



Modelos periódicos de cópulas vine para simulação de cenários de vazões de duas usinas hidrelétricas da região Sul do Brasil

Guilherme A. A. Pereira

Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica
Rua Marquês de São Vicente, 225, Gávea - Rio de Janeiro, RJ, Brasil, 22451-900
gaap@ele.puc-rio.br

Álvaro Veiga

Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica
Rua Marquês de São Vicente, 225, Gávea - Rio de Janeiro, RJ, Brasil, 22451-900
alvf@ele.puc-rio.br

RESUMO

Este trabalho visa avaliar uma metodologia sofisticada baseada em cópulas vine periódicas para representar e simular séries de vazões mensais de duas usinas hidrelétricas localizadas na região Sul do Brasil. O modelo pode ser entendido como uma versão não linear do tradicional modelo periódico autorregressivo (PAR(p)). Dentre as principais características destacam-se a possibilidade de se modelar qualquer tipo de estrutura temporal entre meses subsequentes; o fato do comportamento marginal ser modelado separadamente da estrutura de dependência temporal; a flexibilidade para escolha das distribuições marginais, permitindo que características tradicionalmente observadas em dados hidrológicos como assimetria e curtose sejam adequadamente representadas; os cenários simulados serão sempre válidos (positivos). Os resultados indicam que o modelo é capaz de simular cenários respeitando características presentes na séries históricas tais como média, variância, distribuições de probabilidade e dependência temporal.

PALAVRAS CHAVE. Cópulas vine, Análise de séries temporais, Simulação de cenários de vazões.

EN - PO na Área de Energia; MP - Modelos Probabilísticos.

ABSTRACT

This paper aims to evaluate a sophisticated methodology based on vine copulas to represent and simulate streamflow time series of two hydroelectric power plants located in southern Brazil. The model can be viewed as a non-linear version of the periodic autoregressive model (PAR(p)). Among the main features of the model applied here, we highlight the possibility of modeling any type of temporal structure between adjacent months; the fact that the marginal behavior is modeled separately from the temporal dependence; the flexibility to choose the marginal distributions, allowing features traditionally observed in hydrologic data such as skewness and kurtosis to be properly represented; and the fact that all simulated scenarios will be valid (will have positive values). The results indicate that the model is capable of simulating scenarios preserving features presented in the original streamflow series, such as mean, variance, probability distributions, and temporal dependence.

KEYWORDS. Vine copula, Time series analysis, Streamflow scenario simulation.

EN - OR in Energy; MP - Probabilistic Models.



1. Introdução

No Brasil, a energia gerada é majoritariamente fornecida por usinas hidrelétricas. Um modelo adequado que descreva as vazões dessas usinas é de vital importância para o setor elétrico. Seu planejamento e operação estão fundamentados em plataformas computacionais baseadas em otimização estocástica. Muitas dessas plataformas, são alimentadas por cenários de vazões que buscam caracterizar toda a incerteza associada à disponibilidade de água para a geração. Deste modo, é notória a importância dos modelos de simulação de cenários para o setor elétrico.

A acurácia de um modelo de simulação de cenários pode ser aferida por sua habilidade em gerar dados sintéticos que preservam características históricas observadas. Em sua maioria, tais modelos são baseados nos modelos autorregressivos de [Box e Jenkins, 1976]. Especialmente no caso de vazões brasileiras pode-se citar dentre outros, [Olivera et al., 2011], [Pereira e Souza, 2014], [Souto et al., 2014]. Estes modelos são construídos sob hipóteses rígidas (simplificadoras) no que diz respeito ao tipo de dependência temporal entre vazões adjacentes ou à forma das distribuições marginais ou das distribuições conjuntas, [Sharma e O'Neill, 2002]. Dentre as limitações dessa classe, [Sharma e O'Neill, 2002] e [Zachariah e Reddy, 2013] destacam: (i) Dados de vazões geralmente não seguem uma distribuição Gaussiana; (ii) A hipótese de normalidade limita a capacidade do modelo em representar formas não usuais de distribuições de probabilidade; (iii) Modelos Gaussianos usualmente modelam apenas a dependência linear; (iv) Existe a possibilidade de geração de cenários com valores negativos, uma vez que a distribuição Gaussiana é definida no intervalo $(-\infty, \infty)$.

Sabe-se que com o uso de cópulas, todas as desvantagens supracitadas podem ser contornadas. Desde o trabalho pioneiro de [Sklar, 1959], onde foi demonstrado que uma função de distribuição multivariada pode ser decomposta em uma cópula C e um conjunto de distribuições univariadas, o uso das cópulas se popularizou. Dentre as inúmeras razões, destaca-se a possibilidade de se modelar o comportamento marginal das variáveis separadamente da estrutura de dependência. Além do mais, as cópulas podem representar diversos tipos de associação, tais como efeitos não-lineares, assimetria, dependência de cauda, etc. Existem diversos trabalhos empregando cópulas para modelagem e simulação de séries de hidrológicas. Uma extensa apresentação sobre este tema pode ser encontrada em [Pereira, 2017].

As cópulas também podem ser utilizadas para modelar a dependência serial dentro de uma série temporal. [Joe, 2014] ilustra como estas podem criar versões não-lineares dos modelos AR(p), MA(q) e ARMA(p,q). Em finanças, por exemplo, já existem extensões dos modelos AR(1)-ARCH(1) e do modelo VAR propostas por [Mendes e Aíube, 2011] e [Brechmann e Czado, 2015] respectivamente.

Em termos de simulação univariada de cenários de vazões, [Lee e Salas, 2011] foram os pioneiros. Estes utilizaram cópulas bivariadas e a cópula Gaussiana trivariada para simulação e análise da série anual do Rio Nilo. [Hao e Singh, 2012] propuseram uma metodologia utilizando o princípio de entropia para obtenção das distribuições marginais e cópulas bivariadas para modelar a distribuição conjunta entre dois meses adjacentes. Nesta direção, [Zachariah e Reddy, 2013] utilizaram a cópula Gumbel bivariada para a simulação de cenários de vazões mensais na barragem de Hirakud, Índia. [Kong et al., 2015] também empregaram a cópula Gumbel bivariada em conjunto com o método da máxima entropia para simulação da vazão do Rio Xiangxi, China. [Li et al., 2013] estimaram a distribuição conjunta entre dois meses consecutivos condicionadas a outras variáveis. [Jeong e Lee, 2015] aliaram a abordagem proposta por [Lee e Salas, 2011] ao modelo PDAR (*periodic discrete autoregressive model*) para a simulação de vazões mensais de rios intermitentes.

Em termos de estrutura temporal, o uso de cópulas bivariadas significa que somente a dependência temporal de defasagem 1 (*lag 1*) é considerada. Assim, como pode ser observado, os trabalhos para simulação de cenários hidrológicos supracitados incorporaram somente a dependência temporal de ordem 1. Muitos destes podem ser entendidos como versões não lineares do tradicional modelo periódico autorregressivo Gaussiano de ordem 1, PAR(1). Por outro lado, sabe-se que em



séries periódicas esta hipótese nem sempre será verdadeira.

Para acomodar a dependência temporal periódica de qualquer ordem nos modelos de cópulas para simulação de cenários, [Pereira e Veiga, 2017] propuseram uma metodologia baseada nas cópulas vine. Estas são um tipo de cópula d -dimensional cuja construção ocorre iterativamente através da junção de cópulas bivariadas. Dentre as principais características dessa abordagem destacam-se: (i) a dependência temporal de qualquer ordem pode ser incorporada; (ii) a possibilidade da representação de efeitos não-lineares; (iii) o fato de todos os cenários simulados serem positivos; (iv) a flexibilidade para se modelar as distribuições marginais mensais. O autor avaliou sua metodologia em apenas uma série de vazão do Rio Manso obtendo resultados satisfatórios.

Desta forma, o objetivo deste trabalho é realizar um estudo mais amplo da metodologia proposta por Pereira e Veiga [2017] num conjunto de vazões de usinas localizadas na região Sul do Brasil, uma vez que as mesmas apresentam grandes desafios para a modelagem. Este artigo está organizado da seguinte maneira. Na seção 2, apresentamos uma breve introdução às cópulas e as cópulas vine. Na seção 3 descrevemos brevemente a metodologia aplicada. Na seção 4 realizamos o estudo de caso. Por fim, apresentamos as conclusões obtidas.

2. Aspectos teóricos

2.1. Cópula

Cópula é uma ferramenta extremamente útil na análise de dependência. De forma geral, toda distribuição multivariada carrega informações acerca do comportamento individual de cada variável (distribuições marginais) e da estrutura de dependência que rege a relação entre as mesmas. A abordagem via cópula viabiliza a análise separada destes componentes, permitindo assim, o estudo mais detalhado das estruturas de dependência que governam um vetor aleatório. Uma cópula d -dimensional C é uma função de distribuição acumulada multivariada definida em $[0, 1]^d$ com marginais uniformemente distribuídas.

O teorema de [Sklar, 1959] estabelece que se $\mathbf{Y} = (Y_1, \dots, Y_d)$ é um vetor aleatório com função de distribuição conjunta F e distribuições marginais F_1, \dots, F_d , então existe uma cópula $C : [0, 1]^d \rightarrow [0, 1]$ de forma que para $(y_1, \dots, y_d) \in \mathbb{R}^d$,

$$F(y_1, \dots, y_d) = C(F_1(y_1), \dots, F_d(y_d)). \quad (1)$$

Se as distribuições marginais são contínuas, então a cópula C é única. Esta carrega toda a estrutura de dependência de um vetor aleatório e pode ser entendida como uma função que conecta distribuições marginais para formar uma distribuição multivariada.

A densidade $f(y_1, \dots, y_d)$ correspondente a F pode ser escrita em termos de densidade cópula ao se calcular a derivada parcial da equação (1). Dessa forma, temos

$$f(y_1, \dots, y_d) = c_{1, \dots, d}(F_1(y_1), \dots, F_d(y_d)) \cdot \prod_{i=1}^d f_i(y_i), \quad (2)$$

onde $c_{1, \dots, d}(\cdot)$ é a densidade cópula d -dimensional.

2.2. Cópulas vine

Os grandes avanços na literatura de cópulas ocorreram principalmente no caso das cópulas bivariadas. O número de cópulas bidimensionais disponíveis assim como as estruturas de dependência que estas podem representar são vastos. Por outro lado, em altas dimensões (caso d -dimensional), as poucas cópulas existentes possuem restrições que as tornam menos flexíveis.

A construção pair-cópula (*pair copula construction - PCC*), ou simplesmente cópula vine, representa uma alternativa para a construção de cópulas em altas dimensões. Esta ideia foi inicialmente abordada por [Joe, 1996] e posteriormente desenvolvida em [Bedford e Cooke, 2001] e



[Bedford e Cooke, 2002]. A construção pair-cópula consiste em reescrever uma função de densidade multivariada f em um produto de cópulas incondicionais e condicionais e densidades marginais. Com isso, é possível a construção de cópulas d -dimensionais utilizando toda a flexibilidade das cópulas bivariadas. Para mais detalhes veja [Aas et al., 2009].

A princípio, qualquer função de densidade conjunta pode ser reescrita como

$$f(y_1, \dots, y_d) = f(y_d|y_1, \dots, y_{d-1}) \cdot f(y_{d-1}|y_1, \dots, y_{d-2}) \cdots f(y_1). \quad (3)$$

O princípio geral por trás da construção pair-cópula é reescrever cada uma das densidades condicionais da equação (3) em termos de cópulas bivariadas e densidades marginais. Assim, [Aas et al., 2009] demonstraram que estas densidades condicionais podem ser expressas recursivamente em termos de densidades cópulas através da seguinte relação:

$$f(y|\mathbf{v}) = c_{y,v_j|\mathbf{v}_{-j}}(F(y|\mathbf{v}_{-j}), F(v_j|\mathbf{v}_{-j})) \cdot f(y|\mathbf{v}_{-j}), \quad (4)$$

onde $c_{y,v_j|\mathbf{v}_{-j}}(\cdot, \cdot)$ representa uma densidade cópula bivariada, \mathbf{v} é um vetor d -dimensional, v_j é um dos componentes de \mathbf{v} , e \mathbf{v}_{-j} é um vetor igual à \mathbf{v} eliminando-se o seu j -ésimo componente. Por exemplo, no caso 3-dimensional, uma possível decomposição seria (para mais detalhes veja [Aas et al., 2009]):

$$\begin{aligned} f(y_1, y_2, y_3) &= c_{1,2}(F_1(y_1), F_2(y_2)) \cdot c_{2,3}(F_2(y_2), F_3(y_3)) \cdot \\ & c_{1,3|2}(F_{1|2}(y_1|y_2), F_{3|2}(y_3|y_2)) \cdot f_1(y_1) \cdot f_2(y_2) \cdot f_3(y_3). \end{aligned} \quad (5)$$

[Joe, 1996] demonstrou que as distribuições condicionais, os argumentos das cópulas bivariadas nas equações (4 e 5), são obtidas recursivamente ao se empregar

$$F(y|\mathbf{v}) = \frac{\partial C_{y,v_j|\mathbf{v}_{-j}}(F(y|\mathbf{v}_{-j}), F(v_j|\mathbf{v}_{-j}))}{\partial F(v_j|\mathbf{v}_{-j})}. \quad (6)$$

Quando \mathbf{v} é escalar e $v = u_2$ e $y = u_1$ com $u_1, u_2 \sim U[0, 1]$, temos

$$w = h(u_1, u_2; \theta_{12}) = F_{1|2}(u_1|u_2) = \frac{\partial C_{12}(u_1, u_2; \theta_{12})}{\partial u_2}. \quad (7)$$

Esta equação (7) é denominada *h-function*. Nesta, θ_{12} representa o conjunto de parâmetros da cópula bivariada.

Como mencionado anteriormente, este arcabouço matemático também possui grande utilidade para representar a dependência serial presente dentro de uma série temporal. [Joe, 2014] apresenta como cópulas podem ser empregadas para generalizar alguns modelos Markovianos de ordem p . Com isso, estas extensões são modelos mais flexíveis do que os tradicionais modelos lineares autorregressivos, pois incorporam qualquer tipo de dependência. Sob certas condições, se assumirmos que $Y = \{Y_1, \dots, Y_T\}$ é uma série temporal univariada, um processo Markoviano de ordem 1 pode ser representado pela cópula bivariada $C_{t,t-1}(F(y_t), F(y_{t-1}))$. Considerando $u_t = F(y_t)$, um processo Markoviano de ordem 3 é expresso pela cópula $C_{t,t-1,t-2,t-3}(u_t, u_{t-1}, u_{t-2}, u_{t-3})$.

A construção pair-cópula pode ser organizada através de um modelo gráfico proposto por Bedford e Cooke [2001] e Bedford e Cooke [2002]. Este consiste em um conjunto de árvores $V = (T_1, \dots, T_{d-1})$ que respeitam as seguintes condições: (i) T_1 é uma árvore com o conjunto de vértices $N_1 = \{1, \dots, d\}$ e conjunto de arestas E_1 ; (ii) Para $\ell = 2, \dots, d-1$, T_ℓ é uma árvore cujo os vértices são $N_\ell = E_{\ell-1}$ e as arestas são E_ℓ ; (iii) Para $\ell = 2, \dots, d-1$, e $\{a, b\} \in E_\ell$ onde $a = \{a_1, a_2\}$ e $b = \{b_1, b_2\}$, a condição $\#(a \cap b) = 1$ deve ser satisfeita (condição de proximidade). Esta última assegura que se dois vértices estão conectados por uma aresta em T_ℓ , $\ell > 2$, as correspondentes arestas em $T_{\ell-1}$ devem compartilhar um vértice em comum. A Figura 1 representa graficamente uma representação de um processo Markoviano de ordem 2. Esta estrutura é denominada D-vine, uma vez que cada vértice em T_1 possui grau no máximo igual a dois. Para mais detalhes, veja [Mai e Scherer, 2012].

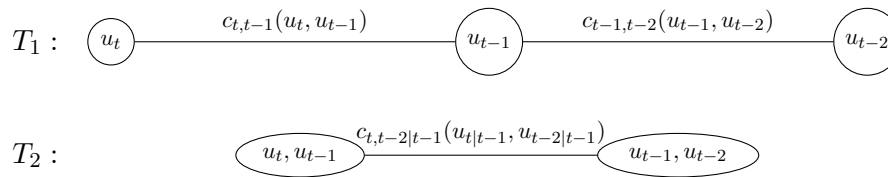


Figura 1: Exemplo de uma D-vine representando um processo Markoviano de ordem 2.

3. Descrição da metodologia

A metodologia empregada para a simulação de cenários mensais de vazões consiste em estimar para cada mês uma D-vine. Nesta, tanto a dimensão quanto as cópulas bivariadas variam de acordo com os meses. Com isso, estaremos construindo modelos autorregressivos não lineares com diferentes ordens para cada mês. Esta metodologia pode ser entendida como uma versão não linear dos tradicionais modelos autorregressivos periódicos (PAR(p)).

Formalmente, temos

$$F^m(y_{t(a,m)-d+1}, \dots, y_{t(a,m)-1}, y_{t(a,m)}) = C_{t, \dots, t-d+1}^m(u_{t(a,m)-d+1}, \dots, u_{t(a,m)-1}, u_{t(a,m)}), \quad (8)$$

onde índice temporal de y_t , $t = 1, \dots, T$, é reescrito como $t = t(a, m) = (a - 1)12 + m$. Neste caso $a = 1, \dots, T/12$ representa os anos e $m = 1, \dots, 12$, denota os meses. Além disso, $u_{t(a,m)} = F^m(y_{t(a,m)})$ e $u_{t(a,m)-i} = F^{m-i}(y_{t(a,m)-i})$.

3.1. Identificação da ordem das cópulas D-vine mensais

Para a identificação da ordem, [Pereira e Veiga, 2017] sugeriram um procedimento iterativo em conjunto com o teste de independência assintótico (veja [Genest e Favre, 2007]). Este pode ser descrito da seguinte maneira.

Para cada período, inicia-se com uma cópula bivariada. Na sequência, estima-se uma cópula tridimensional e verifica-se se a cópula condicional $C_{t-2,t|t-1}$ é uma cópula produto (independência). Caso isto seja verdadeiro, utiliza-se a dimensão igual a 2. Caso contrário, aumenta-se a dimensão para 4. Novamente, testa-se se a cópula condicional de $C_{t-3,t|t-1,t-2}$ é uma cópula de independência ou não. Se existirem evidências de que $C_{t-3,t|t-1,t-2}$ não é uma cópula produto, elege-se a dimensão igual à 4. Por outro lado, se $C_{t-3,t|t-1,t-2}$ é uma cópula de independência, reduzimos a dimensão para 3. Essa ideia é repetida até que a ordem máxima considerada seja alcançada.

Em termos práticos, o objetivo é analisar se a associação entre u_{t-d+1} e u_t , excluído todos os efeitos intermediários é relevante ou não. Então, se u_{t-d+1} não possui associação com u_t , consideramos que a dimensão $t-d+1$ não traz informação adicional para representar a dependência temporal.

3.2. Simulação de cenários de vazões

A simulação de novas observações a partir de uma cópula vine é realizada através de um procedimento iterativo baseado nas cópulas condicionais inversas. Para simplificar a notação, consideremos que t pertence ao mês m . Assim, no caso bidimensional, a simulação da variável u_t condicionada à u_{t-1} é dada por $u_t = C_{t|t-1}^{-1}(w|u_{t-1})$, onde $w \sim U(0, 1)$. Dessa maneira, podemos simular quantas observações desejarmos apenas sorteando novos valores de w de uma uniforme padrão. No caso tridimensional, a variável u_t condicionada aos valores passados de u_{t-1} e u_{t-2} é dada por $u_t = C_{t|t-1,t-2}^{-1}(w|u_{t-1}, u_{t-2})$. No caso d -dimensional, temos

$$u_t := C_{t|t-1,t-2,\dots,t-d+1}^{-1}(w|u_{t-1}, u_{t-2}, \dots, u_{t-d+1}). \quad (9)$$

As cópulas condicionais e as cópulas condicionais inversas são obtidas em termos de cópulas bivariadas ao se iterar as equações (6) e (7). Mais detalhes podem ser encontrados em



[Mai e Scherer, 2012]. Além disso, as variáveis simuladas $u_t \sim U(0, 1)$ devem ser transformadas para escala original através da função de distribuição acumulada inversa apropriada, ou seja, $y_t = F^{-1}(u_t)$.

4. Estudo de caso

A metodologia apresentada foi testada nas séries de vazões mensais (m^3/s) de 2 usinas hidrelétricas localizadas na região Sul do Brasil. Estas são as usinas de Quebra-Queixo e Garibaldi. Os dados foram obtidos junto ao Operador Nacional do Sistema (ONS) e estão disponíveis em seu site oficial [ONS, 2017]. Estas séries possuem início em janeiro de 1931 e término em dezembro de 2015, totalizando 85 anos completos.

Na abordagem via cópulas, o primeiro passo é transformar as variáveis originais (vazões mensais - (m^3/s)) em dados cópula, ou seja, dados que são uniformemente distribuídos. Para esta tarefa, utiliza-se a transformada integral de probabilidade ($u = F(y) \sim U(0, 1)$).

Após uma análise, optou-se pela distribuição gama como distribuição marginal. Assim, para cada mês de cada uma das 2 usinas, uma distribuição gama foi estimada. No total, estimou-se 24 distribuições marginais. A função densidade de probabilidade gama é dada por

$$f(y|\alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} y^{\alpha-1} e^{-y/\beta} \quad (10)$$

onde $y \geq 0$, $\alpha > 0$ é o parâmetro de forma e $\beta > 0$ é o parâmetro de escala. Vale lembrar que existe a possibilidade de se empregar diferentes distribuições para os diferentes meses. A escolha das distribuições marginais ocorre de modo independente e livre.

Inserindo os índices $m = 1, \dots, 12$, para representarem os meses e $s = 1, 2$, para denotar a usinas, os dados cópula são obtidos através da transformada integral de probabilidade,

$$\hat{u}_{t(r,m)}^s = F^{m,s}(y_{t(r,m)}^s | \hat{\alpha}^{m,s}, \hat{\beta}^{m,s}) \quad (11)$$

onde $F^{m,s}$ é a função de distribuição acumulada gama.

Para verificar se os dados transformados seguem realmente uma uniforme padrão empregou-se o teste de Anderson-Darling. A Figura 2 apresenta os p-valores de cada mês para as 2 usinas. A linha preta representa o nível de significância de 5%. Valores acima desse nível indicam que a hipótese de que os dados transformados seguem uma distribuição uniforme padrão não pode ser rejeitada. Como isso, pode-se afirmar que a distribuição gama é uma escolha razoável para as distribuições marginais.

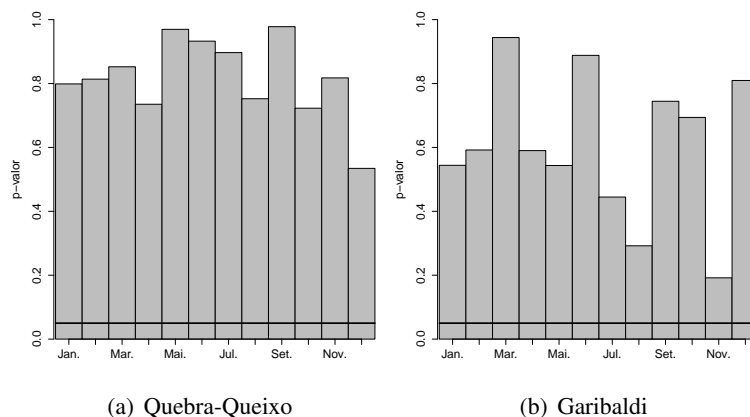


Figura 2: P-valores mensais do teste de Anderson-Darling.

Uma vez transformados os dados originais em uniformes, o próximo passo é estimar as cópulas D-vine periódicas. Como cópulas candidatas, utilizou-se as cópulas de independência (I),



Gaussiana (N), t-Student (S-t), Gumbel (G), Clayton (C) e Frank (F). Empregou-se o BIC (Bayesian Information Criterion) como critério de seleção. Todos os códigos foram implementados em R com suporte do pacote VineCopula [Schepsmeier et al., 2015].

A Tabela 1 apresenta as cópulas bivariadas selecionadas. Como pode ser observado a cópula Clayton ocorre com maior frequência. Esta é uma cópula assimétrica e que possui índice de cauda inferior diferente de zero. Em termos práticos, isto indica que vazões baixas provavelmente serão acompanhadas de vazões baixas. O mesmo não pode ser afirmado para valores altos das vazões, pois esta cópula possui índice de cauda superior igual a zero. Com relação as cópulas Gaussiana e Frank, estas são cópulas simétricas e possuem índices de cauda superior e inferior iguais a zero.

Tabela 1: Número de cópulas selecionadas.

	I	N	S-t	C	G	F
Quebra-queixo	4	2	0	9	0	2
Garibaldi	0	1	0	10	0	2

A Tabela 2 apresenta as ordens selecionadas para os meses. Como pode ser observado, as cópulas bidimensionais são predominantes. Vale lembrar que cópulas D-vine d -dimensionais representam processos Markovianos de ordem $d - 1$. Por exemplo, uma cópula trivariada representa um processo autorregressivo de ordem 2.

Tabela 2: Dimensões selecionadas para os meses.

	Jan.	Fev.	Mar.	Abr.	Mai.	Jun.	Jul.	Ago.	Set.	Out.	Nov.	Dez.
Quebra-queixo	2	2	2	4	2	2	2	2	2	4	3	2
Garibaldi	2	2	2	3	2	2	2	2	2	2	2	2

4.1. Simulação dos cenários

Neste momento apresentamos os cenários simulados e os avaliamos. No total, para cada usina, simulou-se 200 cenários de 60 meses. O último ano histórico foi utilizado para inicializar o procedimento de simulação.

A Figura 3 apresenta todos os 200 cenários simulados (linhas cinza) assim como suas médias mensais (linha azul) e as médias mensais históricas (linha vermelha). Pode-se observar claramente que a média dos cenários praticamente se sobrepõe a média histórica.

Para a análise dos cenários, empregamos um conjunto de 3 testes estatísticos. Estes foram o teste-T, o teste de Levene e o teste de Kolmogorov-Smirnov (K-S). As variáveis de interesse são respectivamente as médias mensais, as variâncias mensais e a distribuição de cada período. Em cada um dos 60 meses simulados, verificamos se estas eram estatisticamente iguais ao correspondente mês histórico. Por exemplo, testamos se a média do primeiro mês simulado era igual a média histórica de janeiro. Em cada usina, estes testes são realizados para cada um dos 60 meses simulados.

A Figura 4 resume os três testes estatísticos em cada uma das usinas analisadas. As barras representam os p-valores de cada teste para cada um dos 60 meses simulados. A linha preta ilustra o nível de significância de 5%. P-valores maiores indicam que a hipótese nula não pode ser rejeitada. Em termos práticos, estes resultados apontam que a metodologia empregada é capaz de simular cenários que replicam as características originais observadas na série temporal. As taxas de aprovação para os cenários da usina de Quebra-Queixo foram 100%, 99% e 100% para o teste-t, o teste de Levene e o teste K-S respectivamente. Para a usina de Garibaldi, os percentuais foram de 99%, 100% e 95% .

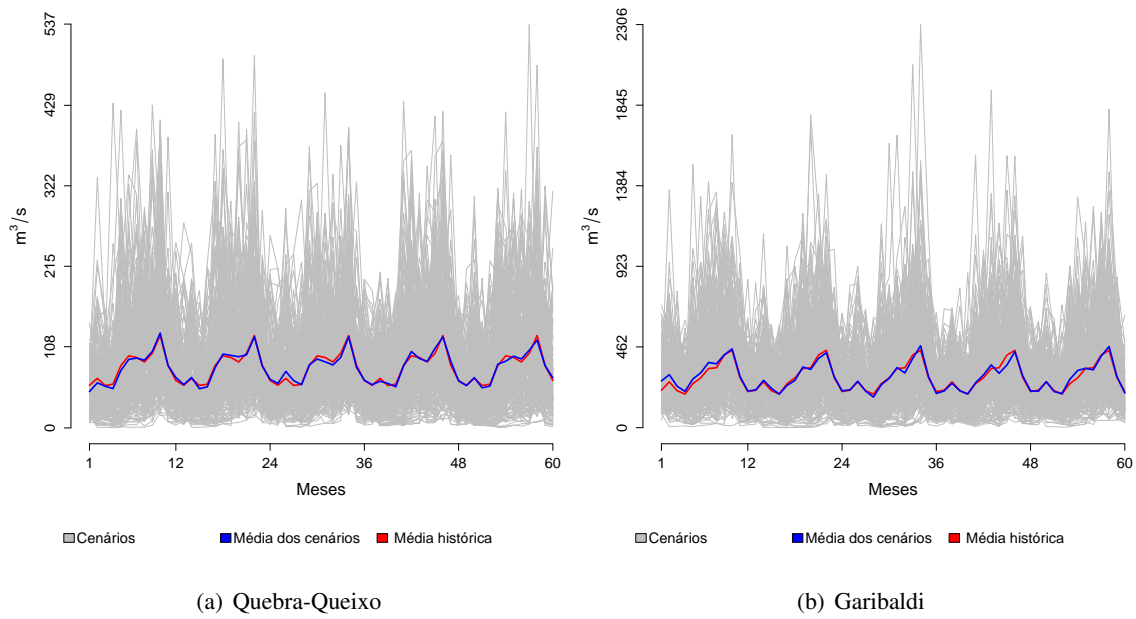


Figura 3: Cenários simulados.

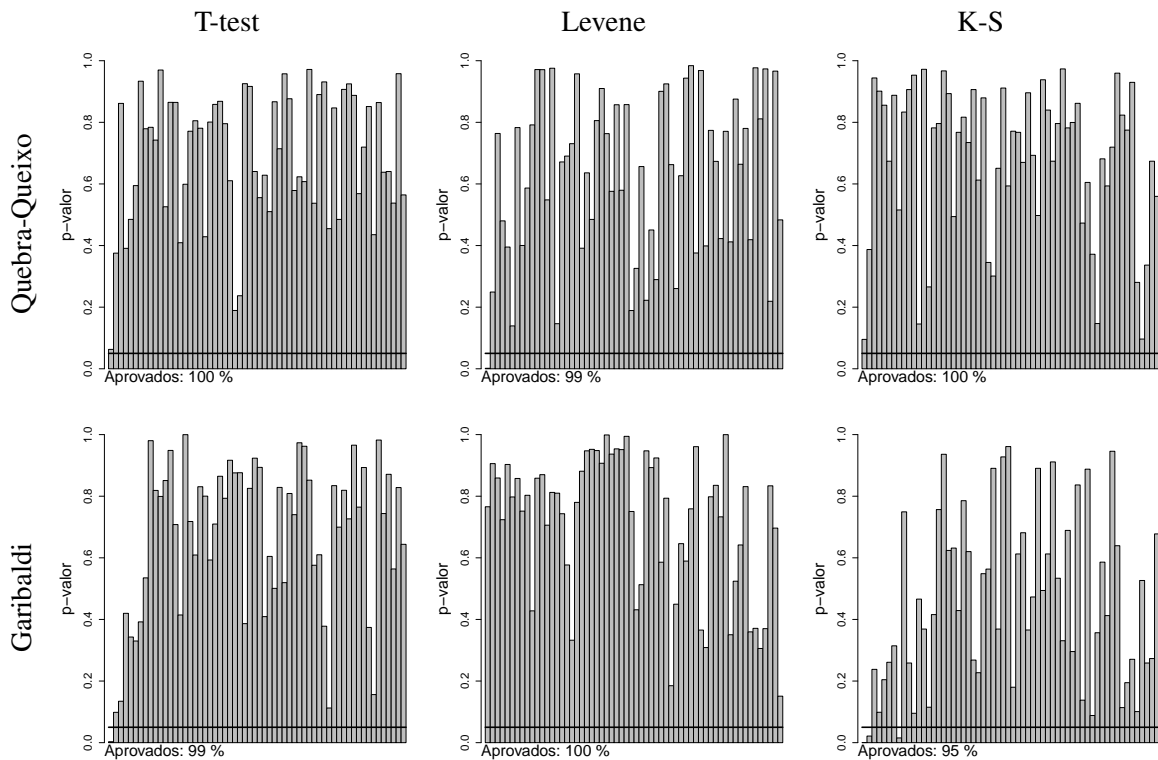


Figura 4: Análise dos cenários: Teste-t, Levene e Kolmogorov-Smirnov.

Por fim, realizamos uma análise da estrutura de dependência dos cenários simulados. O objetivo é averiguar se estes preservam a dependência temporal observada nas séries históricas. Para cada usina, simulou-se 200 cenários de 100 anos. Para cada cenário, calculamos as defasagens (*lags*) de cada modelo mensal e os armazenamos. Por fim, computamos as defasagens médias dessas simulações. As Figuras 5 e 6 apresentam as defasagens 1, 2, 3, 4 e 5 (real e simulado) para cada um dos 12 modelos mensais de cada usina. Como pode ser observado, em praticamente todos os



meses, a média simulada é próxima dos valores históricos. Isto indica que a abordagem periódica com base em cópulas captura corretamente a dependência temporal periódica. Vale lembrar que neste momento, utilizamos o τ de Kendall como medida de associação e não a autocorrelação, tradicionalmente empregada para análise de modelos lineares. Isto ocorre pois o τ de Kendall é uma medida de dependência mais abrangente que a correlação. Mais detalhes sobre suas propriedades e vantagens podem ser encontradas em [McNeil et al., 2010].

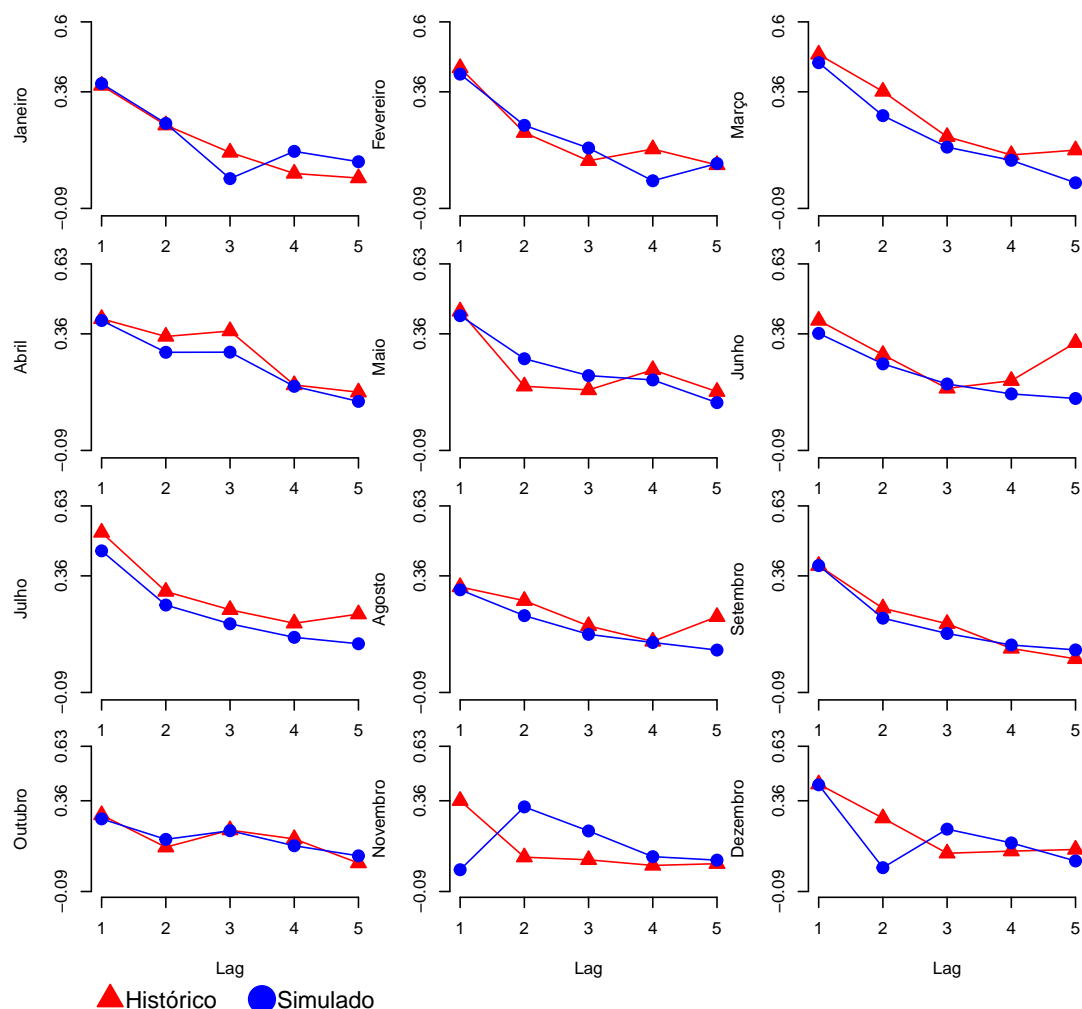


Figura 5: Comparação da auto-dependência de defasagem (*lags*) 1, 2, 3, 4 e 5 para os cenários da usina de Quebra-Queixo.

5. Conclusão

É de conhecimento comum que as vazões da região Sul do Brasil geralmente possuem grandes desafios para a modelagem. Assim este trabalho teve como objetivo avaliar uma metodologia sofisticada baseada em cópulas D-vine proposta por [Pereira e Veiga, 2017] nas séries de vazões mensais de duas usinas hidrelétricas localizadas na região Sul do Brasil.

Até recentemente, os trabalhos para simulação de cenários de vazões baseados em cópulas consideraram apenas a auto-dependência de *lag* 1. Dessa forma, a principal vantagem da metodologia empregada é sua capacidade de modelar adequadamente estruturas temporais de ordens superiores.

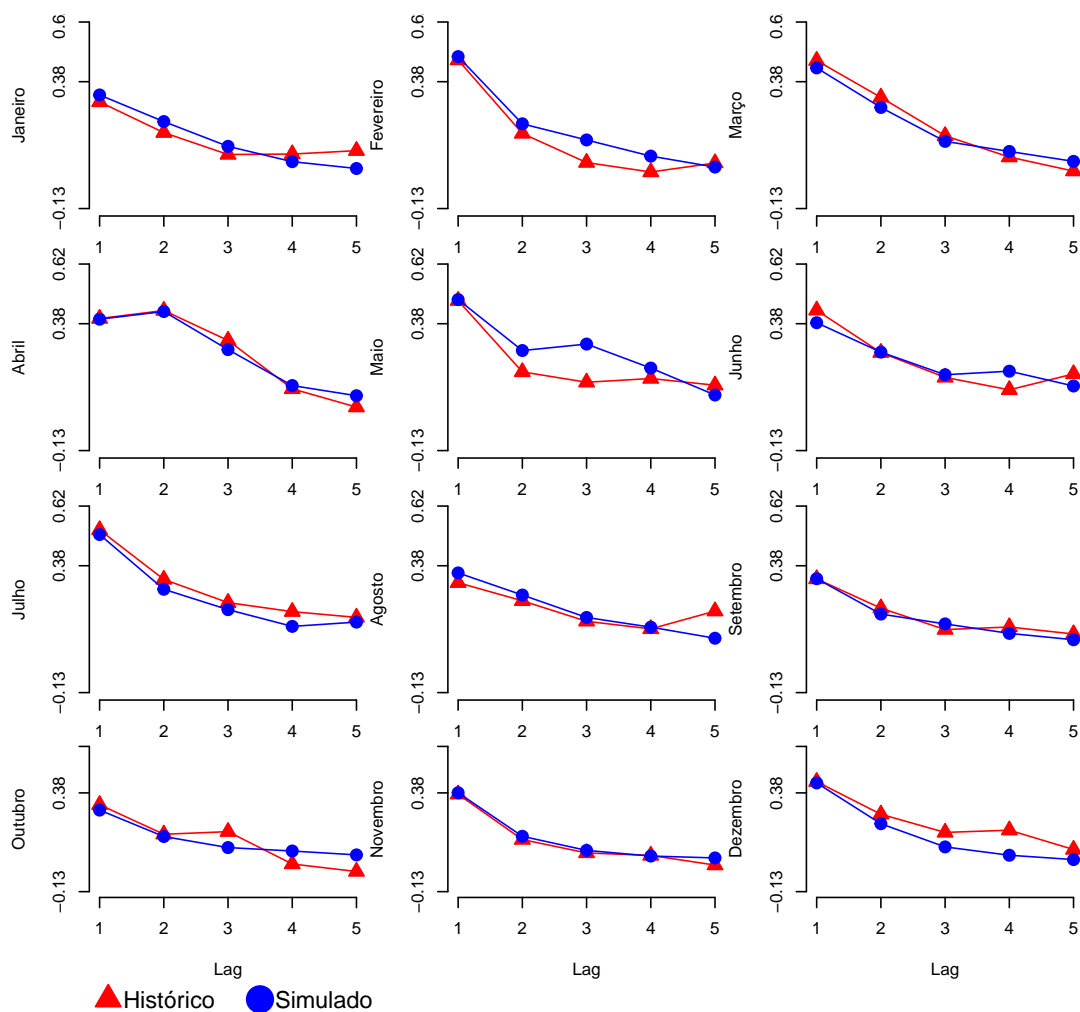


Figura 6: Comparação da auto-dependência de defasagem (*lags*) 1, 2, 3, 4 e 5 para os cenários da usina de Garibaldi.

Além disso, dentre outras características gerais do modelo destacam-se: (i) A possibilidade de se modelar qualquer tipo de associação entre instantes de tempo (efeitos lineares e não lineares, por exemplo); (ii) O comportamento marginal é modelado independentemente da estrutura de dependência temporal; (iii) A flexibilidade para se escolher as distribuições marginais, permitindo que características tradicionalmente observadas em dados hidrológicos como assimetria e curtose possam ser adequadamente representadas. (iv) Os cenários simulados serão sempre válidos (positivos). Para isto basta que as distribuições marginais possuam suporte positivo.

Para as distribuições marginais, optou-se pela distribuição gama. Esta escolha se mostrou adequada uma vez que as variáveis transformadas seguiam aproximadamente uma uniforme padrão. Tal escolha também foi corroborada pelo teste de Anderson-Darling. Ademais, uma vez que a gama é uma distribuição para valores positivos, com esta escolha não existe a possibilidade de se simular cenários com valores negativos.

Em termos das cópulas selecionadas, a cópula Clayton foi a mais empregada. Isto indica que durante a simulação, valores baixos de vazões muito provavelmente serão seguidos por vazões baixas. A mesma afirmação não pode ser feita para vazões altas. Tal motivo é a ausência de índice de cauda superior e a presença de índice de cauda inferior nessa cópula. Com relação a estrutura



temporal, a maior parte das cópulas selecionadas foram cópulas bidimensionais.

Com relação aos resultados das simulações, os conjunto de testes estatísticos empregados indica que o modelo é capaz de gerar cenários que respeitam características observadas nas séries originais, i.e., média, variância e forma das distribuições mensais. Um segundo estudo de simulação indicou que tais cenários também respeitam as estruturas de dependência temporal existentes nas séries históricas.

Referências

- Aas, K., Czado, C., Frigessi, A., e Bakken, H. (2009). Pair-copula constructions of multiple dependences. *Insurance: Mathematics and Economics*, 44(2):182–198.
- Bedford, T. e Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32(1): 245–268.
- Bedford, T. e Cooke, R. M. (2002). Vines: A new graphical model for dependent random variables. *Annals of Statistics*, 30(4):1031–1068.
- Box, G. E. P. e Jenkins, G. M. (1976). *Time series analysis: forecasting and control*. Holden-Day, San Francisco (USA), Revised edition.
- Brechmann, E. C. e Czado, C. (2015). Copar multivariate time series modeling using the copula autoregressive model. *Applied Stochastic Models in Business and Industry*, 31(4):495–514.
- Genest, C. e Favre, A. C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of hydrologic engineering*, 12(4):347–368.
- Hao, Z. e Singh, V. P. (2012). Entropy-copula method for single-site monthly streamflow simulation. *Water Resources Research*, 48(6).
- Jeong, C. e Lee, T. (2015). Copula-based modeling and stochastic simulation of seasonal intermittent streamflows for arid regions. *Journal of Hydro-environment Research*, 9(4):604–613.
- Joe, H. (1996). Families of m -variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. In Rüschendorf, L., Schweizer, B., e Taylor, M. D., editors, *Distributions with fixed marginals and related topics*, volume 28, Lecture Notes-Monograph Series, p. 120–141. Institute of Mathematical Statistics.
- Joe, H. (2014). *Dependence Modeling with Copulas*. CRC Press, Taylor Francis Group, Boca Raton (USA).
- Kong, X. M., Huang, G. H., Fan, Y. R., e Li, Y. P. (2015). Maximum entropy-gumbel-hougaard copula method for simulation of monthly streamflow in xiangxi river, china. *Stochastic Environmental Research and Risk Assessment*, 29(3):833–846.
- Lee, T. e Salas, J. D. (2011). Copula-based stochastic simulation of hydrological data applied to Nile river flows. *Hydrology Research*, 42(4):318–330.
- Li, C., Singh, V. P., e Mishra, A. K. (2013). Monthly river flow simulation with a joint conditional density estimation network. *Water Resources Research*, 49(6):3229–3242.
- Mai, J. F. e Scherer, M. (2012). *Simulating copulas: stochastic models, sampling algorithms, and applications*. Imperial College Press, World Scientific, London (UK).



- McNeil, A. J., Frey, R., e Embrechts, P. (2010). *Quantitative risk management: concepts, techniques, and tools*. Princeton University Press, USA.
- Mendes, B. V. M. e Aíube, C. (2011). Copula based models for serial dependence. *International Journal of Managerial Finance*, 7(1):68–82.
- Olivera, F. L. C., Pereira, G. A. A., Ferreira, P. G., Souza, R. C., e Calili, R. F. (2011). Tratamento estocástico do despacho hidrotérmico no Brasil. In *Anais do XLIII Simpósio Brasileiro de Pesquisa Operacional*, p. 696–707, Ubatuba, São Paulo. SOBRAPO.
- ONS (2017). OPERADOR NACIONAL DO SISTEMA ELÉTRICO, Atualização de séries históricas de vazões. http://www.ons.org.br/operacao/vazoes_naturais.aspx. Acesso em: Fevereiro de 2017.
- Pereira, G. A. A. (2017). Modelos de cópulas para simulação de cenários hidrológicos. Tese de doutorado, Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro.
- Pereira, G. A. A. e Souza, R. C. (2014). Long memory models to generate synthetic hydrological series. *Mathematical Problems in Engineering*, Article ID 823046:1–8. doi:10.1155/2014/823046.
- Pereira, G. e Veiga, Á. (2017). Par (p)-vine copula based model for stochastic streamflow scenario generation. *Stochastic Environmental Research and Risk Assessment*, p. 1–10. doi:10.1007/s00477-017-1411-2.
- Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., Nagler, T., e Erhardt, T. *VineCopula: Statistical Inference of Vine Copulas*, 2015. URL <http://CRAN.R-project.org/package=VineCopula>. R package version 1.5.
- Sharma, A. e O'Neill, R. (2002). A nonparametric approach for representing interannual dependence in monthly streamflow sequences. *Water resources research*, 38(7):5–1–5–10.
- Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Universit? de Paris*, 8:229–231.
- Souto, M., Moreira, A., Veiga, A., Street, A., Garcia, J., e Epprecht, C. (2014). A high-dimensional VARX model to simulate monthly renewable energy supply. In *IEEE Power Systems Computation Conference*, p. 1–7, Wroclaw, Poland.
- Zachariah, M. e Reddy, M. J. (2013). Development of an entropy-copula-based stochastic simulation model for generation of monthly inflows into the hirakud dam. *ISH Journal of Hydraulic Engineering*, 19(3):267–275.