Journal of Hydrology 598 (2021) 126226

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

A GLM copula approach for multisite annual streamflow generation

Victor Costa Porto [a,*], Francisco de Assis de Souza Filho [a], Taís Maria Nunes Carvalho [a], Ticiana Marinho de Carvalho Studart [a], Maria Manuela Portela [b]

[a] Department of Hydraulic and Environmental Engineering, Federal University of Ceará, Fortaleza, CE 60451-970, Brazil
[b] Civil Engineering Research and Innovation for Sustainability (CERIS), Instituto Superior Tecnico (IST), University of Lisbon (UL), Portugal

ARTICLE INFO

This manuscript was handled by A. Bardossy, Editor-in-Chief

Keywords:
Generalized linear models
Copula
Multi-site stochastic streamflow simulation

ABSTRACT

The research presents a multisite annual streamflow generation model that combines the Generalized Linear Model (GLM), for determining the temporal structure, with copulas, for modelling the spatial dependence joint distributions. The performance of the GLM-Copula model was verified by comparing its ability to preserve historical features and simulate drought events with the multivariate auto regressive moving average (ARMA) model and the copula autoregressive (COPAR) model. The statistical measures adopted for the models' performance evaluation include summary statistics (mean, standard deviation, maximum, minimum and skewness coefficient), temporal and spatial correlation, simulation of drought conditions (maximum number of years under drought condition) and copula entropy as a nonlinear measure of total association. The combined GLM-Copula model's main advantages are that (i) it does not require data normalization; ii) it allows the modelling of the dependence structures with different probability functions; and (iii) it is capable of representing non-conventional parsimonious autocorrelation functions. The ability of the GLM-Copula approach to preserve the summary statistics from the historical data was similar to both benchmark models. However, the GLM-Copula was considerably better in reproducing the longest drought duration that was underestimated by the ARMA model and was better in reproducing the copula entropy than both benchmark models. The approach is proposed in its simplest form but can be easily upgraded by combining GLMs with numerical data or extended to predict future streamflow with the incorporation of exogenous climate variables that affect streamflow. The proposed model may be useful in future studies/applications where data normalization jeopardizes the replication of data or/and in drought dependent stochastic applications, like the definition of optimal operation rules of a perennial reservoir system or long-term hydropower dispatch.

1. Introduction

Synthetic streamflow time series generation has an important role in water resources planning and management. It is applied to the design of reservoir systems and to the definition of their optimal operation rules, to drought evaluation and to several other studies with a stochastic nature (McMahon et al., 2006; Rajagopalan et al., 2010; Salas and Lee, 2010). For a correct application, the generated synthetic series must preserve key historical data characteristics, such as statistical moments and dependence structure (e.g. auto and cross-correlation) (Zachariah and Reddy, 2013).

The classical methods for streamflow simulation, such as the ARMA models (Box and Jenkins, 1976), are based on rigid assumptions about the variables' dependence and require them to follow a Gaussian distribution (Sharma and O'Neill, 2002; Prairie et al., 2006). However, some hydrological variables are significantly skewed which requires their normalization, i.e. their transformation into alternatives variables that satisfy those models' assumptions (Salas et al., 1980; Salas, 1993). Most of those models' drawbacks arise from their rigid assumptions (e.g. the Gaussian) and from the limitations of the data transformation techniques resulting in a lack of flexibility that may influence the preservation of the historical characteristics (Sharma et al., 1997; Prairie et al., 2006; Rajagopalan et al., 2010; Hao and Singh, 2011; Lee and Salas, 2011; Pereira et al., 2017).

The stochastic streamflow simulation literature presents several non-Gaussian modeling alternatives. The most famous are the Lag-1 Gamma Autoregressive model (GAR-1) (Fernandez and Salas, 1990) and the nonparametric approaches such as the K-Nearest Neighbor method (KNN) (Lall and Sharma, 1996), and the Kernel Density Estimators (KDE) (Sharma et al., 1997; Sharma and O'Neill, 2002). However, these

* Corresponding author.
E-mail address: victorcporto@gmail.com (V.C. Porto).

https://doi.org/10.1016/j.jhydrol.2021.126226
Received 14 December 2020; Received in revised form 11 March 2021; Accepted 16 March 2021
Available online 23 March 2021

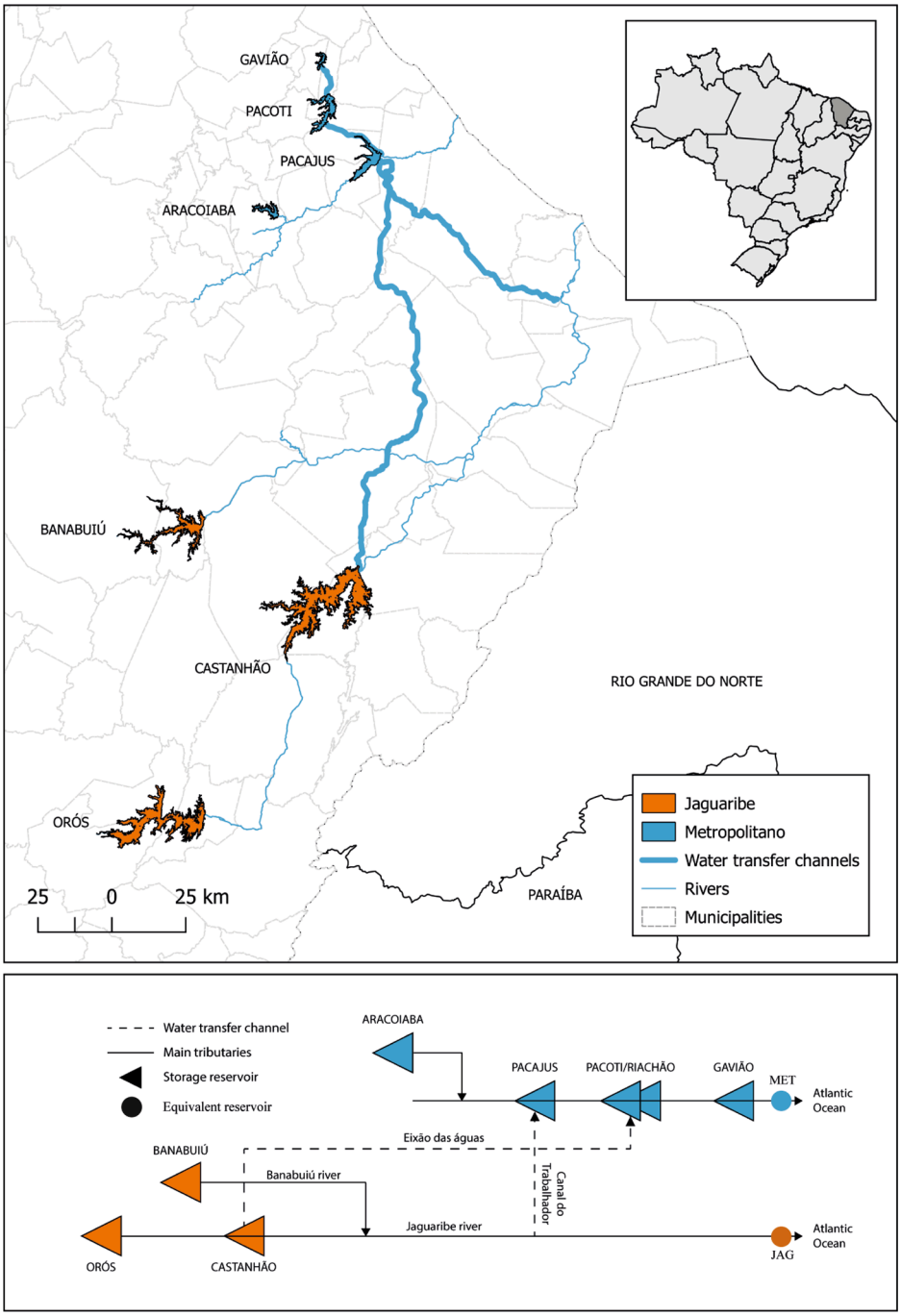

Fig. 1. The Jaguaribe-Metropolitano reservoir system in Ceará, Brazil.

alternative models have their own limitations. The GAR-1 also lacks flexibility and cannot model long term persistency. Furthermore, resampling methods, like the KNN, reproduce only the observed values and the KDE may not be applied to higher dimensions (Rajagopalan et al., 2010; Lee and Salas, 2011).

Recently, copula based approaches have been applied to hydrologic time series generation (Lee and Salas, 2011; Zachariah and Reddy, 2013; Chen et al., 2015; Pereira et al., 2017). The copula methods are parametric approaches that model the dependence structure apart from the marginal distributions, which provides high flexibility by allowing the use of any marginal distributions. Lee and Salas (2011) compared the performance of a copula and an ARMA model applied to single site annual streamflow generation and showed the former had some benefits, if small.

Generalized Linear Models (GLMs), introduced by Nelder and Wedderburn (1972) as an extension of the classical linear regression model, are parsimonious parametric methods that allow the modelling of non-Gaussian variables (McCullagh and Nelder, 1989). As stated by Rajagopalan et al. (2010), GLM approaches may be reasonable alternatives to the traditional parametric methods due to their flexibility and capability to preserve different features of the historical series.

The use of GLMs is recognized in hydrology for stochastic generation of daily weather variables, like precipitation, temperature and potential evapotranspiration (Chandler and Wheater, 2002; Chandler, 2005; Yang

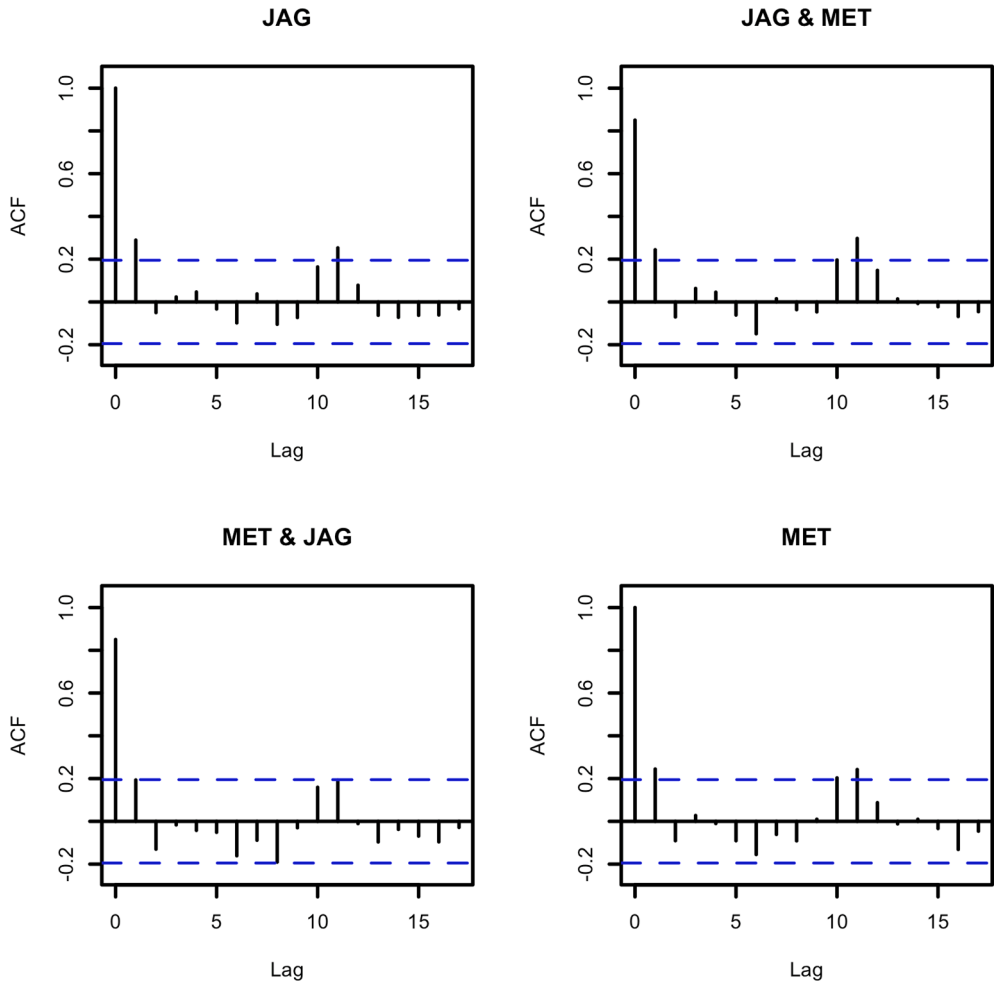

Fig. 2. Temporal autocorrelation (top-left and bottom-right) and spatial cross-correlation (top-right and bottom-left) functions of the annual streamflow series of the JAG and MET equivalent reservoirs. The dashed lines represent the 95% confidence intervals.

et al., 2005; Wheater et al., 2005; Furrer and Katz, 2007; Kleiber et al., 2012; Verdin et al., 2014). However, to the knowledge of the authors, GLMs have not yet been applied for streamflow stochastic generation.

Thus, the first part of this research addresses the applicability of GLM to generate single site annual streamflow and compares its ability to model temporal dependence and preserve historical statistics against a traditional univariate autoregressive (AR) method.

For multisite time series generation, GLM approaches require the specification of the joint probability distributions of the time series which is obtained from a spatial dependence modelling that respects the marginal distributions (Yang et al., 2005). However, modelling the spatial structure is a complex process that is often done in the GLM-based weather generators by multivariate Gaussian assumptions that may need data normalization, reducing the approach's flexibility (Yang et al., 2005; Kleiber et al., 2012; Verdin et al., 2014).

Meanwhile, high dimensional copulas ($d \geq 3$) lose their flexibility to represent the dependence structures as there is a limited set of higher dimensional copula families (Kao and Govindaraju, 2008; Aas et al., 2009; Hao and Singh, 2013). To overcome this limitation, recent copula times series models are mostly built from two approaches: i) vine copulas that decomposes the multidimensional problem into a sequence of bidimensional copulas (Brechmann and Czado, 2015; Pereira et al., 2017; Wang et al., 2019) or ii) maximum entropy copula that based on the concept of maximum entropy distribution from the information theory can fit a flexible high dimensional copula (Hao and Singh, 2013, 2015; Singh and Zhang, 2018;). However, the complexity and the computational burden grows quickly with the dimension for both entropy and vine copula models (Hao and Singh, 2015; Pereira et al., 2017).

The second part of this research presents a multisite annual streamflow generation model that couples GLM and copula, the first to represent the temporal structure and the second, to model the spatial dependence (i.e. the joint distributions). Its performance to reproduce historical statistics and dependence structures is compared with the traditional multivariate ARMA model and the copula autoregressive (COPAR) model (Brechmann and Czado, 2015), a state-of-art copula time series model that extends the vine copula concept to model both spatial and temporal dependence.

The proposed model exploits both methods' flexibility and synergy: the copula provides a flexible way for estimating the joint distribution that GLM needs for multisite generation; while the GLM lowers the problem's dimension and allows copula models to be applied without the need for normalized data.

Despite the existence of several stochastic multisite streamflow time series generation methods, none is universally accepted (Srinivas and Srinivasan, 2005; Chen et al., 2015; Hao and Singh, 2016). In contrast, the proposed model allows us to model both spatial and temporal dependencies without normalization, is computationally efficient and can

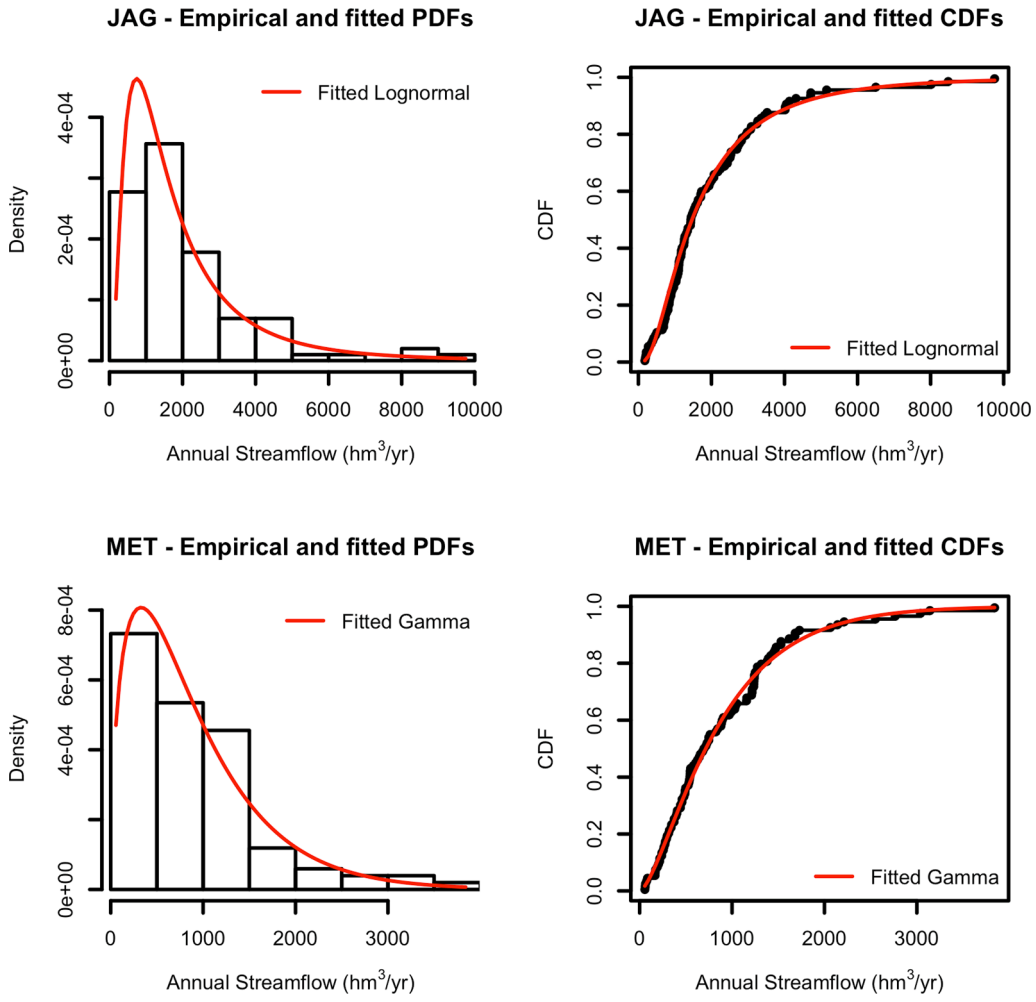

Fig. 3. Empirical and fitted probability density (PDF) and cumulative distribution (CDF) functions of the annual streamflow series at the JAG (top) and MET (bottom) equivalent reservoirs. Note that preferably Lognormal fits JAG and Gamma fits MET.

be used as a dimensional reduction, for vine and maximum entropy copula practitioners, which we suggest justifies its addition to the time series generation toolbox.

2. Case study, data and preliminary analysis

In this paper, the Jaguaribe-Metropolitano reservoir system in Ceará State (Brazil), represented in Fig. 1, was selected as the multi-site synthetic flow series generation case study, since the annual inflows at its seven reservoirs are highly variable and skewed (Souza Filho and Lall, 2003). The annual inflow data to those reservoirs are available for the 1912–2012 period (101 years) from Barros et al. (2013).

The Jaguaribe-Metropolitano is the State's greatest water impounding system and the most important water supply source. It comprises two different basins, the Jaguaribe and the Metropolitano basins, with 72,000 and 15,200 km2 respectively, that were artificially connected by channels allowing the latter to receive water from the former. Most of the system is located in a semiarid region with a highly variable annual streamflow, due to the temporal variability of the precipitation and the predominance of shallow soils (da Silva et al., 2017).

The Jaguaribe basin covers approximately 48% of the State of Ceará and its main water use is irrigation which accounts for about 90% of the state agricultural production. Although the Metropolitano basin is smaller, comprising approximately 10% of the State area, it has a larger population and supplies water to the capital city (Fortaleza) and to its metropolitan region for domestic supply, industry, and tourism. The water demand in the Metropolitano basin is almost uniform throughout the year while the one in Jaguaribe basin is concentrated in the second semester due to the crop irrigation period (i.e. the dry station) (Souza Filho and Lall, 2003; da Silva et al., 2017).

The system is composed of seven major reservoirs (from upstream to downstream): Orós, Castanhão, Banabuiú, Aracoiaba, Pacajus, Pacoti and Gavião. The first three are in the Jaguaribe basin and the last four are in the Metropolitano basin (total storage capacity of 10,240 and 871 hm^3, respectively) (Fig. 1). For water resources planning and management purposes, the Jaguaribe-Metropolitano system can be represented as two equivalent reservoirs, one for each basin, located at the most downstream sections of the main rivers of Jaguaribe and the Metropolitano basins, as illustrated in Fig. 1. In this research, the annual inflows to the Jaguaribe (JAG) and Metropolitano (MET) equivalent reservoirs were obtained by summing the annual inflows to each of their major component reservoirs.

The temporal correlation and the spatial cross-correlation of the annual inflows series thus obtained for JAG and MET are characterized in Fig. 2 for lags 0 until 17.

In spatial terms, there is a high lag-0 correlation (>0.8), as shown by the cross-correlation function, because the rainfall regime in both basins mainly relies on the same climatic process: the Intertropical Convergence Zone (ITCZ) displacement (Moura and Shukla, 1981; Andreoli and Kayano, 2004; Wang et al., 2004).

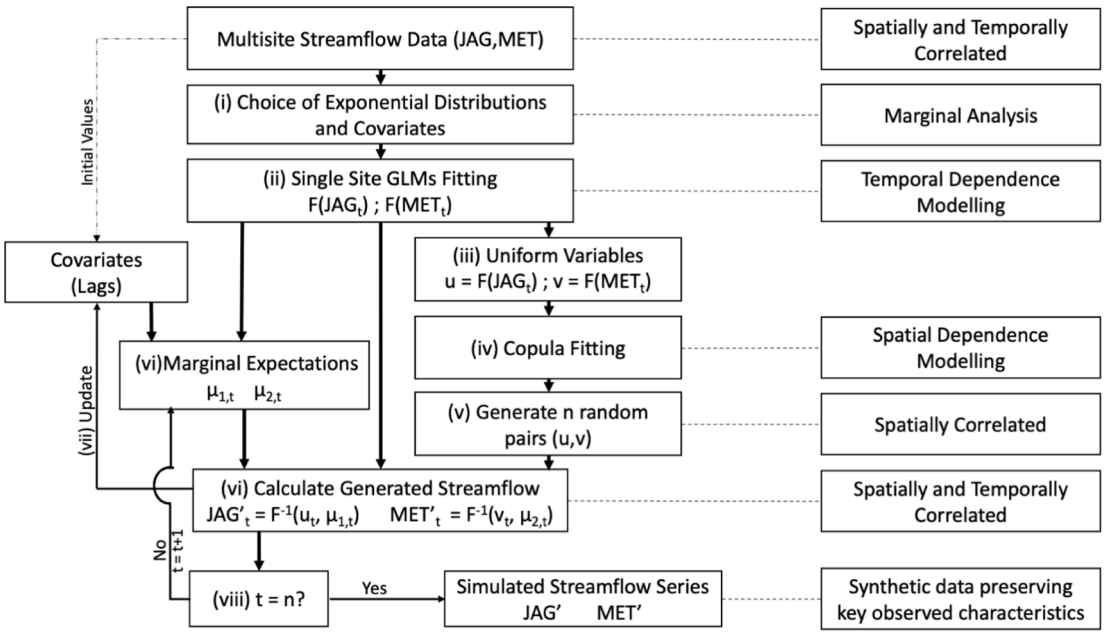

Fig. 4. Flowchart of the GLM-Copula annual streamflow generation procedure.

In temporal terms, both series present a short-term persistency pattern with a fast correlogram decay after the first lag. There is also a long-term dependence pattern with significant positive correlation coefficients for lags 10 and 11. The short memory pattern may be a result of low groundwater flow, since both basins are situated in a crystalline Precambrian basement with shallow soils and poor vegetation cover (Frischkorn et al., 2003; Alexandre et al., 2005; Barros et al., 2013). The long-term persistency may be related to a decadal sea surface temperature variability in the Tropical Atlantic that influences the ITCZ location (Andreoli and Kayano, 2004; Andreoli and Kayano, 2006).

The empirical and fitted cumulative distribution functions (CDF) and probability density functions (PDF) of the annual flows at each equivalent reservoir are shown in Fig. 3, showing convincing similarity both series are non-Gaussian and right-skewed. The inflows to JAG are close to a lognormal distribution and those to MET, to a gamma distribution. The selection of the best-fit distribution was based in the Anderson-Darling test for the maximum likelihood estimated parameters and the fit of the CDFs is remarkably good throughout the range of data.

3. Background and methodology

3.1. Generalized linear models

The classical linear regression model is defined as:

$$y_i = \alpha + \beta x_i + \varepsilon_i \varepsilon_i \; N(0, \sigma^2) i = 1, \cdots, n \tag{1}$$

where, y_i is each value of the response variable, α is the intercept, β is a vector of parameters, x_i is the vector of predictors and ε is a normally distributed error (Salas et al., 1980).

Equation (1) can be rewritten in the following form (Fahrmeir and Tutz, 2001):

$$y_i \; N(\mu_i, \sigma^2) \eta_i = \alpha + \beta x_i \eta_i = g(\mu_i) i = 1, \cdots, n \tag{2}$$

where, μ_i is the expectation of y_i, η_i is a linear predictor, $g(\cdot)$ is the function that links the expectation of the response variable with the predictors (i.e. a link function). The link function is equal to the identity in the model described by Eq. (1).

The model described by Eq. (2) can be extended to a more general case (GLMs) with the assumption that each y_i has a distribution in the exponential family with expectation $E(y_i | x_i) = \mu_i$, a common dispersion parameter ϕ independent of i and function of the response variable variance (Fahrmeir and Tutz, 2001). The density function of these distributions is:

$$f(y_i|\theta_i, \phi) = \exp\left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\} \tag{3}$$

where, θ_i is the natural parameter (dependent on μ_i), $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$ are specific functions related to the type of exponential family.

The exponential family comprises some famous continuous (e.g. Normal, Lognormal and Gamma) and discrete (e.g. Poisson and Bernoulli) distributions and the link between its expectation and the linear predictors may be represented by any monotonic differentiable function (McCullagh and Nelder, 1989). Hence, the flexibility of GLMs to model different types of data (e.g. continuous, discrete and categorical) and Gaussian and non-Gaussian patterns.

Although the GLMs were proposed to model independent variables, they can be extended to time series with lags as covariates (Fahrmeir and Tutz, 2001; Chandler, 2005). A more detailed description of the GLMs theory is in McCullagh and Nelder (1989) and Fahrmeir and Tutz (2001).

3.2. Bivariate copulas

Copulas are parametric functions that are able to combine marginal distributions into a multivariate distribution function. The copula concept allows flexibility to choose the univariate marginal distributions due to its dependence structure that sits within alternative variables that are uniform in the unit square and correspond to the values of the univariate cumulative distributions (Nelsen, 2006).

According to Sklar's theorem (Sklar, 1959), a bivariate distribution function $F(x, y)$ of two correlated random variables X and Y with respective marginal cumulative distributions $F(x)$ and $F(y)$, can be defined as a copula C:

$$F(x, y) = C(F(x), F(y)) = C(u, v) \tag{4}$$

where u and v are uniform and defined in the [0,1] interval and refer

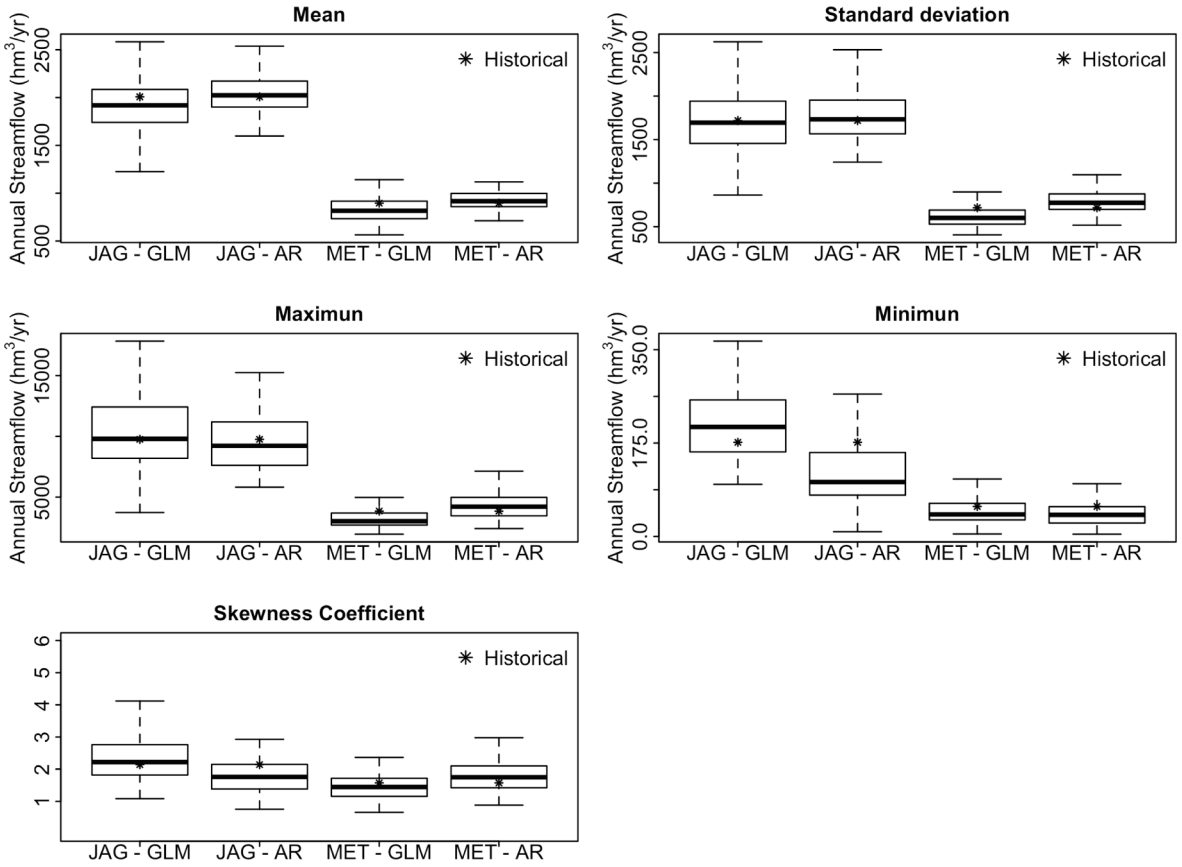

Fig. 5. Comparison, for both equivalent reservoirs (JAG and MET), of some of the annual statistics of the historical series and of the synthetic series obtained by the GLM and AR models. The box ranges from the first to the third quartile and the whiskers have maximum length of 1.5 × IQR (interquartile range).

to the values of *F(x)* and *F(y)*, respectively.

Besides their marginal distribution flexibility, copulas can capture non-linear dependence features, their parameters can be estimated by maximum likelihood and there is a wide range of copula families (e.g. Normal, Frank, Gumbel, Clayton) which allows versatility in the dependence structure modelling as well. Some copula families' descriptions, their formulation and parameter estimation methods can be found in Joe (1997), Nelsen (2006) and Joe (2014).

There are copulas defined for more than two variables; however higher dimension copulas are simply one parameter constructs and result in loss of flexibility and rigid dependence assumptions. It is better to model the joint distributions as a sequence of bivariate copulas (i.e. the vine copula method). Some are applied to pairs of univariate margins and others applied to pairs of univariate conditional distributions (Aas et al., 2009; Joe, 2014). Still, the number of parameters grows exponentially with the number of variables in the vine copula approach. Fortunately, we are only dealing with a pair of time series.

3.3. GLM single site streamflow simulation

The annual streamflow time series for each site is modelled as a univariate GLM with constant variance and the annual streamflow lags with significant correlations (1st,10th,11th) as covariates (Fig. 2). The inflows to JAG were sampled from a Lognormal distribution with identity link and those to MET, from a Gamma with log link (Fig. 3). These models can be described as:

$$f(JAG_t)\,Lognormal\left(\mu_{1,t},\sigma_1^2\right)g(\mu_{1,t})=\beta_{1,0}+\beta_{1,1}JAG_{t-1}+\beta_{1,2}JAG_{t-10}+\beta_{1,3}JAG_{t-11} \quad (5)$$

$$f(MET_t)\,Gamma\left(\mu_{2,t},\sigma_2^2\right)\mathrm{h}(\mu_{2,t})=\beta_{2,0}+\beta_{2,1}MET_{t-1}+\beta_{2,2}MET_{t-10}+\beta_{2,3}MET_{t-11} \quad (6)$$

where JAG_t and MET_t are each equivalent reservoir time series, t is the time, $\mu_{1,t}$ and $\mu_{2,t}$ are each series expected values for time t, $g(\cdot)$ and $h(\cdot)$ are the link functions identity and log respectively, σ_1^2 and σ_2^2 are the series variance and $\beta_{1,i}$ and $\beta_{2,i}$ ($i = 1,2,\ldots,$number of covariates + 1) are each series GLM parameters.

Maximum likelihood GLM parameter estimates are obtained using iterative weighted least squares (McCullagh and Nelder, 1989). This procedure was carried out using the 'base' stats package from the R programming language (R Core Team, 2013).

3.4. Copula GLM multisite streamflow simulation

The joint distribution of both sites' times series is modelled as a bivariate copula:

$$F(JAG_t \le jag_t, MET_t \le met_t) = C(F(JAG_t), F(MET_t)) = C(u,v) \quad (7)$$

This model assumes that the spatial relation between inflows to JAG and MET is temporally stationary. Also, u and v, the marginal's CDFs values, are random variables uniformly distributed between zero and one.

To obtain the random values (u,v), one of the variables may be sampled from the uniform distribution, while the other from the conditional bivariate copula distribution that can be defined as a function of the joint distribution:

$$F(v|u) = C(v|u) = \frac{\partial C(u,v)}{\partial u} \quad (8)$$

The selection of the copula family and parameters estimation was

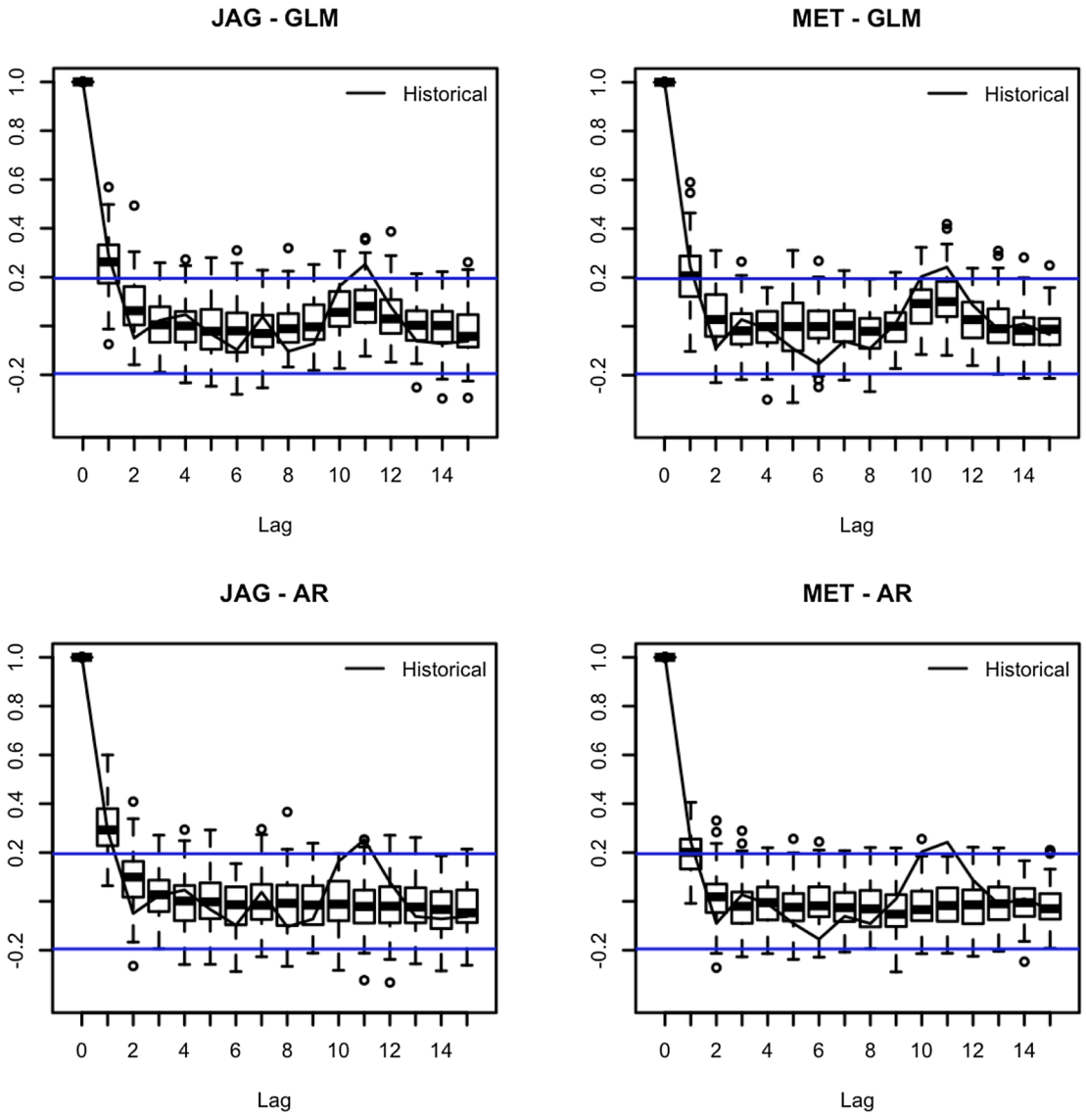

Fig. 6. Comparison, for both equivalent reservoirs (JAG and MET) of the annual autocorrelation function of the historical series and of the univariate synthetic series obtained by the GLM (top) and AR (bottom) models. The boxes range from the first to the third quartile and the whiskers have maximum length of 1.5 × IQR (interquartile range).

done with the 'VineCopula' R-package (Schepsmeier et al., 2018). The package estimates the parameters for different copula families, using the Maximum Likelihood method and then selects the family with the lowest Akaike Information Criterion (AIC); it also verifies the performance of the fit by the reproduction of the Kendall τ correlation coefficient.

Also, a verification of the copula's tail asymmetry is carried out with the lower (ϱ_L) and upper (ϱ_U) tail-weighted bivariate measures of dependence proposed by Krupskii and Joe (2015). The two measures are defined as:

$$\varrho_L(a,p) = Cor\left[a\left(1-\frac{u}{p}\right), a\left(1-\frac{v}{p}\right) \middle| u < p, v < p\right] \quad (9)$$

Table 1
Copula families applied to model the joint distribution of the annual flows at the JAG and MET equivalent reservoirs. Estimated parameters, tail-weighted dependence metrics and fit performance. Ordered from lowest to highest AIC values. Note that the BB1 copula is the one, with asymmetry towards the lower tail ($\varrho_L > \varrho_U$), that is closest to the observed data.

Copula	Parameters names	Parameters values	Kendall τ	AIC	ϱ_L	ϱ_U
Observed Data	–	–	0.67	–	0.81	0.64
Gaussian	ρ	0.85	0.65	−111.76	0.67	0.67
Student t	ρ; ν	0.85; 30	0.64	−109.21	0.68	0.68
BB1	θ; δ	0.73; 1.93	0.62	−105.92	0.76	0.65
Rotated Gumbel (180°)	δ	2.6	0.62	−105.49	0.81	0.50
Frank	δ	8.95	0.64	−99.66	0.41	0.50
Gumbel	δ	2.52	0.60	−97.70	0.47	0.76
Clayton	δ	2.31	0.54	−90.4	0.84	0.17
Joe	δ	2.87	0.50	−73.56	0.10	0.78

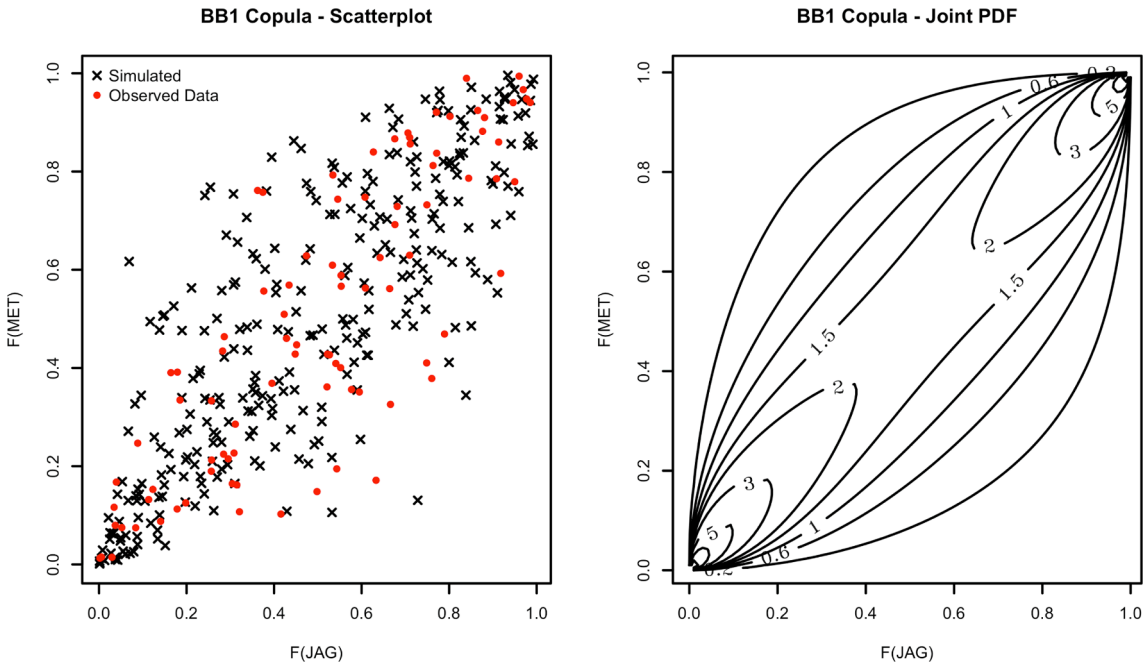

Fig. 7. BB1 copula simulation versus observed data scatterplots (left) and BB1 copula joint probability distribution function contour plot (right). Both plots are in the [0,1] uniform space.

$$\varrho_U(a,p) = Cor\left[a\left(1-\frac{1-u}{p}\right), a\left(1-\frac{1-v}{p}\right)\middle| 1-u<p, 1-v<p\right] \quad (10)$$

Where $a(\cdot)$ is a monotonic increasing function in the [0,1] domain and p is the truncation level with values in the (0,0.5] interval. The monotonic increasing function and truncation level used were respectively $a(x) = x^6$ and $p = 0.5$ as recommended by Krupskii and Joe (2015).

Equations (7) and (8) provide only the spatial correlation structure. The temporal dependence is modelled, within each series marginal distribution, as the single site GLMs described by Eqs. (5) and (6). Thus, after u_t and v_t (spatially correlated variables) are sampled, JAG'_t and MET'_t (temporally and spatially correlated variables) can be determined by the inverse of the respective GLM's CDF.

3.5. Generation algorithm

The proposed multivariate annual streamflow simulation procedure is detailed in Fig. 4. It can be described as an eight-step process:

i The probability distributions of the annual inflow series at the equivalent reservoirs are evaluated to select the respective best fit exponential functions and the lags with relevant correlation coefficients that should compose the covariates.
ii The temporal dependence is modelled as described by Eqs. (5) and (6) by fitting univariate GLMs with the covariates and exponential distributions selected for each series (marginal distributions) in step i.
iii The uniform u-v variables are calculated from the estimated expectations and the GLMs' CDFs at each time position.
iv The spatial correlation structure is established by fitting a copula distribution between the observed u-v uniform variables.
v After modelling both dependence structures, the random simulation starts with the generation of n (the length of the synthetic time series) random u from the uniform [0,1] distribution and then n random v are drawn from the bivariate copula conditional distribution (Eq. (8)).
vi The first ($t = 1$) synthetic generated u_t-v_t pair (spatially correlated) is transformed into the JAG'_t and MET'_t synthetic streamflow pair (temporally and spatially correlated) by the inverse of the marginals GLMs' CDF with expectations determined from the set of covariates (lags) values (the initial set may be a random sample from the historical series).
vii The set of covariates values is updated with the generated JAG'_t and MET'_t.
viii The time position is updated ($t = t + 1$) and the steps vi and vii are repeated for the next u_t-v_t pairs until all pairs are transformed into the synthetic generated streamflow series ($t = n$).

3.6. Performance assessment

In order to demonstrate the performance of the single site GLM and of the multisite GLM Copula annual streamflow time series stochastic simulation models, their efficiency is compared respectively to univariate autoregressive (AR) and multivariate ARMA models described by Salas et al. (1980). The multisite GLM Copula is also compared to the state-of-art copula model, COPAR. The synthetic replicates from the stochastic models should preserve the statistical characteristics (i.e. mean, standard deviation and skewness) and the dependence structures (Srivastav and Simonovic, 2014).

The univariate AR and the multivariate ARMA models were fitted to the normalized JAG and MET inflow series. The data normalization was done using the Box-Cox power transformation. The fitting and sampling procedures were carried out with the R MARIMA package (Spliid, 2017). The first order univariate AR and the (1,1) order multivariate ARMA were selected since they resulted in the lowest AIC values.

The COPAR model (Brechmann and Czado, 2015) applies the Vine Copula theory (Aas et al., 2009) to model both serial and cross-sectional dependences. Vine copulas are based on the decomposition of the multivariate copula density into a product of bivariate copulas, also called pair copula construction.

In this paper, a first order COPAR (1) model is chosen as benchmark. It was reproduced through the original algorithm described in Brechmann and Czado (2015). The application of the COPAR (1) to the case

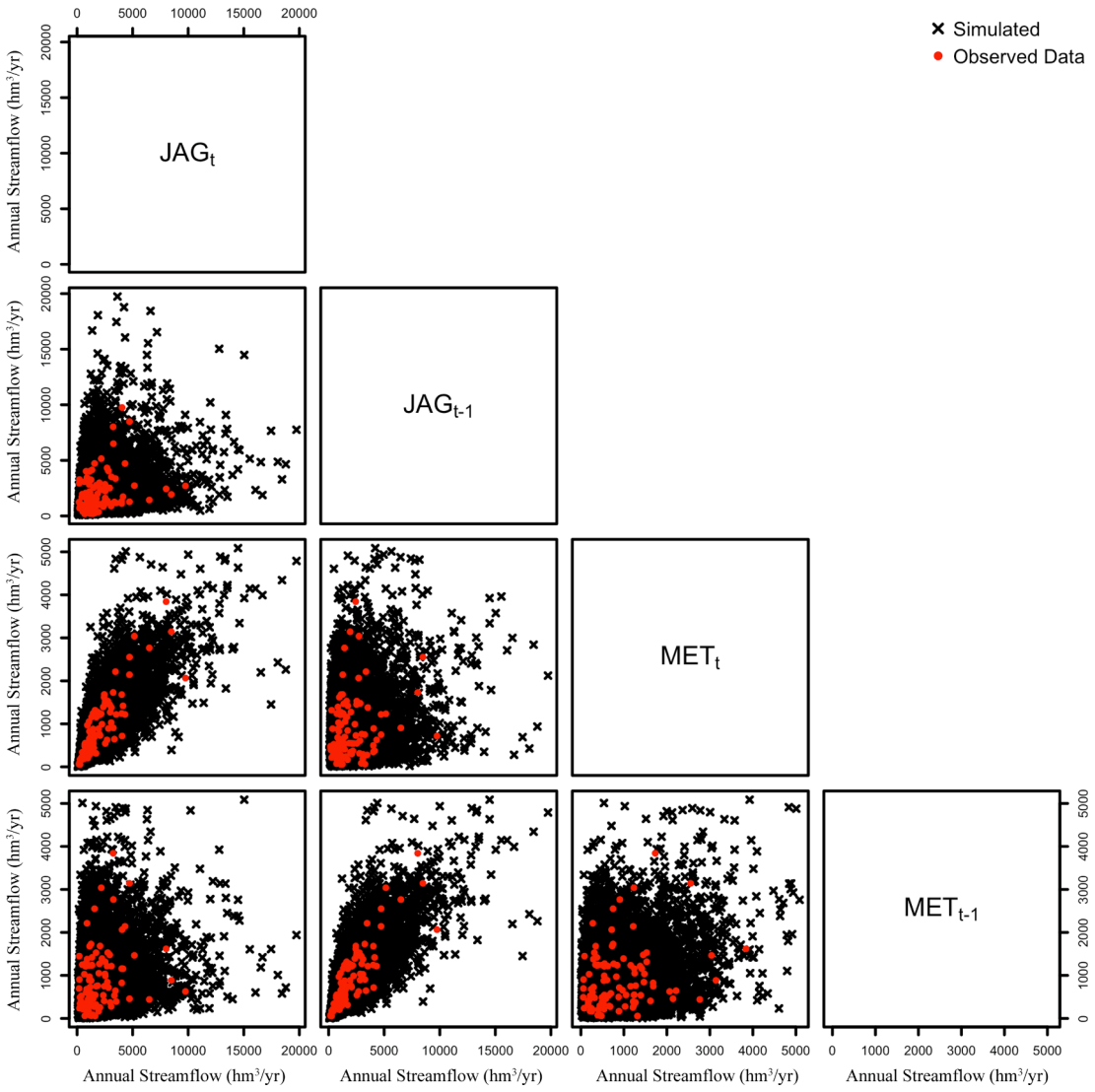

Fig. 8. Temporal and spatial dependency structures scatterplot, 100 points of observed data (red circles) and 10,000 points of simulated data (black crosses) from the GLM-Copula multivariate model. The panels in the off-diagonal show the relationships between the classifications on the diagonal blocks. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

study time series required the sequential fitting of five bivariate copulas: $C(JAG_t, JAG_{t-1})$, $C(JAG_t, MET_t)$, $C(JAG_{t-1}, MET_t \mid JAG_t)$, $C(MET_{t-1}, JAG_t \mid JAG_{t-1})$ and $C(MET_{t-1}, MET_t \mid JAG_{t-1}, JAG_t)$.

The families and parameters of each of the bivariate copulas were estimated using the Two-Stage Maximum Likelihood Estimation method where the parameters of marginal distributions are initially estimated and then the parameters of the copula function are estimated using Maximum Likelihood with the marginals computed from the previously fitted marginal distributions (Singh and Zhang, 2018). The fitting and simulation procedures of the COPAR model were carried out with functions from the 'VineCopula' R-package.

One hundred synthetic annual streamflow series (runs) with the same length as the observed series (n = 101 years) are generated by each model and their statistical characteristics and dependence structures are compared graphically based on boxplots against the historical data. A historical series behavior is judged to be preserved by the synthetic series when its values lie within the box (Salas and Lee, 2010; Lee and Salas, 2011; Hao and Singh, 2013). The use of one hundred runs is in accordance with the streamflow simulation literature (Lee and Salas, 2011; Hao and Singh, 2013; Srivastav and Simonovic, 2014).

As a nonlinear measure of performance, the copula entropy (CE) of the observed variables (JAG_t, JAG_{t-1}, MET_t, MET_{t-1}) was compared to the CE of the model's synthetic series. Based on the definition of Shannon's Entropy (Shannon, 1948), Ma and Sun (2011) proposed the copula entropy as the entropy of the copula function and showed its relation with joint and marginal entropy. They also proved the equivalence between the negative of CE and mutual information (MI).

MI is a traditional non-linear measure of the dependences/association between random variables based on entropy theory (Cover and Thomas, 1991). However, the estimation of MI for more than two variables is a hard task, while CE just requires the variables copula joint distribution (Alpettiyil Krishnankutty et al., 2020). Thus, CE is a useful multivariate estimator of MI.

CE has been used to measure the association between stock market variables (Zhao and Lin, 2011), multiple degradation processes (Sun et al., 2019) and river flows (Chen et al., 2013). It was also used as a performance measure in feature selection for rainfall-runoff modeling and drought prediction (Chen et al., 2014; Huang and Zhang, 2019) and in selecting vine copula structure for multisite streamflow simulation (Ni et al., 2020).

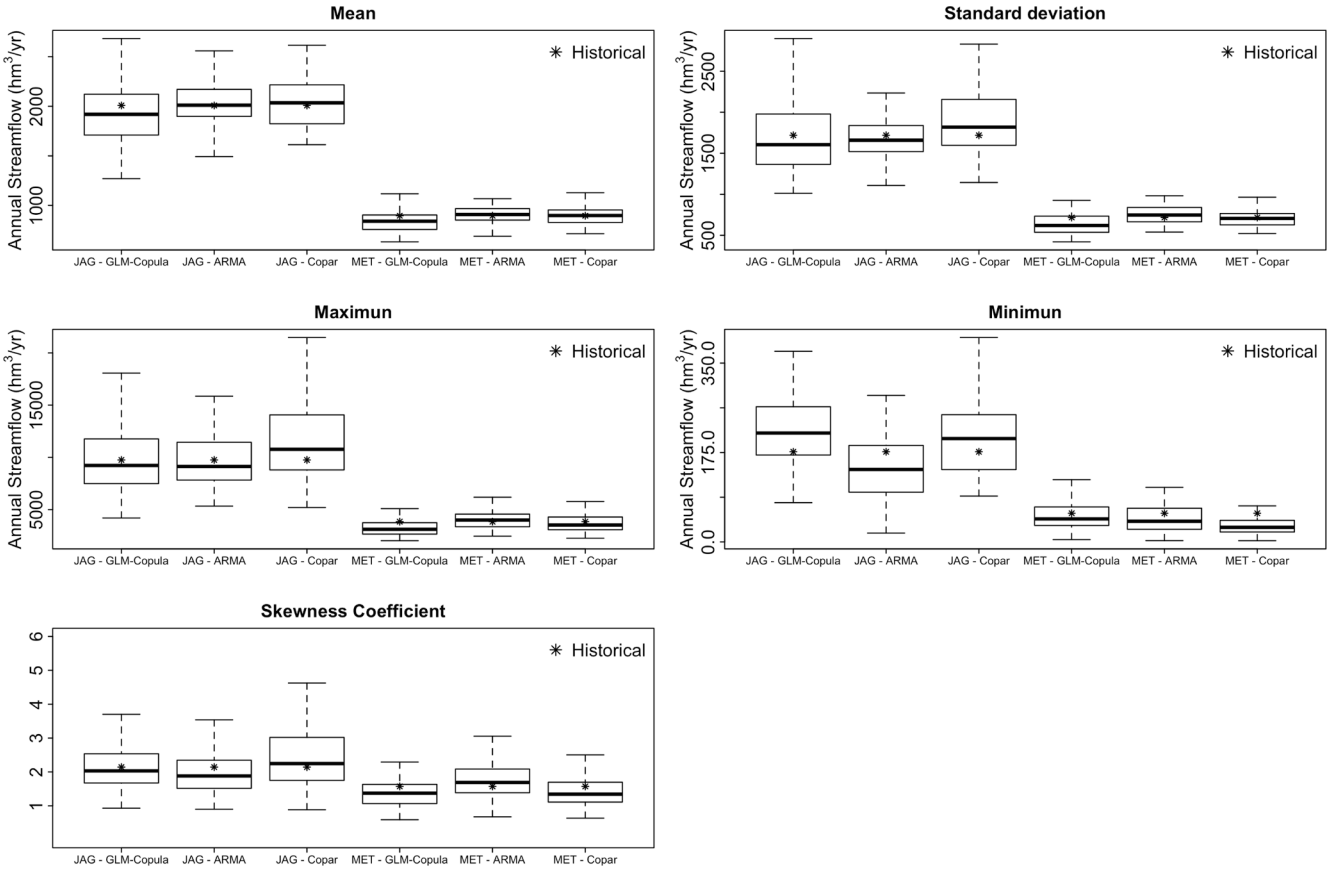

Fig. 9. Comparison, for both equivalent reservoirs (JAG and MET), of some of the annual statistics of the historical series and of the synthetic series obtained by the multivariate GLM-Copula, ARMA and COPAR models. The boxes range from the first to the third quartile and the whiskers have maximum length of 1.5 × IQR (interquartile range).

In this research, the copula entropy of the observed data and the resulted from the synthetic time series were calculated through the 'copent' R package (Ma, 2020) which calculates CE through a nonparametric estimation of the copula function.

As a practical exercise, we also analyzed the performance of the models to simulate drought conditions, simply understood as being below average conditions. For this purpose, in each case study, the years with annual flows below the respective historical mean were assigned to drought conditions.

The maximum number of consecutive years under drought conditions (i.e. the longest drought period) in each of the one hundred synthetic series of the GLM-Copula, ARMA and COPAR models were compared to the longest drought period of the historical series. The longest period under drought conditions is a relevant constraint in the design and in the operation of the artificial reservoirs.

4. Results

4.1. GLM single site streamflow simulation

The ability of the single site GLM and AR methods to preserve the historical statistics and temporal dependence structures is presented respectively in Figs. 5 and 6. For this purpose and like all the other figures, boxplots were drawn of the generated values of the annual statistics and these were compared with the historical statistics. For the boxplots, the whiskers have maximum length of 1.5 × IQR (interquartile range) and the values outside the whiskers are considered to be outliers (Robbins, 2004).

Fig. 5 shows that all the historical statistics are well reproduced by both methods except for the standard deviation of the inflows to MET from the GLM model and for the minimum of the inflows to JAG from the AR model. Individually, the GLM reproduced better the minimum and the skewness coefficient, the first since the AR model underestimated the minimum of both series and the latter may be due to not requiring data normalization. Also, the methods were better in reproducing the maxima than the minima and, overall, presented the same performance in both JAG and MET regarding the replication of the sample statistics.

From the autocorrelation function (ACF) analysis in Fig. 6, both methods were similarly efficient in representing the first lag autocorrelation (the short-term temporal dependence) and also had similar dispersions. The GLM also depicted the long-term dependence peak at the 10th and 11th lags, however with better performance for the MET inflow series (gamma distribution). An interesting advantage presented by the GLM series is that their ACF is not represented as an exponential decay like the autoregressive models but instead it is able to capture the lagged correlations used as covariates (1st,10th and 11th). Thus, the GLM can be applied to mimic some complex ACF designs by considering

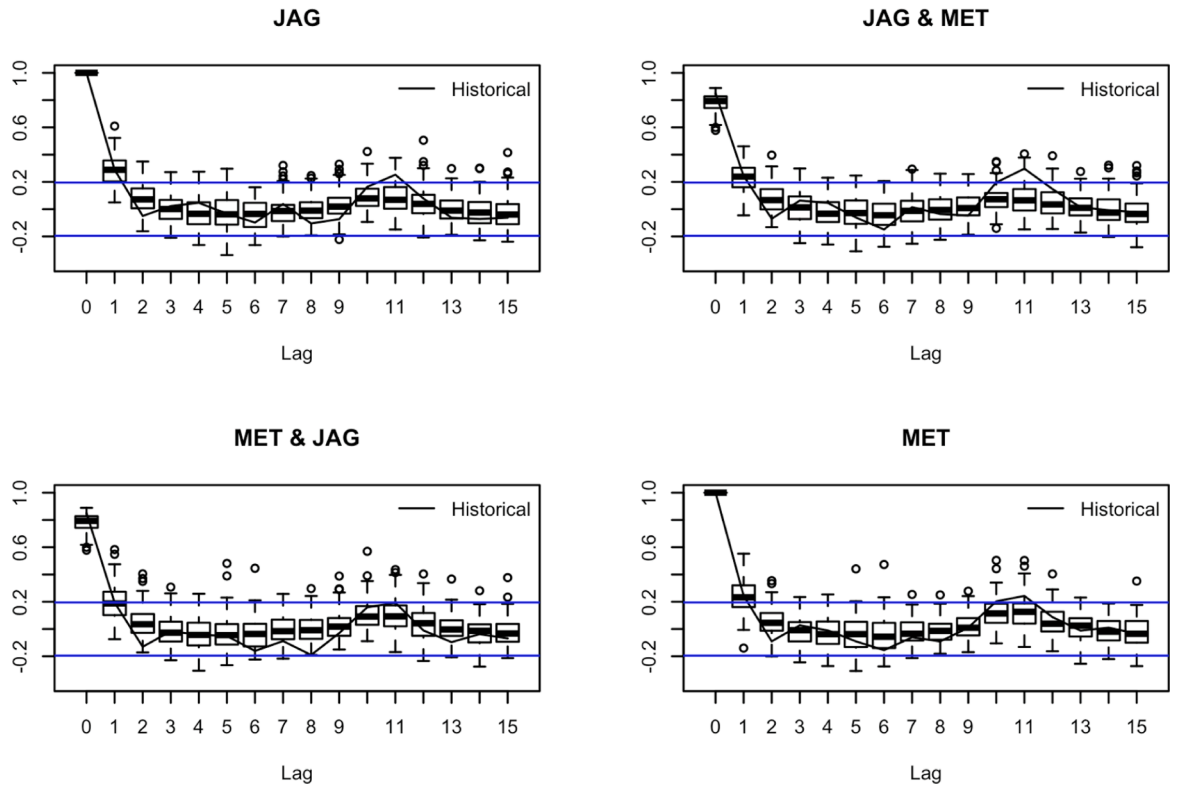

Fig. 10. Comparison, for both equivalent reservoirs (JAG and MET) of the annual auto and cross-correlation functions of the historical series and of the multivariate synthetic series obtained by the GLM-Copula model. The boxes range from the first to the third quartile and the whiskers have maximum length of 1.5 × IQR (interquartile range).

a larger set of lags as covariates. However, the increase of the set of covariates increases the number of parameters and reduces the parsimony of the model which can be a drawback depending on the ratio between the length of the observed series and the number of parameters.

4.2. GLM copula multisite streamflow simulation

4.2.1. Copula fitting and family selection

Table 1 presents the estimated parameters, the tail-weighted dependence metrics (ϱ_L, ϱ_U) and the performance metrics of different copula families to model the joint distribution of the annual inflows at JAG and MET equivalent reservoirs. The Kendall τ is a nonlinear measure of correlation between two series.

The copula family defines how the joint distribution is modeled. For instance, the Gaussian copula represents the joint distribution with the same association intensity despite the values, while the Gumbel copula has a higher association for large values and the Clayton copula for lower values. The Frank copula has a high association for the middle values and low in the extremes (Lee and Salas, 2008). More details of the tested copula families can be found in Nelsen (2006) and Joe (2014).

Copula tail asymmetry can be inferred from the ϱ_L and ϱ_U values: i) if ϱ_L is stronger than ϱ_U then the joint probability distribution might have greater values in the joint lower tail, i.e. there is tail asymmetry toward the joint lower tail; ii) if ϱ_L is weaker than ϱ_U then the joint probability distribution might have greater values in the joint upper tail, i.e. there is tail asymmetry toward the joint upper tail; iii) if ϱ_L is about equal ϱ_U then the joint probability distribution values in both tails might be similar, i.e. there is no tail asymmetry. With these definitions, it is possible to understand the tail asymmetry of the observed data and of each of the fitted copulas presented in the Table 1.

The Gaussian and Student copulas presented the minima AIC values and the Kendall τ values closest to the observed. However, as the observed data presented a small tail asymmetry towards the lower joint tail, symmetrical copulas like the Gaussian and Student *t* copulas might not be the best suited to model the observed data. Therefore, the BB1 copula, which is the fitted copula with asymmetry towards the lower tail that presented the lowest AIC value, was the copula selected to model the spatial dependence for the case study.

Fig. 7 illustrates the scatterplot of 300 random uniform pairs (F(JAG), F(MET)) simulated from the BB1 copula versus the observed values. It also depicts the contour plot of the BB1 copula joint probability distribution values. Fig. 7 shows that dependence structure was well preserved by the BB1 copula.

4.2.2. Preservation of the dependency structure and historical statistics

Fig. 8 presents, as an illustration of the dependency structure, the pair plot of the JAG_t, JAG_{t-1}, MET_t, MET_{t-1} variables for the observed data and for the synthetic series generated by the GLM-Copula model. The GLM-Copula model visually reproduced the dependency structure of the observed variables.

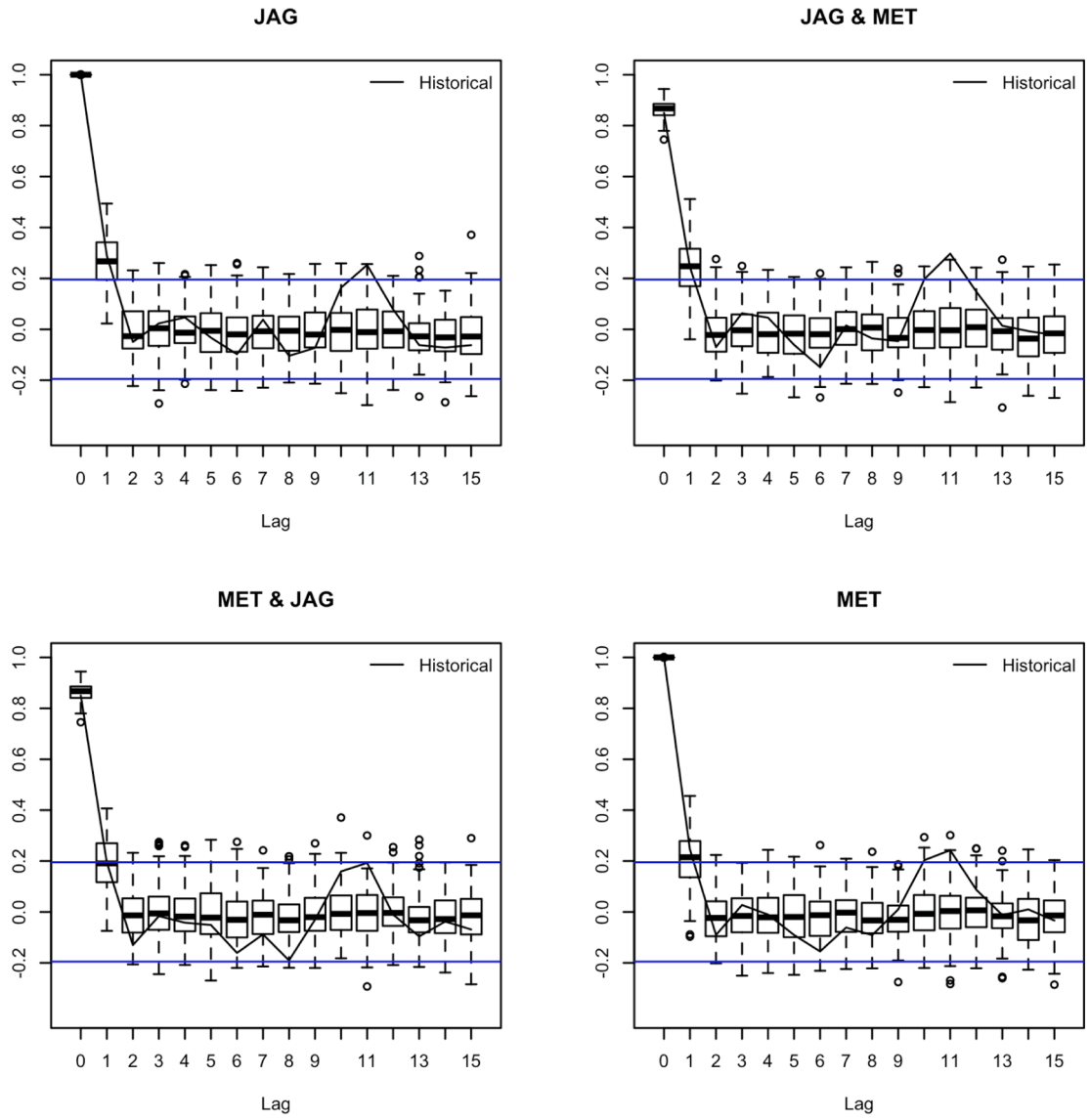

Fig. 11. Comparison, for both equivalent reservoirs (JAG and MET) of the annual auto and cross-correlation functions of the historical series and of the multivariate synthetic series obtained by the ARMA model. The boxes range from the first to the third quartile and the whiskers have maximum length of 1.5 × IQR (inter-quartile range).

For both equivalent reservoirs, Figs. 9–11 compare some statistical characteristics of the historical series with those related to the dependence structure of the synthetic annual flow series generated by the GLM-Copula, ARMA and COPAR multivariate models.

All three methods were able to reproduce the historical statistics with similar efficiency. Like the univariate case, the GLM based model carried out the skewness coefficient better than ARMA and the ARMA model underestimated the minimum values.

From the auto and cross-correlation functions, Figs. 10–12 for the GLM-Copula, ARMA and COPAR models respectively, the methods represented the short-term dependence structure, i.e. the first lag autocorrelation (temporal dependence) and the lag 0 cross-correlation (spatial dependence), with matched efficiency and spreads.

Like the univariate GLM model, the GLM-Copula model also has the ability to represent the long terms dependencies as its correlation function do not necessarily follow an exponential decay (Fig. 10), which is not the case of the ARMA model regarding the long-term dependencies (Fig. 11). Thus, the proposed model showed the same performance in representing the short-term dependencies than the ARMA and COPAR models, while having the advantage over ARMA of being flexible and intuitive to depict isolated peaks in the auto and cross-correlation functions by choosing the appropriate lags as covariates and being simpler than the COPAR model.

The relative errors (Eq. (11)) of the statistical characteristics between the synthetic and the historical series were computed as another performance measure (Silva and Portela, 2012) and are presented in Table 2.

$$Relativeerror(\%) = \frac{Gen - Hist}{Hist} x100 \tag{11}$$

where Gen denotes the mean of the 100 statistics estimated from the

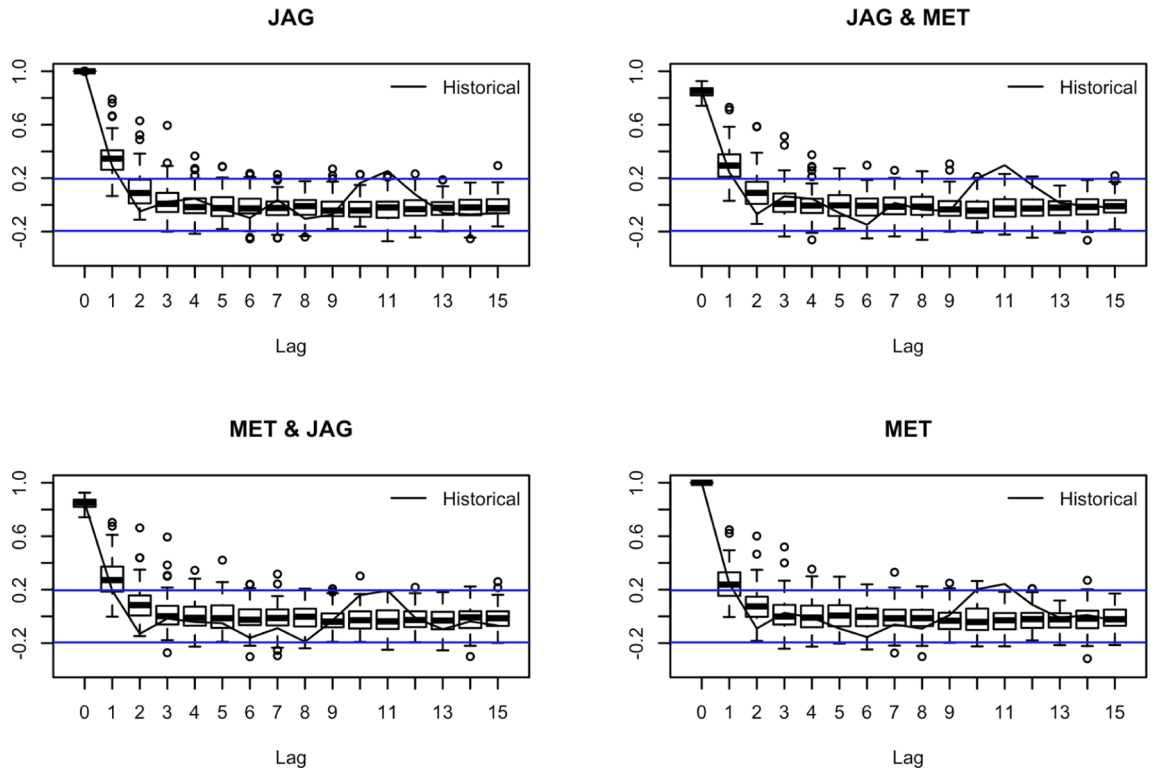

Fig. 12. Comparison, for both equivalent reservoirs (JAG and MET) of the annual auto and cross-correlation functions of the historical series and of the multivariate synthetic series obtained by the COPAR model. The boxes range from the first to the third quartile and the whiskers have maximum length of 1.5 × IQR (inter-quartile range).

Table 2
Relative errors of the annual parameters and historical values for both equivalent reservoirs. The values in **bold** identify the criterion with the best performance.

Statistics	JAG – Relative error %			MET – Relative error %			Historical values	
	GLM-Copula	ARMA	COPAR	GLM-Copula	ARMA	COPAR	JAG	MET
Mean	−4.09	**1.01**	2.71	−6.03	1.18	**0.80**	2007.91	896.24
Standard deviation	**−1.35**	−1.42	14.31	−10.42	5.63	**−0.17**	1719.42	717.27
Maximum	2.02	**1.20**	25.89	−14.79	10.15	**−2.59**	9751.80	3841.52
Minimum	24.25	−15.71	**12.49**	**−3.8**	−17.66	−38.92	176.38	56.13
Skewness	**2.08**	−6.84	16.94	−11.86	16.30	**−7.31**	2.11	1.55
Lag-1 auto-correlation	**−1.59**	−8.73	18.46	**−1.74**	−14.07	2.54	0.29	0.24
Lag-0 cross-correlation	−8.32	1.76	**−0.57**	–			0.85	

generated synthetic series and Hist the same statistic estimated from the corresponding historical sample.

The results of Table 2 corroborate those denoted by the previous boxplots and are not conclusive as to whether there is a better model for simulating annual streamflow when comparing the preservation of historical statistics. However, it can be noticed that GLM-Copula was the best in reproducing the temporal dependence statistic for both series which is likely to be due to the GLM's margin modeling. In contrast, the GLM-Copula was the worst in reproducing the spatial dependence statistics.

In comparison to the ARMA model, the GLM-Copula reproduced the Skewness better for both series while being worse reproducing the mean. By comparing GLM-Copula and COPAR, it can be noticed that the former was better in reproducing JAG statistics while the latter was better in reproducing MET's, which also happened in the comparison between ARMA and COPAR.

Fig. 13 presents the comparison of the copula entropy for the original data and the models' synthetic series as a nonlinear measure of total association between the variables (JAG_t, JAG_{t-1}, MET_t, MET_{t-1}). It shows that the three models resulted in higher association than the observed data and that the GLM-Copula model was the closest to the observed total association and reasonable better than both benchmark models.

4.2.3. *Drought conditions*

In Fig. 14, the longest drought period from each of the synthetic series are compared with the historical values. Although the methods applied matched performance in reproducing the historical statistics, the GLM-Copula model was significantly superior in simulating drought duration for the JAG inflow series, however with greater spreads. The ARMA longest drought showed lower durations than the GLM-Copula's and led to an underestimation in the JAG series. Both copula models (GLM-Copula and COPAR) were better than ARMA in this criterion for both series. For MET, the GLM-Copula and the COPAR presented matched performance while ARMA was slightly worse.

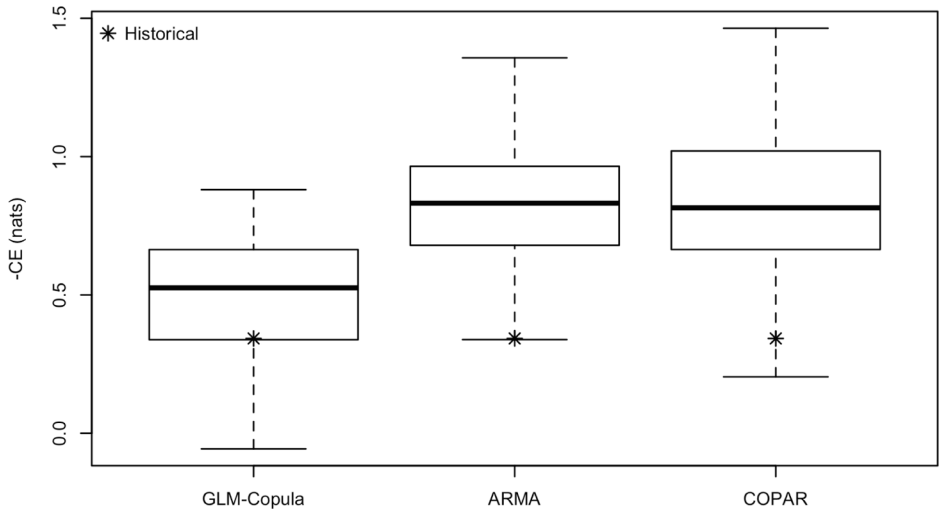

Fig. 13. Comparison of the copula entropy (CE) for the observed data and the synthetic series obtained by the multivariate GLM-Copula, ARMA and COPAR models. The boxes range from the first to the third quartile and the whiskers have maximum length of 1.5 × IQR (interquartile range).

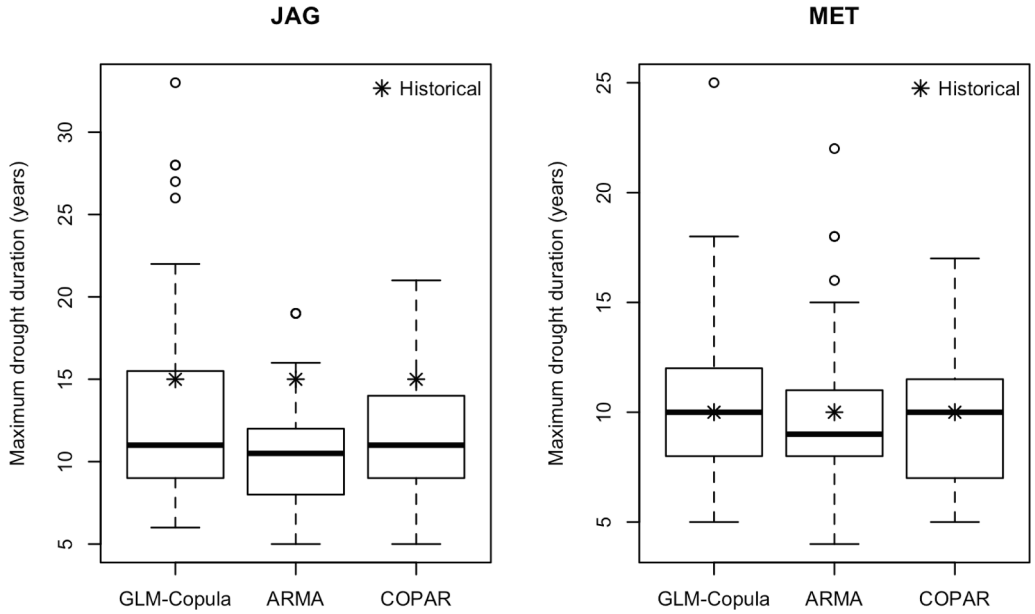

Fig. 14. Maximum drought duration comparison for both equivalent reservoirs (JAG and MET) of the historical and of the multivariate synthetic time series generated by the GLM-Copula, ARMA and COPAR models. The boxes range from the first to the third quartile and the whiskers have maximum length of 1.5 × IQR (interquartile range).

While the greater dispersion presented by the GLM-Copula and COPAR time series models would not be a problem in the stochastic optimization of the reservoir system operation rules, the use of the ARMA generated series would imply a false higher water availability for the JAG series, increasing the vulnerability of the system to longer droughts and the risk of water shortage due to the pluriannual characteristic of the reservoirs.

5. Summary and conclusions

This paper presents the implementation of a new multisite stochastic annual streamflow simulation approach based on the combination of bivariate copulas (spatial dependence) and Generalized Linear Models (temporal dependence). This research also brought a simple application of GLM to generate univariate streamflow time series. The authors believe this work is the first research to apply Generalized Linear Models to stochastic streamflow simulation.

The GLM-Copula time series model efficiently exploits synergies and the flexibility of both techniques and its main advantages are that they:

i Do not require data normalization, hence the GLMs flexibility to deal with any exponential family distribution.
ii Have capacity to represent non-conventional long-term ACF designs intuitively by just considering significant lags as covariates of the GLMs.
iii And have flexibility to model spatial dependence by defining the copula family

The results showed that the GLM-Copula approach ability to preserve summary statistics from the historical data was similar to the classical

multivariate ARMA and the state-of-art COPAR models. For the dependency structures, the GLM-Copula reproduced what was narrowly the best in reproducing the short-term temporal dependence (lag-1 autocorrelation), narrowly the worst in reproducing the spatial dependence (lag-0 cross-correlation) and reasonable the best in reproducing the total association (copula entropy). Thus, the proposed GLM-Copula model can be an alternative with matched, if not better, performance when compared to the existing time series simulation methods.

In comparison with the ARMA model, the GLM-Copula was better in reproducing both the skewness coefficient and the maximum drought duration, and the latter was underestimated by the ARMA model. COPAR was also better reproducing the maximum drought duration than the ARMA model.

These results are similar to those obtained by other copula approach studies of skewness coefficient replication (Lee and Salas, 2011; Chen et al., 2019) and for drought representation (Lee and Salas, 2011). Although Lee and Salas (2011) considered these results as "marginal benefits", we suggest that in both works it is clear that the ARMA models lead to an underestimation of drought conditions while the copula-based models do not. For an underdeveloped semiarid region like that considered in the case study, water resources planning with misleading drought information could result in heavy economical losses. Thus, the copula based GLM-Copula and COPAR synthetic series are preferable to those resulting from the ARMA model in drought dependent stochastic applications. In addition, the better reproduction of the skewness coefficient might be related to the lack of data normalization. Thus, GLM or copula based methods like the GLM-Copula are flexible parametric approaches that can be applied even when data normalization fails.

Compared to the COPAR model, the GLM-Copula has the advantages of being simpler and reducing the computational burden for multisite and/or greater-than-one lag applications, while maintaining the flexibility of the marginal distributions modeling. The main drawback is that it does not model the temporal dependence nonlinearly.

Despite the existence of multiple time series simulation methods, this research showed that there is still space for improvement. The proposed method is intuitive, robust, requires low computational effort and can be easily replicated with open-source R packages. Therefore, the authors consider that the proposed model might be useful in future studies/applications due to its flexibility and the solid results presented.

The model was created in its simplest form and due to its flexibility, can be easily extended by combining Generalized Linear Models with numerical data or by extending them to include exogenous climate variables that affect streamflow. The extension of the GLM-Copula to higher dimensions (more than two spatially dependent time series) is straightforward by combining GLMs for temporal dependence modeling with vine or maximum entropy copulas for spatial dependence modeling. Modeling temporal dependencies in combinationwith GLMs would soften the curse of dimensionality in vine and maximum entropy copulas applications.

Funding

The research was supported by grants from the Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil (CNPq).

CRediT authorship contribution statement

Victor Costa Porto: Conceptualization, Methodology, Software, Visualization, Writing - original draft. **Francisco de Assis de Souza Filho:** Conceptualization, Validation, Supervision, Funding acquisition. **Taís Maria Nunes Carvalho:** Software, Validation, Visualization. **Ticiana Marinho de Carvalho Studart:** Supervison, Writing - Review & Editing. **Maria Manuela Portela:** Writing - Review & Editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Dr. Geoff Pegram and the other anonymous reviewer for their helpful comments and suggestions. The authors also benefitted from discussions from the colleagues of the Grupo de Gerenciamento de Risco Climático (GRC) of the Federal University of Ceará.

References

Aas, K., Czado, C., Frigessi, A., Bakken, H., 2009. Pair-copula constructions of multiple dependence. Insur. Math. Econ. 44, 182–198. https://doi.org/10.1016/j.insmatheco.2007.02.001.

Alexandre, A.M., Martins, E.S., Clarke, R.T., Reis, D.S., 2005. Regionalização de parâmetros de modelos hidrológicos. An. do XVI Simpósio Bras. Recur. Hídricos. ABRH, João Pessoa- PB, Brasil.

Alpettiyil Krishnankutty, B., Ganapathy, R., Sankaran, P.G., 2020. Non-parametric estimation of copula based mutual information. Commun. Stat. – Theory Methods 49, 1513–1527. https://doi.org/10.1080/03610926.2018.1563180.

Andreoli, R.V., Kayano, M.T., 2004. Multi-scale variability of the sea surface temperature in the Tropical Atlantic. J. Geophys. Res. C Ocean. 109, 1–12. https://doi.org/10.1029/2003JC002220.

Andreoli, R.V., Kayano, M.T., 2006. Tropical Pacific and South Atlantic effects on rainfall variability over Northeast Brazil. Int. J. Climatol. 26 (13), 1895–1912. https://doi.org/10.1002/joc.1341.

Barros, F.V.F., Martins, E.S.P.R., Souza Filho, F.A., 2013. Regionalização de parâmetros do modelo chuva-vazão SMAP das bacias hidrográficas do Ceará. In: Souza Filho, F. A. (Ed.), Gerenciamento de Recursos Hídricos no Semiárido, 1st ed. Expressão Gráfica e Editora, Fortaleza, pp. 186–207.

Brechmann, E.C., Czado, C., 2015. COPAR - Multivariate time series modeling using the copula autoregressive model. Appl. Stoch. Model. Bus. Ind. 31, 495–514. https://doi.org/10.1002/asmb.2043.

Box, G.E.P., Jenkins, G.M., 1976. Time Series Analysis: Forecasting and Control, revised ed. Holden-Day, San Francisco, California, USA.

Chandler, R.E., 2005. On the use of generalized linear models for interpreting climate variability. Environmetrics 16, 699–715. https://doi.org/10.1002/env.731.

Chandler, R.E., Wheater, H.S., 2002. Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland. Water Resour. Res. 38, 10-1-10–11. 10.1029/2001wr000906.

Chen, L., Qiu, H., Zhang, J., Singh, V.P., Zhou, J., Huang, K., 2019. Copula-based method for stochastic daily streamflow simulation considering lag-2 autocorrelation. J. Hydrol. 578, 123938 https://doi.org/10.1016/j.jhydrol.2019.123938.

Chen, L., Singh, V.P., Guo, S., 2013. Measure of correlation between river flows using the copula-entropy method. J. Hydrol. Eng. 18, 1591–1606. https://doi.org/10.1061/(asce)he.1943-5584.0000714.

Chen, L., Singh, V.P., Guo, S., Zhou, J., Ye, L., 2014. Copula entropy coupled with artificial neural network for rainfall–runoff simulation. Stoch. Environ. Res. Risk Assess. 28, 1755–1767. https://doi.org/10.1007/s00477-013-0838-3.

Chen, L., Singh, V.P., Guo, S., Zhou, J., Zhang, J., 2015. Copula-based method for multisite monthly and daily streamflow simulation. J. Hydrol. 528, 369–384. https://doi.org/10.1016/j.jhydrol.2015.05.018.

Cover, T.M., Thomas, J.A., 1991. Elements of Information Theory, 2nd ed. John Wiley & Sons.

Fahrmeir, L., Tutz, G., 2001. Multivariate Statistical Modelling Based on Generalized Linear Models, Springer Series in Statistics. Springer New York, New York, NY. 10.1007/978-1-4757-3454-6.

Fernandez, B., Salas, J.D., 1990. Gamma-autoregressive models for stream-flow simulation. J. Hydraul. Eng. 116, 1403–1414. https://doi.org/10.1061/(asce)0733-9429(1990)116:11(1403).

Frischkorn, H., Araújo, J.C., Santiago, M.M.F., 2003. Water Resources of Ceará and Piauí. In: Global Change and Regional Impacts. Springer, Berlin Heidelberg, pp. 87–94. https://doi.org/10.1007/978-3-642-55659-3_6.

Furrer, E., Katz, R., 2007. Generalized linear modeling approach to stochastic weather generators. Clim. Res. 34, 129–144. https://doi.org/10.3354/cr034129.

Hao, Z., Singh, V.P., 2011. Single-site monthly streamflow simulation using entropy theory. Water Resour. Res. 47 https://doi.org/10.1029/2010WR010208.

Hao, Z., Singh, V.P., 2013. Modeling multisite streamflow dependence with maximum entropy copula. Water Resour. Res. 49, 7139–7143. https://doi.org/10.1002/wrcr.20523.

Hao, Z., Singh, V.P., 2015. Integrating entropy and copula theories for hydrologic modeling and analysis. Entropy 17, 2253–2280. https://doi.org/10.3390/e17042253.

Hao, Z., Singh, V.P., 2016. Review of dependence modeling in hydrology and water resources. Prog. Phys. Geogr. Earth Environ. 40, 549–578. https://doi.org/10.1177/0309133316632460.
Huang, C.Y., Zhang, Y.P., 2019. Prediction based on copula entropy and general regression neural network. Appl. Ecol. Environ. Res. 17, 14415–14424. 10.15666/aeer/1706_1441514424.
Joe, H., 1997. Multivariate Models and Multivariate Dependence Concepts, Multivariate Models and Multivariate Dependence Concepts. Chapman and Hall/CRC. 10.1201/9780367803896.
Joe, H., 2014. Dependence modeling with copulas, Dependence Modeling with Copulas. CRC Press. 10.1201/b17116.
Kao, S.C., Govindaraju, R.S., 2008. Trivariate statistical analysis of extreme rainfall events via the Plackett family of copulas. Water Resour. Res. 44 https://doi.org/10.1029/2007WR006261.
Kleiber, W., Katz, R.W., Rajagopalan, B., 2012. Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes. Water Resour. Res. 48 https://doi.org/10.1029/2011WR011105.
Krupskii, P., Joe, H., 2015. Tail-weighted measures of dependence. J. Appl. Stat. 42, 614–629. 10.1080/02664763.2014.980787.
Lall, U., Sharma, A., 1996. A nearest neighbor bootstrap for resampling hydrologic time series. Water Resour. Res. 32, 679–693. https://doi.org/10.1029/95WR02966.
Lee, T., Salas, J.D., 2008. Using Copulas for Stochastic Streamflow Generation. In: World Environmental and Water Resources Congress 2008. American Society of Civil Engineers, Reston, VA, pp. 1–10. https://doi.org/10.1061/40976(316)572.
Lee, T., Salas, J.D., 2011. Copula-based stochastic simulation of hydrological data applied to Nile River flows. Hydrol. Res. 42, 318–330. https://doi.org/10.2166/nh.2011.085.
Ma, J., 2020. Copent: Estimating Copula Entropy in R. R package version 1. https://CRAN.R-project.org/package=copent.
Ma, J., Sun, Z., 2011. Mutual information is copula entropy. TSINGHUA Sci. Technol. 16, 51–54. https://doi.org/10.1016/S1007-0214(11)70008-6.
McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models, 2nd ed. Chapman and Hall/CRC. 10.1201/9780203753736.
McMahon, T.A., Adeloye, A.J., Zhou, S.L., 2006. Understanding performance measures of reservoirs. J. Hydrol. 324, 359–382. https://doi.org/10.1016/j.jhydrol.2005.09.030.
Moura, A.D., Shukla, J., 1981. On the dynamics of droughts in northeast Brazil: Observations, theory and numerical experiments with a general circulation model. J. Atmos. Sci. 38, 2653–2675. https://doi.org/10.1175/1520-0469(1981)038<2653:OTDODI>2.0.CO;2.
Nelsen, R.B., 2006. An Introduction to Copulas, An Introduction to Copulas. Springer, New York. 10.1007/0-387-28678-0.
Nelder, J.A., Wedderburn, R.W.M., 1972. Generalized Linear Models. J. R. Stat. Soc. 135, 370–384. https://doi.org/10.2307/2344614.
Ni, L., Wang, D., Wu, J., Wang, Y., Tao, Y., Zhang, J., Liu, J., Xie, F., 2020. Vine copula selection using mutual information for hydrological dependence modeling. Environ. Res. 186, 109604 https://doi.org/10.1016/j.envres.2020.109604.
Pereira, G.A.A., Veiga, Á., Erhardt, T., Czado, C., 2017. A periodic spatial vine copula model for multi-site streamflow simulation. Electr. Power Syst. Res. 152, 9–17. https://doi.org/10.1016/j.epsr.2017.06.017.
Prairie, J.R., Rajagopalan, B., Fulp, T.J., Zagona, E.A., 2006. Modified K-NN Model for Stochastic Streamflow Simulation. J. Hydrol. Eng. 11, 371–378.
R Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria http://www.R-project.org/.
Rajagopalan, B., Salas, J.D., Lall, U., 2010. Stochastic methods for modeling precipitation and streamflow, in: Advances in Data-Based Approaches for Hydrologic Modeling and Forecasting. World Scientific Publishing Co., pp. 17–52. 10.1142/9789814307987_0002.
Robbins, N.B., 2004. Creating More Effective Graphs, Creating More Effective Graphs. John Wiley & Sons, Inc., Hoboken, NJ, USA. 10.1002/9780471698180.
Salas, J.D., 1993. Analysis and modeling of hydrologic time series. In: Maidment, D.R. (Ed.), Handbook of Hydrology. McGraw-Hill, New York.
Salas, J.D., Delleur, J.W., Yevjevich, V., Lane, W.L., 1980. Applied Modeling of Hydrologic Time Series. Water Resources Publications, Littleton, Colorado.
Salas, J.D., Lee, T., 2010. Nonparametric simulation of single-site seasonal streamflows. J. Hydrol. Eng. 15, 284–296. https://doi.org/10.1061/(asce)he.1943-5584.0000189.
Schepsmeier, U., Stoeber, J., Brechmann, E.C., Graeler, B., Nagler, T., Erhardt, T., Killiches, M., 2018. Package 'VineCopula'. R package version 2 (5).
Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27, 623–656. https://doi.org/10.1002/j.1538-7305.1948.tb00917.x.
Sharma, A., O'Neill, R., 2002. A nonparametric approach for representing interannual dependence in monthly streamflow sequences. Water Resour. Res. 38, 5-1-5–10. 10.1029/2001wr000953.
Sharma, A., Tarboton, D.G., Lall, U., 1997. Streamflow simulation: A nonparametric approach. Water Resour. Res. 33, 291–308. https://doi.org/10.1029/96WR02839.
Silva, A.T., Portela, M.M., 2012. Disaggregation modelling of monthly streamflows using a new approach of the method of fragments. Hydrol. Sci. J. 57, 942–955. https://doi.org/10.1080/02626667.2012.686695.
da Silva, S.M.O., de Souza, F., Filho, A., Aquino, S.H.S., 2017. Avaliação do risco da alocação de água em período de escassez hídrica: o caso do Sistema Jaguaribe-Metropolitano. Eng. Sanit. e Ambient. 22, 749–760. https://doi.org/10.1590/s1413-41522017161303.
Singh, V.P., Zhang, L., 2018. Copula–entropy theory for multivariate stochastic modeling in water engineering. Geosci. Lett. 5 https://doi.org/10.1186/s40562-018-0105-z.
Sklar, M., 1959. Fonctions de repartition an dimensions et leurs marges. Publ. inst. statist. univ. Paris 8, 229–231.
Spliid, H., 2017. marima: Multivariate ARIMA andARIMA-X Analysis. R package version 2, 2. https://CRAN.R-project.org/package=marima.
Souza Filho, F.A., Lall, U., 2003. Seasonal to interannual ensemble streamflow forecasts for Ceara, Brazil: Applications of a multivariate, semiparametric algorithm. Water Resour. Res. 39 https://doi.org/10.1029/2002WR001373.
Srinivas, V.V., Srinivasan, K., 2005. Hybrid moving block bootstrap for stochastic simulation of multi-site multi-season streamflows. J. Hydrol. 302, 307–330. https://doi.org/10.1016/j.jhydrol.2004.07.011.
Srivastav, R.K., Simonovic, S.P., 2014. An analytical procedure for multi-site, multi-season streamflow generation using maximum entropy bootstrapping. Environ. Model. Softw. 59, 59–75. https://doi.org/10.1016/j.envsoft.2014.05.005.
Sun, F., Zhang, W., Wang, N., Zhang, W., 2019. A copula entropy approach to dependence measurement for multiple degradation processes. Entropy 21. https://doi.org/10.3390/e21080724.
Verdin, A., Rajagopalan, B., Kleiber, W., Katz, R.W., 2014. Coupled stochastic weather generation using spatial and generalized linear models. Stoch. Environ. Res. Risk Assess. 29, 347–356. https://doi.org/10.1007/s00477-014-0911-6.
Wang, W., Dong, Z., Lall, U., Dong, N., Yang, M., 2019. Monthly streamflow simulation for the headwater catchment of the yellow river basin with a hybrid statistical-dynamical model. Water Resour. Res. 55, 7606–7621. https://doi.org/10.1029/2019WR025103.
Wang, X., Auler, A.S., Edwards, R.L., Cheng, H., Cristalli, P.S., Smart, P.L., Richards, D. A., Shen, C.-C., 2004. Wet periods in northeastern Brazil over the past 210 kyr linked to distant climate anomalies. Nature 432, 740–743. https://doi.org/10.1038/nature03067.
Wheater, H.S., Chandler, R.E., Onof, C.J., Isham, V.S., Bellone, E., Yang, C., Lekkas, D., Lourmas, G., Segond, M.L., 2005. Spatial-temporal rainfall modelling for flood risk estimation. Stoch. Environ. Res. Risk Assess. 19, 403–416. https://doi.org/10.1007/s00477-005-0011-8.
Yang, C., Chandler, R.E., Isham, V.S., Wheater, H.S., 2005. Spatial-temporal rainfall simulation using generalized linear models. Water Resour. Res. 41, 1–13. https://doi.org/10.1029/2004WR003739.
Zachariah, M., Reddy, M.J., 2013. Development of an entropy-copula-based stochastic simulation model for generation of monthly inflows into the Hirakud Dam. ISH J. Hydraul. Eng. 19, 267–275. https://doi.org/10.1080/09715010.2013.804697.
Zhao, N., Lin, W.T., 2011. A copula entropy approach to correlation measurement at the country level. Appl. Math. Comput. 218, 628–642. https://doi.org/10.1016/j.amc.2011.05.115.