, X_{t-1}}^{-1}\left(F(x_{t-1}), v, \Theta\right)\right] \tag{3}$$

where  $F^{-1}$  is the inverse marginal distribution. The process, briefly described in this section, has been used previously, for example, to model traffic loads for bridge reliability in [46,47].

### 2.2. First-Order Bivariate Processes

In this section, the univariate model briefly described in Section 2.1 is extended to account for dependence between two variables. Let us consider two dependent processes  $\{X_t\}$  and  $\{Y_t\}$  that present additionally serial correlation. Similarly, as in the univariate case,  $t \in \mathbb{N}$ . A good alternative to model these kinds of processes are vine copulas or simply vines. Vines were originally introduced in Joe [44], Cooke [36], and Bedford and Cooke [48] (see [49]). Vines are a graphical way of representing multivariate joint distributions. They are a generalization of dependence trees. Roughly, a vine on  $n$  elements  $V(n) = \{T_1, \dots, T_n\}$  is a nested set of trees where the edges of tree  $j$  are nodes on the tree  $j + 1$  for  $j = 1, \dots, n - 1$ . In particular, *regular vines* are of interest. These are vines whose edges in tree  $j$  are connected as nodes in tree  $j + 1$  only if they share a common node in tree  $j$ . Vine copulas as dependence models assign a bivariate copula to each edge in the first tree of the vine and conditional bivariate copulas to the edges of every tree  $> 1$ . For a formal definition and their statistical treatment see for example [45]. A graphical representation of a regular vine on 3 nodes is presented in Figure 1. The edges in Figure 1 are assigned copulas that are parametrized by rank correlation coefficients and a conditional rank correlation of  $X_t, Y_t|Y_{t-1}$ . Rank correlations are denoted by  $r$  followed by a subscript indicating the respective variables and time step. For example,  $r_{Y_t Y_{t-1}}$  denotes the rank correlation between process  $Y$  at times  $t$  and  $t - 1$ .

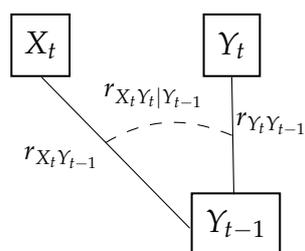


Figure 1. Vine representation of a joint distribution in 3 nodes.

Notice that the essential correlations that the model needs to capture are the temporal dependencies  $\{X_t\}$  and  $\{Y_t\}$  and the cross-correlation  $\{X_t, Y_t\}$ . If the process  $\{X_t\}$  is generated independently with the univariate conditional copula model discussed in Section 2.1, then all that is left is to induce the correct dependence  $\{X_t, Y_t\}$  and  $\{Y_t\}$ . This can be achieved by modeling  $Y_t$  conditional on both  $Y_{t-1}$  and  $X_t$  with a vine as the one presented in Figure 1.  $Y_t$  can be sampled according to Equation (4) where  $F_X$  and  $F_Y$  are the marginal distributions underlying  $X_t$  and  $Y_t$  ( $F_X^{-1}$  and  $F_Y^{-1}$  their inverses). Notice that the time index  $t$  is omitted since both processes are stationary. Moreover,  $h$  is the conditional copula (and  $h^{-1}$  its inverse) operator, as it was established earlier, and  $v$  is as before, an independent uniform random variable on  $[0,1]$ .

$$y_t = F_Y^{-1} \left[ h_{Y_t, Y_{t-1}}^{-1} \left( F_Y(y_{t-1}), h_{Y_t, X_t | Y_{t-1}}^{-1} (F_X(x_t), v, \Theta_1, \Theta_2) \right) \right] \tag{4}$$

The modeling procedure for the generation of a continuous bivariate stochastic process may be summarized as:

- i Fit the appropriate marginal distributions characterising the random variables (or use the empirical distribution).
- ii Select suitable copulas  $C_{X_t, X_{t-1}}$ ,  $C_{Y_t, Y_{t-1}}$ , and  $C_{Y_t, X_t | Y_{t-1}}$ .
- iii Generate the first variable ( $X_t$ ) using the univariate copula model from the previous subsection.
- iv Generate the second variable ( $Y_t$ ) using Equation (4).

Similar models to the one briefly described in this section have been used in [33,38].

### 2.3. First-Order Intermittent-Continuous Bivariate Processes

In this section, the algorithm introduced in Section 2.2 is expanded to take into account intermittency. Results are based on the methodology presented in Torres-Alves et al. [41]. The basic idea is the coupling of the copula models with Markov chains. Specifically, the intermittent process is split into two sub-processes, (i) a discrete state-time Markov Chain accounting for the occurrence of a certain environmental condition, and (ii) copula models accounting for “amounts”. For example, in the case of daily precipitation, the Markov chain model would simulate days with or without rain while the copula models would account for the amount of rain (mm) per day. In an intermittent process, a Markov Chain model accounts for two states: one for zero amount (i.e., a dry day) and one for larger than zero amount (i.e., a wet day). The first-order discrete state-time Markov process is described by:

$$P(X_t = s | X_{t-1}, X_{t-2}, \dots, X_1) = P(X_t = s | X_{t-1}) \quad (5)$$

Similar to previous subsections,  $t \in \mathbb{N}$  while  $s \in S$  for a countable set  $S$ . The probabilities that express the chance of transition between states are called transition probabilities. Continuing with our example, the two states of precipitation could be *wet*  $\Rightarrow 1$  and *dry*  $\Rightarrow 0$ , thus  $S = \{0, 1\}$ . The probabilities of interest are defined as follows:

$$\begin{aligned} P(X_t = 1 | X_{t-1} = 1) &= P_{1|1} \\ P(X_t = 0 | X_{t-1} = 1) &= P_{0|1} \\ P(X_t = 1 | X_{t-1} = 0) &= P_{1|0} \\ P(X_t = 0 | X_{t-1} = 0) &= P_{0|0} \end{aligned} \quad (6)$$

For the above probabilities it holds that  $P_{1|1} = 1 - P_{0|1}$  and  $P_{1|0} = 1 - P_{0|0}$ . These probabilities comprise the so-called transition-probability matrix:

$$P = \begin{bmatrix} P_{0|0} & P_{1|0} \\ P_{0|1} & P_{1|1} \end{bmatrix} \quad (7)$$

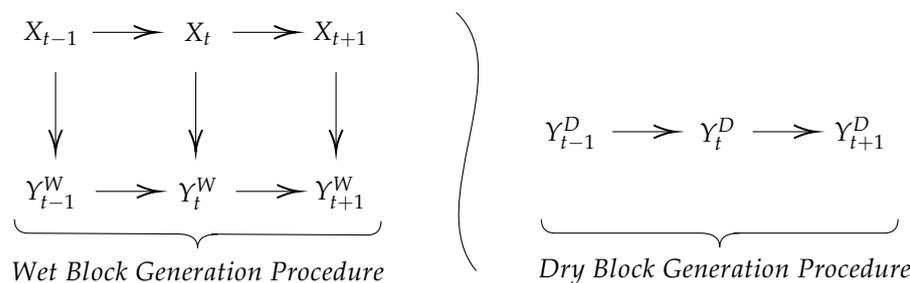
This matrix characterizes the Markov chain, providing all the information necessary to reproduce the process. In this way, the probability of each state is explicitly simulated. Another way to approach intermittency could be by adopting mixed type distributions that consider the dry state as an atom of probability. These have been adopted in literature [50,51] but they will not be discussed in this text; however, they are theoretically compatible with the proposed copula models.

The way that intermittency has been approached, by splitting the main process into two sub-processes of occurrence and amount dictates that the generation algorithm cannot be continuous. Specifically, the generation of the “occurrence” process must precede the “amount” process. In other words, the Markov chain must precede the copula time series.

Consequently, to generate a synthetic time series, the first step is to generate realizations of the Markov chain. This will define the different state blocks (i.e., sequences of wet and dry periods of time) of the time series. For the dry (zero amount state) block of the intermittent process, no further action is needed. What follows is the generation of the amount process for the wet blocks. This can be performed for each block by employing the univariate model of Equation (3). Then the continuous process can be generated conditionally on the intermittent one of the same time step using the vine copula model of Equation (4); however, dry blocks require special treatment. Naturally, zero amount does not allow the expression of dependence via a copula. To overcome this, the cross-correlation is treated implicitly by conditionalizing the continuous process distribution according to the two states, wet and dry.

On a final note, to preserve correctly the first-order auto-dependence structure of the continuous process the generation of blocks must be sequential. This means that the first realization of each block should be conditional on the final realization of the previous block.

The generation procedure is schematized in Figure 2. Superscripts  $W$  and  $D$  denote the conditional wet and dry continuous process marginal distributions, respectively.



**Figure 2.** Proposed bivariate (intermittent process  $X_t$ , continuous process  $Y_t$ ) generation algorithm explanatory diagram. Superscripts  $W$  and  $D$  denote the conditional wet and dry continuous process marginal distributions, respectively.

The generation algorithm is summarized below:

- i Fit the appropriate marginal distributions characterizing the RV's.
- ii Select and fit the suitable copulas  $C_{X_t, X_{t-1}}$ ,  $C_{Y_t, Y_{t-1}}$ , and  $C_{Y_t, X_t | Y_{t-1}}$ .
- iii Calculate the transition-probability matrix of the Markov chain for the intermittent process  $X_t$ . Simulate a desired length of the Markov sequence.
- iv Split the time series into dry and wet blocks.
- v Generate “blocks” representing  $\{X_t\}$  using the univariate copula model (Equation (3)).
- vi Generate the wet block  $Y_t$  using the vine model of Equation (4) and the appropriate marginal distribution. Use as seed for the first value of the wet block the last value of the previous dry block.
- vii Generate the dry block  $Y_t$  using the vine model (Equation (3)) and the appropriate marginal distribution. Use as seed for the first value of the dry block the last value of the previous wet block.

#### 2.4. Multivariate Processes

Equations (3) and (4) define univariate and bivariate models, respectively; however, using the vine decomposition, higher-order models can be defined. As it was discussed during the introduction, while such models that account for temporal and spatial dependence exist, they require the calculation of a great number of parameters and are limited in the dependence structures they can represent. The methodology described herein aims to approximate dependent processes in time and space (for example precipitation and evaporation from an arbitrary number of stations) while remaining parsimonious and flexible. The key of this methodology is to approximate the conditional distribution of the spatial and temporal dependence through repetitive sampling instead of inferring it theoretically. It should be noted that the use of repetitive sampling in the context of hydrology has been used in [52,53] for the coupling of different temporal scales.

For simplicity, let us first assume a two-dimensional case. Let us define a two-dimensional process  $\{Y_t^1, Y_t^2\}$  where the superscripts denote, for example, a corresponding station and the subscripts the time step. Moreover, let us assume that the temporal and spatial dependence can be described by the first-order model of Equation (3). The methodology can be described in four steps:

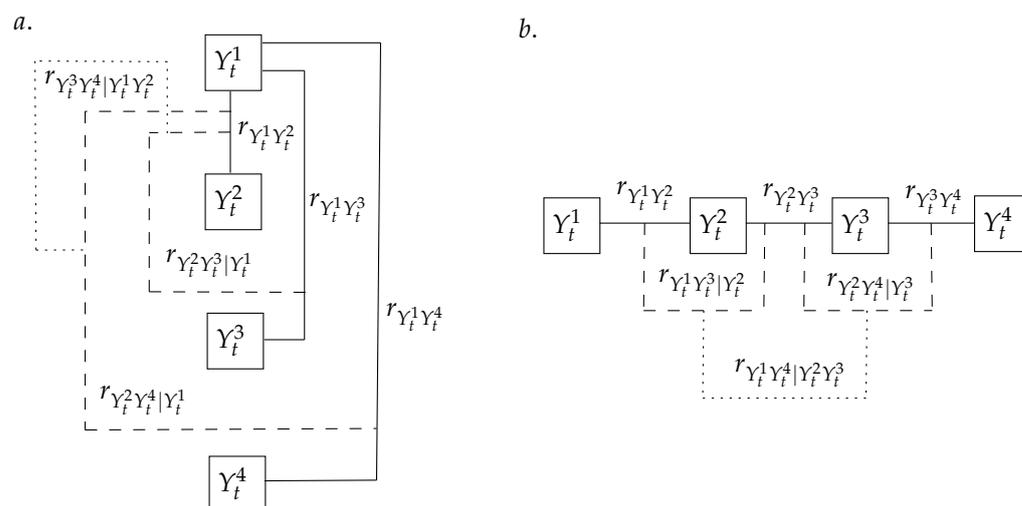
- i Generate one realization of  $\{Y_t^1\}$  from  $Y_{t-1}^1$  based on the temporal model.
- ii Generate  $n$  realizations of  $Y_t^2$  based on  $Y_t^1$  according to the spatial model. These are denoted with a tilde and the superscript  $S2$  as  $[\tilde{Y}_{t,1}^{S2}, \dots, \tilde{Y}_{t,n}^{S2}]$ .
- iii Generate  $n$  realizations of  $Y_t^2$  based on  $Y_{t-1}^2$  according to the temporal model. These are denoted with a tilde and the superscript  $T2$  as  $[\tilde{Y}_{t,1}^{T2}, \dots, \tilde{Y}_{t,n}^{T2}]$ . The spatial and

- temporal realizations are all plausible realizations of  $Y_t^2$ , which implies that a section between the two sets exists.
- iv Identify the common space. This is performed by identifying the realizations that minimize the root mean squared error (RMSE).

$$\min_i \{ \sqrt{(\tilde{Y}_{t,i}^{T2} - \tilde{Y}_{t,i}^{S2})^2} \} \tag{8}$$

Notice that RMSE in step iv above is not used as a goodness of fit measure but rather as a selection criterion between the spatial and temporal components of the model. Another way to approach this could be a Monte Carlo procedure where a convergence criterion is targeted; however, in the context of time series generation this would result in non-practical computational demand. Other measures for selection may be very well possible; however, investigating them is out of the scope of this paper since we focus mainly on the general methodology for simulating spatially diverse data sources.

The above procedure can easily be extended to more dimensions. Let us consider another case where a set of four correlated variables  $[Y_t^1, Y_t^2, Y_t^3, Y_t^4]$  is generated. It is reminded that the vine representation in four variables is not unique [54]. Two very common vines are the canonical or a C-Vine and the D-Vine. Their graphical representations are found in Figure 3.



**Figure 3.** Graphical representation of four-dimensional C-vine (a) and D-vine (b). Figures reconstructed from [54], with permission from Elsevier.

It can be seen that the C-Vine representation has a unique variable, in our case  $Y_t^1$ , from which the other variables depend. On the contrary in the D-Vine representation, the variables depend serially. Both could be probable representations of spatial dependence; however, the algorithm depends on the vine’s sampling order. Let us assume that a C-Vine is selected. Further, let  $v_1, v_2, v_3, v_4$  denote independent random draws from the uniform (0, 1) distribution. The sampling algorithm is given by the following equations:

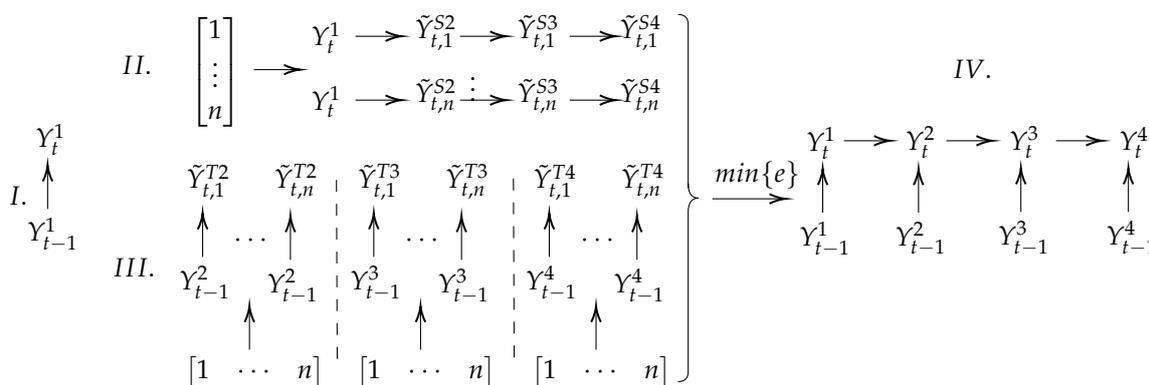
$$\begin{aligned}
 y_t^1 &= F_Y^{-1}(v_1) \\
 y_t^2 &= F_Y^{-1} \left[ h_{Y_t^2, Y_t^1}^{-1} \left( F_Y(y_t^1), v_2, \Theta_1 \right) \right] \\
 y_t^3 &= F_Y^{-1} \left[ h_{Y_t^3, Y_t^1}^{-1} \left( F_Y(y_t^1), h_{Y_t^2, Y_t^1}^{-1} \left( F_Y(y_t^2), v_3, \Theta_2 \right), \Theta_3 \right) \right] \\
 y_t^4 &= F_Y^{-1} \left\{ h_{Y_t^4, Y_t^1}^{-1} \left[ F_Y(y_t^1), h_{Y_t^2, Y_t^1}^{-1} \left( F_Y(y_t^2), h_{Y_t^3, Y_t^1}^{-1} \left( F_Y(y_t^3), v_4, \Theta_4 \right), \Theta_5 \right), \Theta_6 \right] \right\}
 \end{aligned} \tag{9}$$

Furthermore, let us assume that the temporal dependence can be described by the first-order model of Equation (3). The first step is to generate a basis realization on which the

repetitive sampling will be based. This decision depends on the selected spatial dependence model. It is noted that according to Equation (9), the C-Vine sampling order is 1-2-3-4. Thus, in this case, the algorithm should start by computing  $Y_t^1$ . If a D-Vine was selected instead, the algorithm should start by computing  $Y_t^3$  since the sampling order of a D-Vine is 3-2-4-1 [54]. The rest of the steps remain the same.  $n$  realizations are generated according to the spatial and temporal models of which the one that minimizes the RMSE (Equation (10)) is selected.

$$\min_i \left\{ \sqrt{\frac{\sum_{j=1}^m (\tilde{Y}_{t,i}^{Tj} - \tilde{Y}_{t,i}^{Sj})^2}{m}} \right\} \tag{10}$$

where  $m = 4$  in our case. This procedure is schematized in Figure 4.



**Figure 4.** Four-dimensional generation algorithm. The roman numerals denote the steps of the algorithm. Superscripts S and T denote a realization according to spatial and temporal models, respectively. The tilde denotes a possible realization.

The methodology above can be extended for  $m$  stations. The required number of parameters of such a model is:

$$N_{\Theta} = m(m + 1)/2 \tag{11}$$

The multivariate generation algorithm is summarized below:

- i Fit the appropriate marginal distributions characterizing the RV's.
- ii Identify suitable models to describe spatial and temporal dependence.
- iii Select and fit  $N_{\Theta}$  suitable copulas to model dependencies.
- iv Select the number of trials  $n$  for the repetitive sampling.
- v Generate the first temporal realization according to the sampling order of the selected spatial model.
- vi Generate  $n$  realizations from the spatial and temporal models independently.
- vii Select the realization which minimizes the error (Equation (10)).

### 2.5. Admissible Marginal Distributions and Copula Fitting

One of the major benefits of modeling stochastic time series with copulas is the direct use of the marginal distribution, instead of moment approximations. This allows the utilization of classical fitting techniques such as maximum likelihood estimation [55] and L-Moments [56] and state-of-the-art methods such as K-Moments [14] and the metastatistical extreme value framework (MEV) [57]. Moreover, any finite variance distribution is admissible, which increases the flexibility of the models.

The same flexibility can be demonstrated in the admissible copula fitting techniques. In this study, the estimation of copula parameters is performed via the Spearman or Rank correlation coefficient accompanied by measures to test the goodness of fit, such as the Cramer–Von Mises statistic [58] and semi-correlations [43]; however, other methods such as maximum likelihood, Bayesian information criteria, or Kendall's tau can be supported. A more comprehensive review can be found in [25,59].

### 3. Case Study

To demonstrate the effectiveness of the proposed vine copula model, two scenarios are presented: (i) simulation of daily evaporation and precipitation at one station and (ii) simulation of daily evaporation at four stations, all located at the Valley of Mexico. Additionally, the first scenario illustrates the influence of the dependence structure on the reliability of flood defenses.

The proposed methodology is validated on the accuracy of the reproduced historical characteristics, such as the approximation of the historical marginal distributions, the first-order auto-correlation dependence, and the cross-correlation dependence structure.

#### 3.1. Area of Study

The case study described herein is presented in Torres-Alves & Morales-Nápoles (2020) [41]. It concerns the reliability analysis due to overflow of the ancient Nezahualcoyotl dike that was once located in the Valley of Mexico (Figure 5). The Aztecs built this structure around 1450 AD to protect their capital, the city of Tenochtitlan, from rising water levels in the lacustrine system.

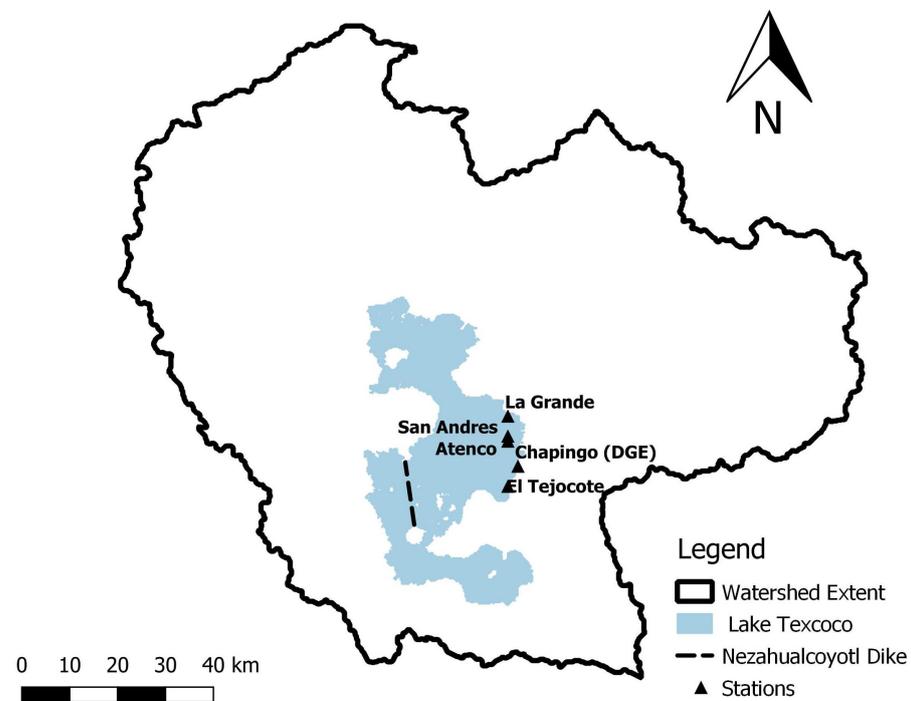


Figure 5. Outline of the study area. Reconstructed from information taken from [41].

Around 1519, the Valley of Mexico was covered by a lacustrine system comprising of six interconnected lakes. These lakes connected during high water and because of the salinity of the lakes, the rising water posed a threat to agriculture and freshwater supplies in Tenochtitlan. Nowadays, there are no remains of this dike and the lakes were almost completely drained by the end of the 19th century. Ref. [41] reconstructed the geometry and position of the ancient structure based on historical sources. The dike was 16 km long with a height of 8 m and a width of 3.5 m. It was made out of wood, stone, and mud. Moreover, ref. [41] estimated the extent of the lacustrine system and proposed a simple hydrological balance equation (Equation (12)) to compute the water level fluctuation at the lake. Finally, an assessment of the reliability of the dike due to overflow was presented.

$$X_t A_{tr} - Y_t A_l = \frac{dV_t}{dt} \quad (12)$$

where:

- $X_t$  is the precipitation at time  $t$  (m/day).
- $Y_t$  is the evaporation at time  $t$  (m/day).
- $A_{tr}$  is the tributary area of the basin (m<sup>2</sup>).
- $A_l$  is the surface area of the lake (m<sup>2</sup>).
- $\frac{dV}{dt}$  is the daily change in volume (m<sup>3</sup>/day).

The hydrological balance was simulated only for the wet seasons. For this reason, an initial water level ( $L_0$ ) needed to be assumed. The dike's probability of failure due to overflow was computed for six initial water levels (1 to 6 m).

### 3.2. Data

To simulate the precipitation and evaporation time series, records from five stations situated within the Valley of Mexico are analyzed. This analysis was originally performed in [41]. Therefore, only a brief overview is provided in the present work. For a detailed presentation of the historical records the reader is referred to the original publication. The data are available from Mexico's national database [60]. Information about the stations is given in Table 1.

**Table 1.** Summary of precipitation and evaporation information for each station.

Code	Name	Coordinates		Time Period
		Longitude	Latitude	
15008	Atenco	−98.92	19.54	1 January 1961–22 August 2013
15044	La Grande	−98.92	19.58	1 January 1964–31 August 2014
15083	San Andres	−98.92	19.53	1 May 1967–31 December 2014
15167	El Tejocote	−98.92	19.44	1 December 1957–1 January 2007
15170	Chapingo (DGE)	−98.9	19.48	13 March 1957–1 January 2000

In this region, two seasons are identified, (i) wet season (May–October), and (ii) dry season (November–April). The data from the stations were analyzed and appropriate distributions were fitted using Maximum likelihood (Table 2). The respective cumulative distribution functions are presented in Appendix A.

**Table 2.** Selected distributions for all variables and stations.

Station	Probability Distribution Function	
	Precipitation	Evaporation
Atenco	Weibull	Generalized Extreme Value
La Grande	Weibull	Generalized Extreme Value
San Andres	Weibull	Generalized Extreme Value
El Tejocote	Weibull	Generalized Extreme Value
Chapingo (DGE)	Weibull	Generalized Extreme Value

Furthermore, bivariate copulas were fitted for each data pair (precipitation–evaporation). The results are presented in Table 3. In [41] the authors use different goodness of fit (GOF) techniques to show that the pairs in Table 3 are adequate choices for the (conditional) copulas of interest. For further details about the GOF techniques used, the reader is referred to [41] and references therein. Their functional representations are given in Appendix B.

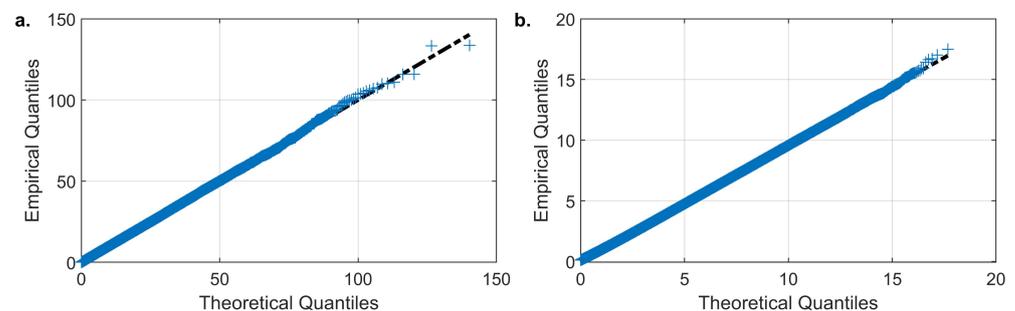
**Table 3.** Selected copula pairs for all stations.

Pairs	Atenco	La Grande	San Andres	El Tejocote	Chapingo (DGE)
$X_t - X_{t-1}$	Gauss	Gauss	Gauss	Gauss	Gauss
$Y_t - Y_{t-1}$	Gumbel	Gumbel	Gumbel	Gumbel	Gumbel
$X_t - Y_t$	Gumbel	Gumbel	Gumbel	Gumbel	Gumbel

### 3.3. Simulation of Daily Evaporation and Precipitation

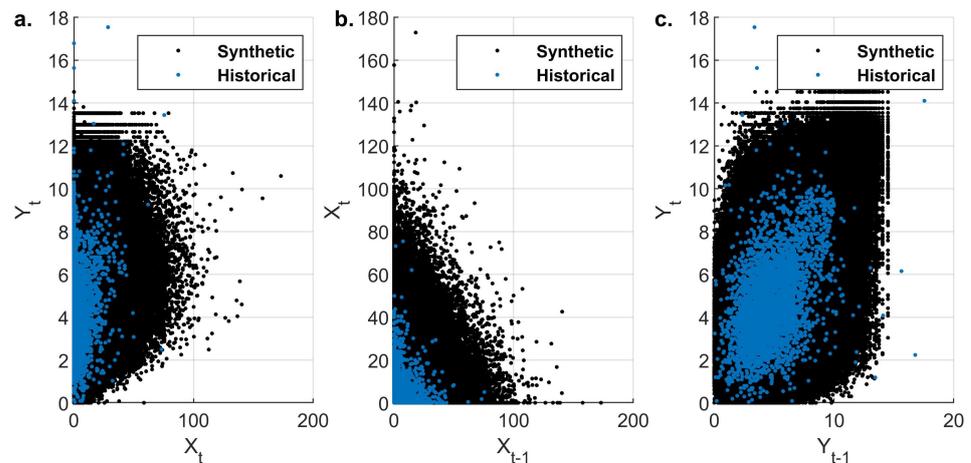
The simulation of precipitation  $\{X_t\}$  and evaporation  $\{Y_t\}$  poses a particular challenge due to the intermittent behavior exhibited by precipitation. Moreover, these variables are the basic drivers for many hydrological models (i.e., [61]). Producing correlated time series enables the quantification of uncertainty in these drivers and a probabilistic description of the results. For this scenario, the data from Atenco station were analyzed (for further details on why this station is selected, refer to [41]). From these data, 5000 hypothetical wet seasons were generated, each with a length of 185 days.

The ability of the model to reproduce the historical distributions is exhibited in Figure 6 with the use of QQ plots. These depict the relationship between the theoretical and empirical quantiles of the cumulative distribution function. In the presented case, the plotted quantiles form an almost straight line; thus, good agreement is found between historical and simulated CDFs for both variables. Figure 7 provides a comparison between historical and synthetic dependence structure in the real domain. Precipitation and evaporation exhibit a moderate correlation in the range of 0.3 to 0.4. Additionally, between the variables, a low negative correlation is observed. These dependencies are quantified in terms of rank correlations in Table 4. Overall, the model reproduces adequately the first-order auto-dependence as well as the dependence between variables.

**Figure 6.** Comparison between empirical and theoretical quantiles for Precipitation  $\{X_t\}$  (a) and Evaporation  $\{Y_t\}$  (b).

#### 3.3.1. The Effect of the Choice of Copula

For the description of the historical first-order auto-dependence, a Gumbel copula was employed; however, in non-copula models addressing directly the characteristics of the observed dependence structure is not straightforward. To demonstrate the importance of a correct dependence description, a reliability calculation will be performed using two sets of the synthetic time series. The first one will be the one generated during the previous section. The second one is a precipitation and evaporation pair with the auto-dependence of evaporation described by a Gaussian copula. The latter is selected for comparison purposes because it represents a simple dependence structure that is reproduced by commonly used linear Gaussian models and more recent models such as [50,51,53,62]. The results for both cases are summarized in Table 5.



**Figure 7.** Scatter plot presenting the dependence structure of historical and synthetic data. (a) Precipitation–Evaporation  $\{X_t, Y_t\}$ , (b) Precipitation–Precipitation  $\{X_t, X_{t-1}\}$ , (c) Evaporation–Evaporation  $\{Y_t, Y_{t-1}\}$ .

**Table 4.** Comparison of historical and simulated rank correlations.

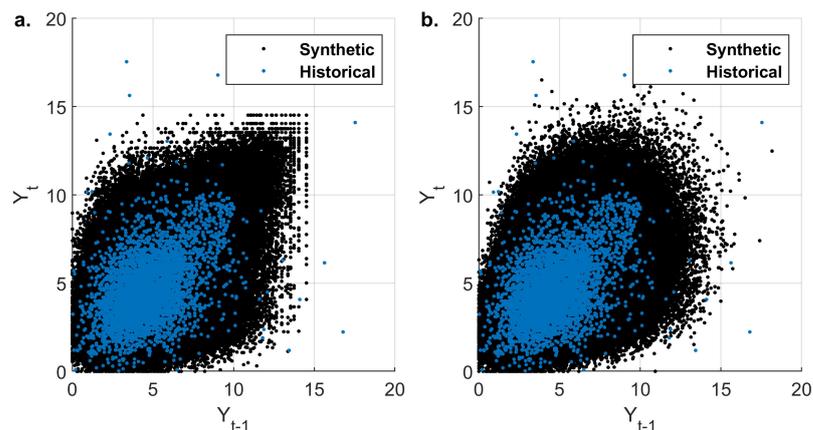
	$X_t - X_{t-1}$	$Y_t - Y_{t-1}$	$X_t - Y_t$
Historical	0.27	0.4	−0.22
Synthetic	0.27	0.37	−0.18

**Table 5.** Probability of failure for different initial water levels ( $L_0$ ) as calculated with the Gaussian and Gumbel time series.

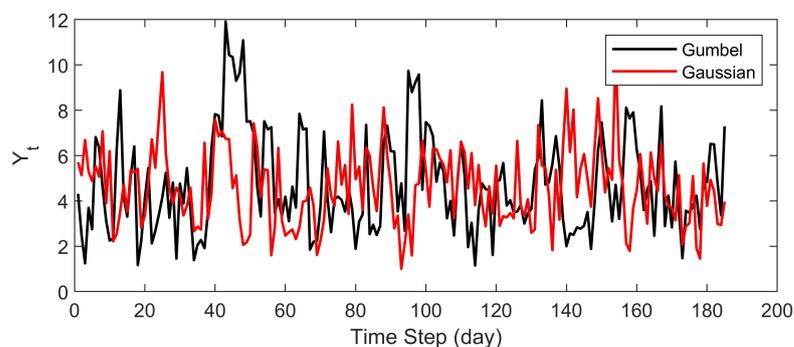
Water Level at the Foot of the Dike (m) ( $L_0$ )	m.a.s.l.	Gaussian Copula		Gumbel Copula	
		$P_f$	Return Period (Years)	$P_f$	Return Period (Years)
1	2231	0.00171	586	0.00099	1006
2	2232	0.02641	38	0.02068	48
3	2233	0.16189	6	0.14956	7
4	2234	0.45131	2	0.44657	2
5	2235	0.74568	1	0.75102	1
6	2236	0.91468	1	0.92038	1

Given an initial water level of  $L_0 = 1$  m. the probability of failure computed for the Gaussian copula case is 73% higher than the Gumbel case.

Note that the Gaussian copula reproduces the same first-order rank auto-correlation as the Gumbel. The difference can be entirely attributed to the change in the shape of the dependence structure (Figure 8). In the absence of tail dependence, there is no clustering of high evaporation events. This moderates the outgoing flux of the lake resulting in more frequent high-water levels and thus, a larger probability of failure. This can be illustrated by comparing the two evaporation time series (Figure 9).



**Figure 8.** Scatter plot comparison of generated Evaporation  $\{Y_t\}$  auto-dependence structures with (a) Gumbel copula and (b) Gaussian copula. Historical data plotted as reference.

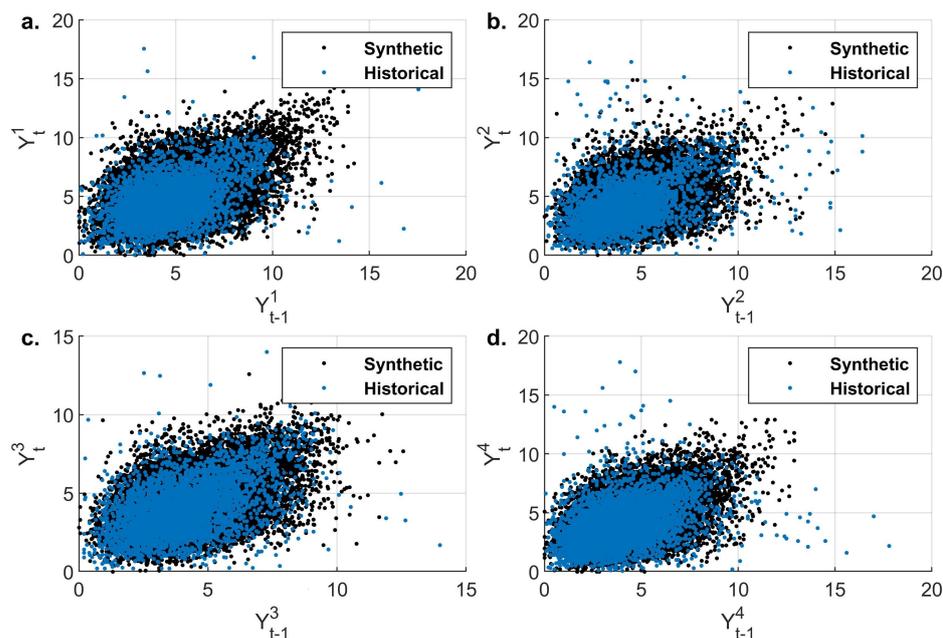


**Figure 9.** Evaporation  $\{Y_t\}$  time series generated by Gumbel and Gaussian copulas.

In Figure 9, the time series generated by the Gumbel copula presents clusters of high evaporation values around time steps 40 to 70, 100 and 160. On the contrary, the time series generated by the Gaussian copula demonstrate a more symmetrical dependence pattern. This is in accordance with the scatter plots provided in Figure 8 where the difference of tail dependence between both copulas is demonstrated. Additionally, it is observed that the differences between the computed probabilities decrease as the initial water level  $L_0$  increases (Table 5). This is because larger probabilities (smaller return periods) correspond to events further from the tail, where the differences between the copula functions are found. These results provide evidence on the importance of selecting an appropriate model that characterizes the dependence structure of extreme events.

### 3.4. Simulation of Evaporation across Multiple Stations

This section presents the application of the multivariate methodology. The developed multivariate algorithm is implemented for the simulation of daily evaporation in four of the five stations at the Valley of Mexico. Evaporation data from Atenco  $\{Y_t^1\}$ , La Grande  $\{Y_t^2\}$ , San Andres  $\{Y_t^3\}$ , and Chapingo (DGE)  $\{Y_t^4\}$  were used to generate 100 seasons (each of 185 days). The spatial dependence is described by a Gaussian copula while the temporal dependence is described by a Gumbel copula. For the latter, only the first-order auto-dependence is taken into account. The sampling number is set to  $n = 100$ . A comparison between historical and simulated first-order auto-dependence is provided in Figure 10.



**Figure 10.** Scatter plot of historical and synthetic auto-dependence structures. (a) Atenco  $\{Y_t^1\}$ , (b) La Grande  $\{Y_t^2\}$ , (c) San Andres  $\{Y_t^3\}$ , and (d) Chapingo (DGE)  $\{Y_t^4\}$  stations.

On the temporal level, evaporation in all stations demonstrates a rank correlation in the range of 0.35 to 0.45. Moreover, the historical data exhibit upper tail dependence between consecutive time steps. In all cases, the simulated data reproduce the shape of the historical dependence structure. The results are quantified in Table 6 where a very good agreement is found between historical and simulated rank correlations.

**Table 6.** Comparison of historical and simulated rank auto-correlations.

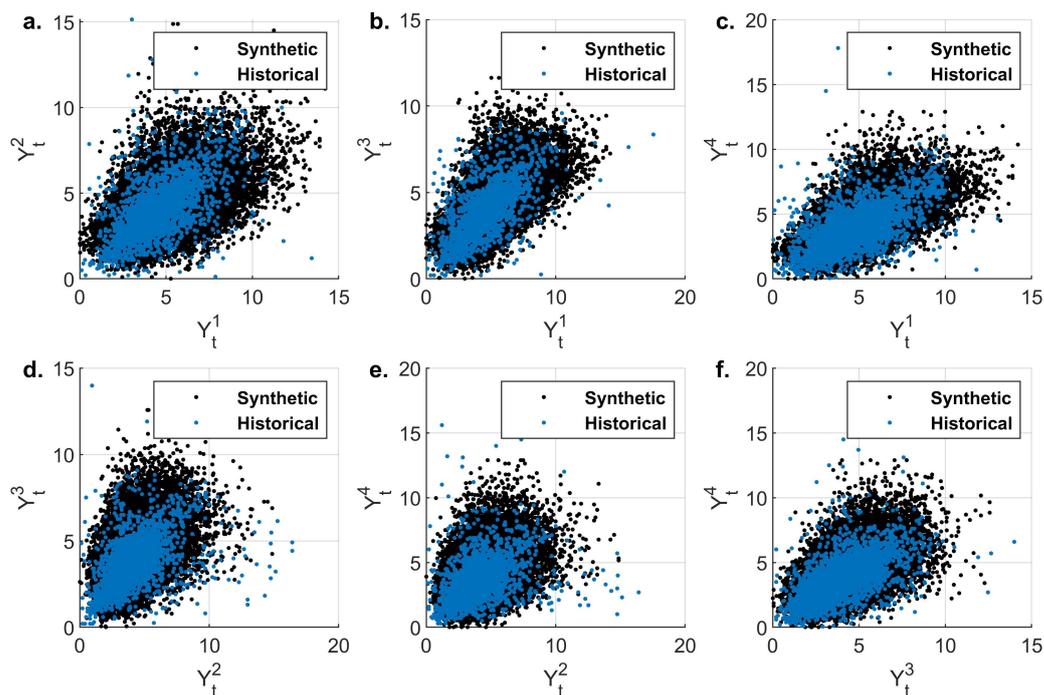
	$Y_t^1 - Y_{t-1}^1$	$Y_t^2 - Y_{t-1}^2$	$Y_t^3 - Y_{t-1}^3$	$Y_t^4 - Y_{t-1}^4$
Historical	0.41	0.36	0.45	0.44
Synthetic	0.41	0.37	0.46	0.40

Furthermore, evaporation exhibits a high spatial correlation, in the order of  $r \approx 0.6$  among the stations. From a physical point of view, this can be explained by the proximity of the stations and their position within the same watershed. The proposed algorithm was able to reproduce these dependencies to a good extent (Figure 11); however, a small deviation is observed between pairs  $Y_t^1 - Y_t^4$  and  $Y_t^2 - Y_t^3$  (Table 7). This difference can be attributed to the copula choice.

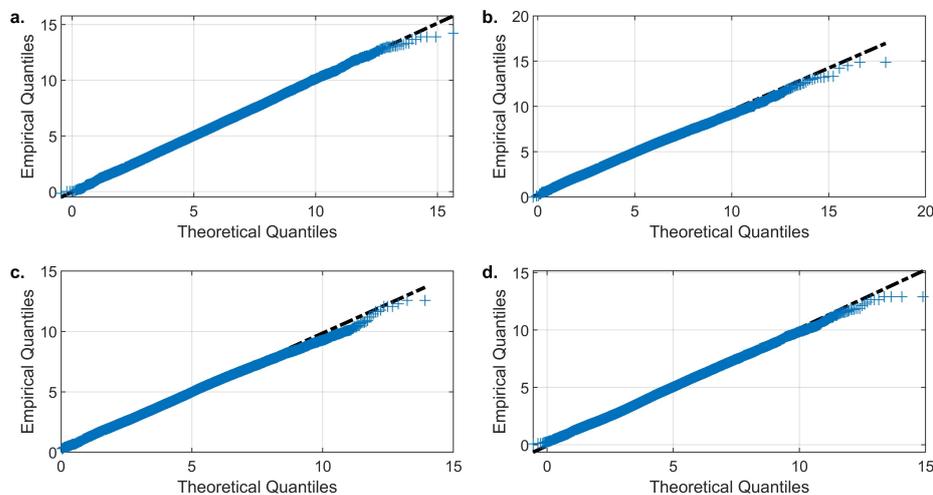
**Table 7.** Comparison of historical and simulated rank cross-correlations.

	$Y_t^1 - Y_t^2$	$Y_t^1 - Y_t^3$	$Y_t^1 - Y_t^4$	$Y_t^2 - Y_t^3$	$Y_t^2 - Y_t^4$	$Y_t^3 - Y_t^4$
Historical	0.57	0.70	0.70	0.50	0.48	0.63
Synthetic	0.60	0.70	0.62	0.56	0.50	0.63

Finally, the model’s ability to reproduce the marginal distributions is presented in Figure 12 where the empirical and synthetic quantiles of the cumulative distribution function are compared. Overall, a good agreement is observed for all variables; however, small differences are found near the tails. The size of this disagreement is not atypical in hydrological research or other fields where uncertainty in modeling or observations is, in general, significant.



**Figure 11.** Scatter plot of historical and synthetic cross-dependence structure. (a) Atenco–La Grande  $\{Y_t^1, Y_t^2\}$ , (b) Atenco–San Andres  $\{Y_t^1, Y_t^3\}$ , (c) Atenco–Chapingo (DGE)  $\{Y_t^1, Y_t^4\}$  (d) La Grande–San Andres  $\{Y_t^2, Y_t^3\}$ , (e) La Grande–Chapingo (DGE)  $\{Y_t^2, Y_t^4\}$ , and (f) San Andres–Chapingo (DGE)  $\{Y_t^3, Y_t^4\}$ .

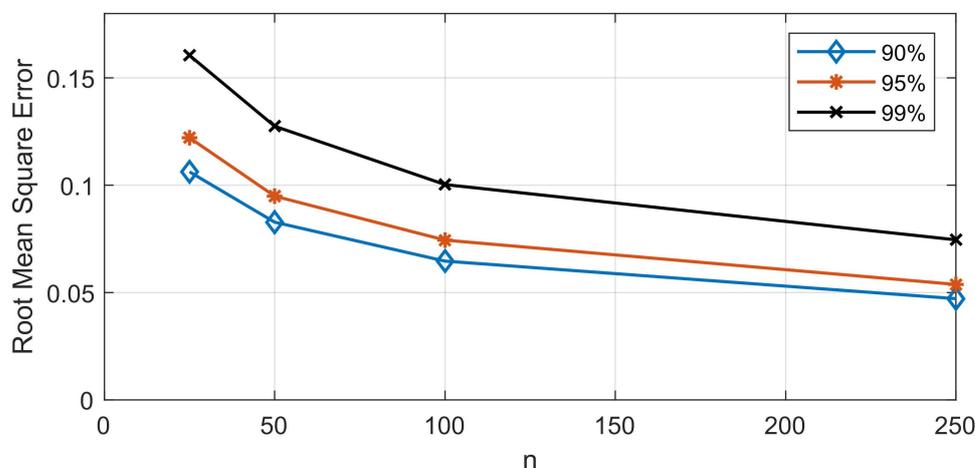


**Figure 12.** Empirical and theoretical quantiles comparison for (a) Atenco  $\{Y_t^1\}$ , (b) La Grande  $\{Y_t^2\}$ , (c) San Andres  $\{Y_t^3\}$ , and (d) Chapingo (DGE)  $\{Y_t^4\}$  stations.

### Exploration of the Repetitive Sampling Length

As with any repetitive sampling approach, the accuracy of the method depends upon the sampling length. To determine the relationship between accuracy and sampling length, four simulations were performed with  $n = [25, 50, 100, 250]$  and the empirical cumulative distributions of their RMSE (Appendix B) were computed. In Figure 13, the relationship between RMSE and sampling length for 90%, 95%, and 99% percentiles is plotted. As the sampling length increases, the RMSE decreases. The largest decrease can be found between  $n = 25$  and  $n = 50$ . For  $n = 100$ , the RMSE of the 99% percentile is equal to 0.1. The previous simulation suggests that this level of error yields adequate results. Furthermore, increasing the sampling length does not offer significant improvements in accuracy while

at the same time it drastically increases the computational cost. From this perspective, a sampling length between 50 to 100 is a reasonable choice.



**Figure 13.** RMSE as a function of sampling length (n) for 90%, 95%, and 99% quantiles.

#### 4. Conclusions

This paper presents a methodology to simulate hydroclimatic variables through copula-based models. The proposed methodologies focus on the use of vine copulas to characterize complex temporal and spatial probabilistic dependence. Two cases are presented: (i) the generation of correlated precipitation and evaporation on a daily scale and (ii) the generation of correlated daily evaporation time series from four stations. The first case consists of a trivariate vine copula to handle temporal dependence of evaporation and cross-dependence with precipitation. To capture intermittency, Markov chains were coupled with a copula-based model. In the second case, the methodology relied on repetitive sampling to couple the temporal and spatial vine copula models. The result is a parsimonious and flexible model capable of simulating accurately observations from an arbitrary number of stations.

In both case studies discussed herein, the models proved their capability to capture the underlying dependence structures, reproduce the marginal distributions, and intermittency. In fact, for the dependence structure, important asymmetries (such as tail dependence) may be incorporated. These asymmetries are often found but overlooked in the modeling of hydroclimatic variables. One of our case studies shows that by ignoring these asymmetries, the reliability of hydraulic structures, for example, could be underestimated by a factor of 2 in some cases (Table 5).

We have shown the ability of our approach combining Markov chains and vine copula models to reproduce short-range dependence (SRD) structures [63]. Many hydroclimatic processes however exhibit long-range dependence (LRD) [64–66]. The performance of our approach for these kind of dependence may be a subject for future research.

**Author Contributions:** Conceptualization, G.P., G.A.T.-A. and O.M.-N.; Formal analysis, G.P.; Methodology, G.P.; Supervision, G.A.T.-A. and O.M.-N.; writing—original draft preparation, G.P., writing—review and editing, G.P., G.A.T.-A. and O.M.-N. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was partially funded by the Eugenides Foundation.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The precipitation and evaporation timeseries analysed herein are available from the Center for Scientific Research and Higher Education at Ensenada (Centro de Investigacion Cientifica y de Educacion Superior de Ensenada—CISECE) at <http://clicom-mex.cicese.mx/> accessed on 10 September 2019.

**Acknowledgments:** The first author was partially funded by the Eugenides Foundation for this research, as part of his MSc studies scholarship.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Appendix A. Distribution Functions

### Weibull

$$F(x) = 1 - e^{-(\kappa \frac{x}{\lambda})^{1/\kappa}}$$

where  $\kappa$  is the shape parameter and  $\lambda$  is the scale parameter.

### Generalized Extreme Value

$$F(x) = \begin{cases} e^{-[1+\kappa(\frac{x}{\lambda}-\psi)]^{-\frac{1}{\kappa}}} & \text{if } \kappa \neq 0 \\ e^{-e^{(\frac{x}{\lambda}-\psi)}} & \text{if } \kappa = 0 \end{cases}$$

where  $\kappa$  is the shape parameter,  $\lambda$  is the scale parameter and  $\psi$  is the location parameter.

## Appendix B. Inverse Conditional Copulas

### Gaussian Copula

$$h(u, v, \rho) = \Phi\left(\frac{\Phi^{-1}(u) - \rho\Phi^{-1}(v)}{\sqrt{1 - \rho^2}}\right) \quad (\text{A1})$$

$$h^{-1}(u, v, \rho) = \Phi\left(\Phi^{-1}(u)\sqrt{1 - \rho^2} + \rho\Phi^{-1}(v)\right) \quad (\text{A2})$$

### Gumbel Copula

$$h(u, v, \delta) = C(u, v) \frac{1}{v} (-\log v)^{\delta-1} ((-\log u)^\delta + (-\log v)^\delta)^{1/\delta-1} \quad (\text{A3})$$

where  $C$  is the copula function  $C(u, v) = \exp[-\{(-\log u)^\delta + (-\log v)^\delta\}^{1/\delta}]$ .

In the case of the Gumbel copula, the inverse of the  $h$ -function cannot be given analytically. The use of a numerical method is necessary, such as Secant, Newton–Raphson, or even brute force. It should be noted that convergence of the solution close to the tails can be difficult and can affect the reproduction of the marginal distribution and the desired dependence structure. Moreover, accuracy in the order of at least  $10^{-4}$  is required for reliable simulation results. This task can be computationally very expensive.

## References

1. Fiering, M.B. *Streamflow Synthesis*; Harvard University Press: Cambridge, MA, USA, 1967. [CrossRef]
2. Klemeš, V.; Srikanthan, R.; McMahon, T.A. Long-memory flow models in reservoir analysis: What is their practical value? *Water Resour. Res.* **1981**, *17*, 737–751. [CrossRef]
3. Vogel, R.M.; Stedinger, J.R. The value of stochastic streamflow models in overyear reservoir design applications. *Water Resour. Res.* **1988**, *24*, 1483–1490. [CrossRef]
4. Tsoukalas, I.; Makropoulos, C. A Surrogate Based Optimization Approach for the Development of Uncertainty-Aware Reservoir Operational Rules: The Case of Nestos Hydrosystem. *Water Resour. Manag.* **2015**, *29*, 4719–4734. [CrossRef]
5. Tsoukalas, I.; Makropoulos, C. Multiobjective optimisation on a budget: Exploring surrogate modelling for robust multi-reservoir rules generation under hydrological uncertainty. *Environ. Model. Softw.* **2015**. [CrossRef]

6. Papoulakos, K.; Pollakis, G.; Moustakis, Y.; Markopoulos, A.; Iliopoulou, T.; Dimitriadis, P.; Koutsoyiannis, D.; Efstratiadis, A. Simulation of water-energy fluxes through small-scale reservoir systems under limited data availability. *Energy Procedia* **2017**, *125*, 405–414. [[CrossRef](#)]
7. Koutsoyiannis, D.; Economou, A. Evaluation of the parameterization-simulation-optimization approach for the control of reservoir systems. *Water Resour. Res.* **2003**, *39*. [[CrossRef](#)]
8. Koskinas, A.; Tegos, A. StEMORS: A Stochastic Eco-Hydrological Model for Optimal Reservoir Sizing. *Open Water J.* **2020**, *6*, 1.
9. Haberlandt, U.; Hundecha, Y.; Pahlow, M.; Schumann, A.H. Rainfall Generators for Application in Flood Studies. In *Flood Risk Assessment and Management: How to Specify Hydrological Loads, Their Consequences and Uncertainties*; Schumann, A.H., Ed.; Springer: Dordrecht, The Netherlands, 2011; pp. 117–147.
10. Papalexiou, S.M.; Koutsoyiannis, D. A global survey on the seasonal variation of the marginal distribution of daily precipitation. *Adv. Water Resour.* **2016**, *94*, 131–145. [[CrossRef](#)]
11. Kossieris, P. Exploring the Statistical and Distributional Properties of Residential Water Demand at Fine Time Scales. *Water* **2018**, *10*, 1481. [[CrossRef](#)]
12. Thomas, H.A.; Fiering, M. Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. In *Design of Water Resources-Systems*; Harvard University Press: Cambridge, UK, 1962; pp. 459–493.
13. Lombardo, F.; Volpi, E.; Koutsoyiannis, D.; Papalexiou, S.M. Just two moments! A cautionary note against use of high-order moments in multifractal models in hydrology. *Hydrol. Earth Syst. Sci.* **2014**, *18*, 243–255. [[CrossRef](#)]
14. Koutsoyiannis, D. Knowable moments for high-order stochastic characterization and modelling of hydrological processes. *Hydrol. Sci. J.* **2019**. [[CrossRef](#)]
15. Koutsoyiannis, D.; Manetas, A. Simple Disaggregation by Accurate Adjusting Procedures. *Water Resour. Res.* **1996**, *32*, 2105–2117. [[CrossRef](#)]
16. Tsoukalas, I.; Papalexiou, S.M.; Efstratiadis, A.; Makropoulos, C. A Cautionary Note on the Reproduction of Dependencies through Linear Stochastic Models with Non-Gaussian White Noise. *Water* **2018**, *10*, 771. [[CrossRef](#)]
17. Beran, J. *Statistics for Long-Memory Processes*; Routledge: London, UK, 1994. [[CrossRef](#)]
18. Lloyd, E.H.; Hurst, H.E.; Black, R.P.; Simaika, Y.M. Long-Term Storage: An Experimental Study. *J. R. Stat. Society. Ser. A Gen.* **1966**, *129*, 591. [[CrossRef](#)]
19. Mandelbrot, B.B. Une classe de processus stochastiques homothétiques a soi: application à la loi elimatoloieque de H. E. Hurst. *C. R. Acad. Sci.* **1965**, *260*, 3274–3277.
20. Koutsoyiannis, D. The Hurst phenomenon and fractional Gaussian noise made easy. *Hydrol. Sci. J.* **2002**, *47*, 573–595. [[CrossRef](#)]
21. Dimitriadis, P.; Koutsoyiannis, D.; Iliopoulou, T.; Papanicolaou, P. A Global-Scale Investigation of Stochastic Similarities in Marginal Distribution and Dependence Structure of Key Hydrological-Cycle Processes. *Hydrology* **2021**, *8*, 59. [[CrossRef](#)]
22. Koutsoyiannis, D. A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series. *Water Resour. Res.* **2000**, *36*, 1519–1533. [[CrossRef](#)]
23. Efstratiadis, A.; Dialynas, Y.G.; Kozanis, S.; Koutsoyiannis, D. A multivariate stochastic model for the generation of synthetic time series at multiple time scales reproducing long-term persistence. *Environ. Model. Softw.* **2014**, *62*, 139–152. [[CrossRef](#)]
24. Dupuis, D.J. Using Copulas in Hydrology: Benefits, Cautions, and Issues. *J. Hydrol. Eng.* **2007**, *12*, 381–393. [[CrossRef](#)]
25. Genest, C.; Favre, A.C. Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. *J. Hydrol. Eng.* **2007**, *12*, 347–368. [[CrossRef](#)]
26. Salvadori, G.; Michele, C.D. On the Use of Copulas in Hydrology: Theory and Practice. *J. Hydrol. Eng.* **2007**, *12*, 369–380. [[CrossRef](#)]
27. Daneshkhah, A.; Remesan, R.; Chatrabgoun, O.; Holman, I.P. Probabilistic modeling of flood characterizations with parametric and minimum information pair-copula model. *J. Hydrol.* **2016**, *540*, 469–487. [[CrossRef](#)]
28. Lu Chen, S.G. *Copulas and Its Application in Hydrology and Water Resources*, 1st ed.; Springer: Singapore, 2019; p. 290. [[CrossRef](#)]
29. Zhang, L.; Singh, V.P. *Copulas and Their Applications in Water Resources Engineering*; Cambridge University Press: Cambridge, UK, 2019. [[CrossRef](#)]
30. Lee, T.; Salas, J.D. Copula-based stochastic simulation of hydrological data applied to Nile River flows. *Hydrol. Res.* **2011**, *42*, 318–330. [[CrossRef](#)]
31. Jeong, C.; Lee, T. Copula-based modeling and stochastic simulation of seasonal intermittent streamflows for arid regions. *J. Hydro-Environ. Res.* **2015**, *9*, 604–613. [[CrossRef](#)]
32. Serinaldi, F. A multisite daily rainfall generator driven by bivariate copula-based mixed distributions. *J. Geophys. Res. Atmos.* **2009**. [[CrossRef](#)]
33. Leontaris, G.; Morales-Nápoles, O.; Wolfert, A. Probabilistic scheduling of offshore operations using copula based environmental time series—An application for cable installation management for offshore wind farms. *Ocean Eng.* **2016**, *125*, 328–341. [[CrossRef](#)]
34. Bedford, T.; Daneshkhah, A.; Wilson, K.J. Approximate Uncertainty Modeling in Risk Analysis with Vine Copulas. *Risk Anal.* **2016**, *36*, 792–815. [[CrossRef](#)]
35. Chatrabgoun, O.; Hosseini-Far, A.; Chang, V.; Stocks, N.G.; Daneshkhah, A. Approximating non-Gaussian Bayesian networks using minimum information vine model with applications in financial modelling. *J. Comput. Sci.* **2018**, *24*, 266–276. [[CrossRef](#)]
36. Cooke, R.M. Markov and entropy properties of tree and vines-dependent variables. In Proceedings of the ASA Section of Bayesian Statistical Science, Istanbul, Turkey, 16–18 August 1997.

37. 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]
38. Jäger, W.S.; Nápoles, O.M. A Vine-Copula Model for Time Series of Significant Wave Heights and Mean Zero-Crossing Periods in the North Sea. *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part Civ. Eng.* **2017**. [CrossRef]
39. Sarmiento, C.; Valencia, C.; Akhavan-Tabatabaei, R. Copula autoregressive methodology for the simulation of wind speed and direction time series. *J. Wind. Eng. Ind. Aerodyn.* **2018**. [CrossRef]
40. Brechmann, E.C.; Czado, C. COPAR—multivariate time series modeling using the copula autoregressive model. *Appl. Stoch. Model. Bus. Ind.* **2015**, *31*, 495–514. [CrossRef]
41. Torres-Alves, G.A.; Morales-Nápoles, O. Reliability Analysis of Flood Defenses: The Case of the Nezahualcoyotl Dike in the Aztec City of Tenochtitlan. *Reliab. Eng. Syst. Saf.* **2020**, 107057. [CrossRef]
42. Sklar, A. Fonctions de répartition à  $n$  dimensions et leurs marges. *Publ. L'Institut De Stat. L'Université Paris* **1959**, *8*, 229–231.
43. Joe, H. *Dependence Modeling with Copulas*; Chapman & Hall/CRC Monographs on Statistics & Applied Probability; Taylor and Francis: Hoboken, NJ, USA, 2014.
44. Joe, H. Families of  $m$ -variate distributions with given margins and  $m(m - 1)/2$  bivariate dependence parameters. In *Distributions with Fixed Marginals and Related Topics*; Rüschendorf, L., Schweizer, B., Taylor, M.D., Eds.; Lecture Notes–Monograph Series; Institute of Mathematical Statistics: Hayward, CA, USA, 1996; Volume 28, pp. 120–141. [CrossRef]
45. Aas, K.; Czado, C.; Frigessi, A.; Bakken, H. Pair-copula constructions of multiple dependence. *Insur. Math. Econ.* **2009**. [CrossRef]
46. Morales-Nápoles, O.; Steenbergen, R.D.J.M. Large-Scale Hybrid Bayesian Network for Traffic Load Modeling from Weigh-in-Motion System Data. *J. Bridge Eng.* **2015**, *20*, 04014059. [CrossRef]
47. Morales-Nápoles, O.; Steenbergen, R.D. Analysis of axle and vehicle load properties through Bayesian Networks based on Weigh-in-Motion data. *Reliab. Eng. Syst. Saf.* **2014**, *125*, 153–164. [CrossRef]
48. Bedford, T.; Cooke, R.M. Vines—A new graphical model for dependent random variables. *Ann. Statist.* **2002**, *30*, 1031–1068. [CrossRef]
49. Cooke, R.M.; Joe, H.; Aas, K. Vines arise. In *Dependence Modeling: Vine Copula Handbook*; World Scientific: Singapore, 2010. [CrossRef]
50. Papalexiou, S.M. Unified theory for stochastic modelling of hydroclimatic processes: Preserving marginal distributions, correlation structures, and intermittency. *Adv. Water Resour.* **2018**, *115*, 234–252. [CrossRef]
51. Tsoukalas, I.; Makropoulos, C.; Koutsoyiannis, D. Simulation of Stochastic Processes Exhibiting Any-Range Dependence and Arbitrary Marginal Distributions. *Water Resour. Res.* **2018**, *54*, 9484–9513. [CrossRef]
52. Koutsoyiannis, D. Coupling stochastic models of different timescales. *Water Resour. Res.* **2001**, *37*, 379–391. [CrossRef]
53. Tsoukalas, I.; Efstratiadis, A.; Makropoulos, C. Building a puzzle to solve a riddle: A multi-scale disaggregation approach for multivariate stochastic processes with any marginal distribution and correlation structure. *J. Hydrol.* **2019**, *575*, 354–380. [CrossRef]
54. Kurowicka, D.; Cooke, R. Sampling algorithms for generating joint uniform distributions using the vine-copula method. *Comput. Stat. Data Anal.* **2007**, *51*, 2889–2906. [CrossRef]
55. Coles, S. *An Introduction to Statistical Modeling of Extreme Values*; Springer: London, UK, 2001.
56. Hosking, J.R.M. L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. (English). *J. R. Stat. Soc. Ser. B Methodol.* **1990**, *52*, 105–124. [CrossRef]
57. Zorzetto, E.; Botter, G.; Marani, M. On the emergence of rainfall extremes from ordinary events. *Geophys. Res. Lett.* **2016**, *43*, 8076–8082. [CrossRef]
58. Genest, C.; Rémillard, B.; Beaudoin, D. Goodness-of-fit tests for copulas: A review and a power study. *Insur. Math. Econ.* **2009**, *44*, 199–213. [CrossRef]
59. Choroś, B.; Ibragimov, R.; Permiakova, E. Copula Estimation. In *Copula Theory and Its Applications*; Jaworski, P., Durante, F., Härdle, W.K., Rychlik, T., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 77–91.
60. CLICOM. CISECE-Centro de Investigacion Cientifica y de Educacion Superior de Ensenada. Available online: <http://clicom-mex.cicese.mx/> (accessed on 10 September 2019).
61. Savenije, H.H.G. HESS Opinions “Topography driven conceptual modelling (FLEX-Topo)”. *Hydrol. Earth Syst. Sci.* **2010**, *14*, 2681–2692. [CrossRef]
62. Tsoukalas, I.; Efstratiadis, A.; Makropoulos, C. Stochastic Periodic Autoregressive to Anything (SPARTA): Modeling and Simulation of Cyclostationary Processes With Arbitrary Marginal Distributions. *Water Resour. Res.* **2018**, *54*, 161–185. [CrossRef]
63. Ibragimov, R.; Lentzas, G. Copulas and long memory. *Probab. Surv.* **2017**, *14*, 289–327. [CrossRef]
64. Koutsoyiannis, D. HESS Opinions “A random walk on water”. *Hydrol. Earth Syst. Sci.* **2010**, *14*, 585–601. [CrossRef]
65. Koutsoyiannis, D. Climate change, the Hurst phenomenon, and hydrological statistics. *Hydrol. Sci. J.* **2003**, *48*, 3–24. [CrossRef]
66. Markonis, Y.; Moustakis, Y.; Nasika, C.; Sychova, P.; Dimitriadis, P.; Hanel, M.; Máca, P.; Papalexiou, S. Global estimation of long-term persistence in annual river runoff. *Adv. Water Resour.* **2018**, *113*, 1–12. [CrossRef]