



Uma nova abordagem de cópulas vine para a simulação de cenários de vazões

Guilherme Pereira
FGV ENERGIA

Álvaro Veiga
DEE/PUC-Rio

18 de janeiro de 2018



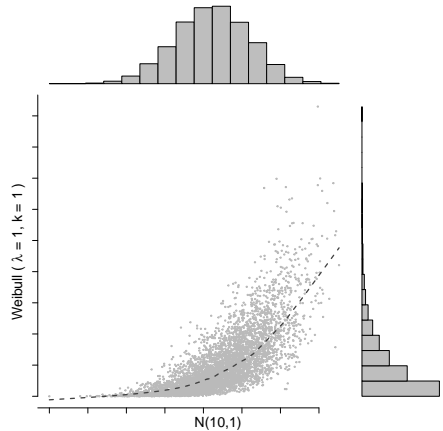
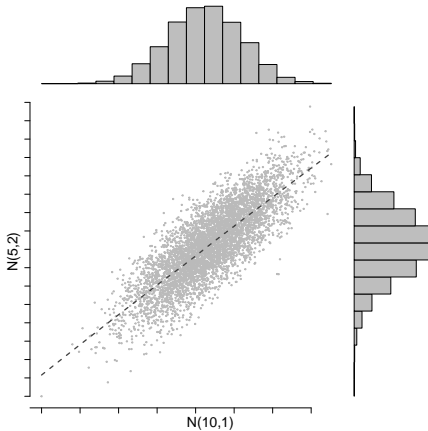
- 1 Cópulas
 - Motivação
 - Uma breve introdução
 - Construção pair-copula
 - Regular vine (R-vine)

- 2 Metodologia e aplicações
 - Univariado
 - Multivariado

- 3 Considerações finais

Cópuas: motivação

Essas variáveis possuem a mesma estrutura de dependência ?



Cópuas: uma breve introdução

Transformação Integral de Probabilidade

$$u_1 = F_1(x_1) \sim \text{Uniforme}(0, 1)$$

Cópuas: uma breve introdução

Transformação Integral de Probabilidade

$$u_1 = F_1(x_1) \sim \text{Uniforme}(0, 1)$$

Cópula d -dimensional: C

É uma *f.d.a* definida em $[0, 1]^d$, $C : [0, 1]^d \rightarrow [0, 1]$, com marginais uniformes.

Cópuas: uma breve introdução

Transformação Integral de Probabilidade

$$u_1 = F_1(x_1) \sim \text{Uniforme}(0, 1)$$

Cópuła d -dimensional: C

É uma *f.d.a* definida em $[0, 1]^d$, $C : [0, 1]^d \rightarrow [0, 1]$, com marginais uniformes.

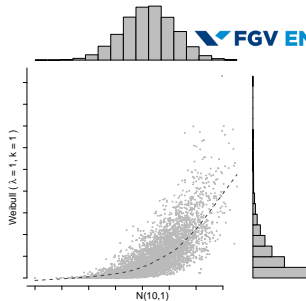
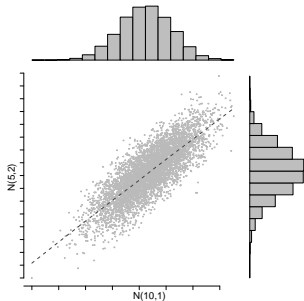
Teorema de Sklar (1959) - (caso bidimensional)

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)) = C(u_1, u_2)$$

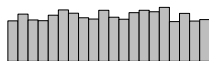
$$f(x_1, x_2) = c((F_1(x_1), F_1(x_2)))f_1(x_1)f_2(x_2)$$

O teo. de Sklar permite que separemos esta dist. bivariada em:

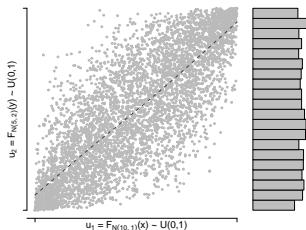
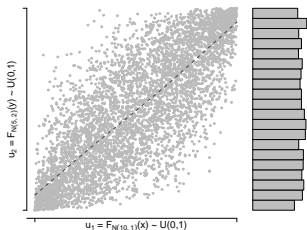
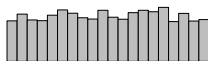
- um conjunto de distribuições marginais, $F_1(x_1)$ e $F_2(x_2)$.
- uma cópuła C que fornece toda a informação acerca da estrutura de dependência de $F(x)$.



$$\downarrow u_i = F_i(x), i=1,2 \downarrow$$



$$\downarrow u_i = F_i(x), i=1,2 \downarrow$$

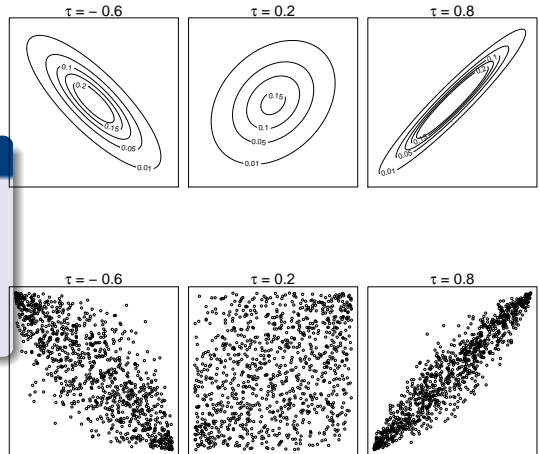


Essas variáveis possuem a mesma estrutura de dependência? **Sim.** 

Exemplos de cópulas bivariadas: Gaussiana

Propriedades:

- dep. **simétrica**.
- dep. negativa.
- dep. positiva.
- **ausência** de dep. de cauda.

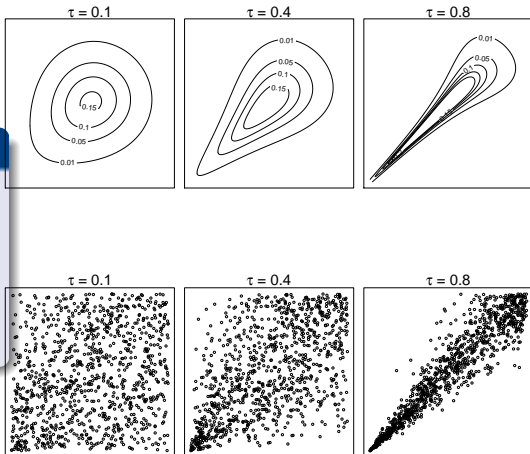


Exemplos de cópulas bivariadas: Clayton

Propriedades:

- dep. **assimétrica**.
- dep. positiva.
- dep. de cauda **inferior**, λ_I :

$$\lim_{q \rightarrow 0^+} P(X_2 \leq F_2^{-1}(q) | X_1 \leq F_1^{-1}(q))$$

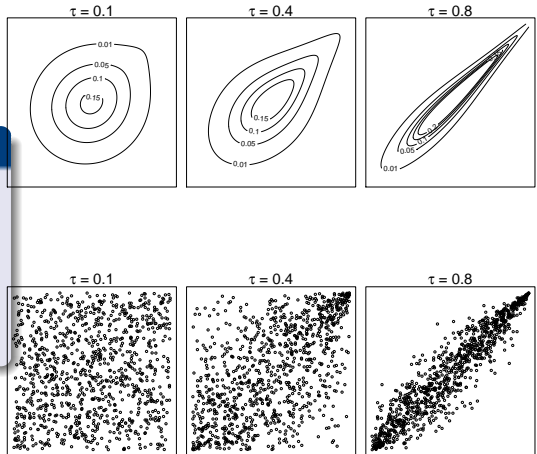


Exemplos de cópulas bivariadas: Gumbel

Propriedades:

- dep. **assimétrica**.
- dep. positiva.
- dep. de cauda **superior**, λ_U .

$$\lim_{q \rightarrow 1^-} P(X_2 > F_2^{-1}(q) | X_1 > F_1^{-1}(q))$$



Cópuas

Cópuas para $d > 2$.

- O número de cópuas d-dimensionais ($d > 2$) é reduzido.
- Possuem restrições que tornam o seu uso menos atrativo.
- Menor flexibilidade para modelar diferentes tipos de associação.

Construção *pair-copula*, Aas et al. (2009).

- Decompõe uma densidade multivariada $f(x) = f(x_1, \dots, x_d)$ em um produto de $d(d - 1)$ cópuas bivariadas (pair-cópuas) e d densidades marginais univariadas.

Construção *pair-copula*

Exemplo com 4 variáveis:

$$f(x_1, x_2, x_3, x_4) = f(x_1) f_{2|1}(x_2|x_1) f_{3|1,2}(x_3|x_1, x_2) f_{4|1,2,3}(x_4|x_1, x_2, x_3)$$

$$\begin{aligned} f_{2|1}(x_2|x_1) &= \frac{f_{12}(x_1, x_2)}{f_1(x_1)} \\ &= \frac{c_{12}(F_1(x_1), F_2(x_2)) f_1(x_1) f_2(x_2)}{f_1(x_1)} \quad \leftarrow \text{Teo. Sklar} \\ &= c_{12}(F_1(x_1), F_2(x_2)) f_2(x_2) \end{aligned}$$

Construção *pair-copula*

Exemplo com 4 variáveis:

$$f(x_1, x_2, x_3, x_4) = f(x_1) f_{2|1}(x_2|x_1) f_{3|1,2}(x_3|x_1, x_2) f_{4|1,2,3}(x_4|x_1, x_2, x_3)$$

$$\begin{aligned} f_{3|1,2}(x_3|x_1, x_2) &= \frac{f_{1,2,3}(x_1, x_2, x_3)}{f_{1,2}(x_1, x_2)} = \frac{f_{3,2|1}(x_3, x_2|x_1)}{f_{2|1}(x_2|x_1)} \frac{1}{f_{2|1}(x_2|x_1)} \\ &= \frac{c_{3,2|1}(F_{3|1}(x_3|x_1), F_{2|1}(x_2|x_1)) f_{3|1}(x_3|x_1) f_{2|1}(x_2|x_1)}{f_{2|1}(x_2|x_1)} \\ &= c_{3,2|1}(F_{3|1}(x_3|x_1), F_{2|1}(x_2|x_1)) f_{3|1}(x_3|x_1) \end{aligned}$$

De maneira similar...

$$f_{3|1}(x_3|x_1) = c_{31}(F_3(x_3), F_1(x_1)) f_3(x_3).$$

Assim...

$$f_{3|1,2}(x_3|x_1, x_2) = c_{3,2|1}(F_{3|1}(x_3|x_1), F_{2|1}(x_2|x_1)) c_{31}(F_3(x_3), F_1(x_1)) f_3(x_3)$$

Construção *pair-copula*

Exemplo com 4 variáveis:

$$f(x_1, x_2, x_3, x_4) = f(x_1) f_{2|1}(x_2|x_1) f_{3|1,2}(x_3|x_1, x_2) f_{4|1,2,3}(x_4|x_1, x_2, x_3)$$

$$\begin{aligned} f_{4|1,2,3}(x_4|x_1, x_2, x_3) &= c_{34|1,2}(F_{3|1,2}(x_3|x_1, x_2), F_{4|1,2}(x_4|x_1, x_2)) \\ &\cdot c_{24|1}(F_{2|1}(x_2|x_1), F_{4|1}(x_4|x_1)) \\ &\cdot c_{14}(F_1(x_1), F_4(x_4)) \cdot f_4(x_4) \end{aligned}$$

Assim, a decomposição completa é dada por:

$$\begin{aligned} f(x_1, x_2, x_3, x_4) &= \\ &f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot f_4(x_4) \cdot \\ &c_{12}(F_1(x_1), F_2(x_2)) \cdot c_{13}(F_1(x_1), F_3(x_3)) \cdot c_{14}(F_1(x_1), F_4(x_4)) \cdot \\ &c_{32|1}(F_{3|1}(x_3|x_1), F_{2|1}(x_2|x_1)) \cdot c_{24|1}(F_{2|1}(x_2|x_1), F_{4|1}(x_4|x_1)) \cdot \\ &c_{34|1,2}(F_{3|1,2}(x_3|x_1, x_2), F_{4|1,2}(x_4|x_1, x_2)). \end{aligned}$$

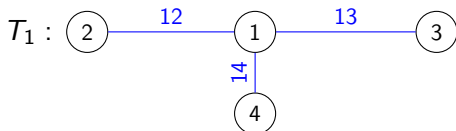
Cópuas: R-vine Bedford e Cooke (2001,2002).

Densidade

$$f = f_1 \cdot f_2 \cdot f_3 \cdot f_4 \\ \cdot c_{12} \cdot c_{13} \cdot c_{14}$$

Condição de proximidade

Se dois vértices estão conectados por uma aresta em T_ℓ , $\ell > 2$, as correspondentes arestas em $T_{\ell-1}$ devem dividir um vértice em comum.



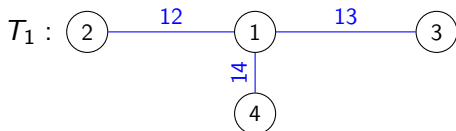
Cópuas: R-vine Bedford e Cooke (2001,2002).

Densidade

$$f = f_1 \cdot f_2 \cdot f_3 \cdot f_4 \\ \cdot c_{12} \cdot c_{13} \cdot c_{14}$$

Condição de proximidade

Se dois vértices estão conectados por uma aresta em T_ℓ , $\ell > 2$, as correspondentes arestas em $T_{\ell-1}$ devem dividir um vértice em comum.



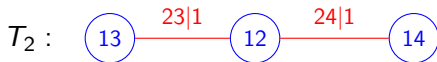
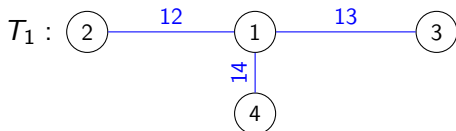
Cópuas: R-vine Bedford e Cooke (2001,2002).

Densidade

$$f = f_1 \cdot f_2 \cdot f_3 \cdot f_4 \\ \cdot c_{12} \cdot c_{13} \cdot c_{14} \\ \cdot c_{23|1} \cdot c_{24|1}$$

Condição de proximidade

Se dois vértices estão conectados por uma aresta em T_ℓ , $\ell > 2$, as correspondentes arestas em $T_{\ell-1}$ devem dividir um vértice em comum.



Cópuas: R-vine Bedford e Cooke (2001,2002).

Densidade

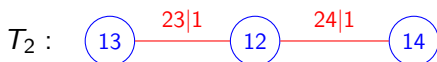
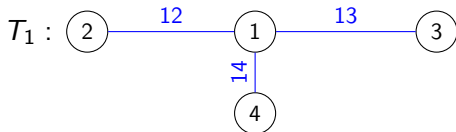
$$f = f_1 \cdot f_2 \cdot f_3 \cdot f_4$$

$$\cdot c_{12} \cdot c_{13} \cdot c_{14}$$

$$\cdot c_{23|1} \cdot c_{24|1}$$

Condição de proximidade

Se dois vértices estão conectados por uma aresta em T_ℓ , $\ell > 2$, as correspondentes arestas em $T_{\ell-1}$ devem dividir um vértice em comum.



Cópuas: R-vine Bedford e Cooke (2001,2002).

Densidade

$$f = f_1 \cdot f_2 \cdot f_3 \cdot f_4$$

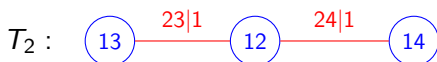
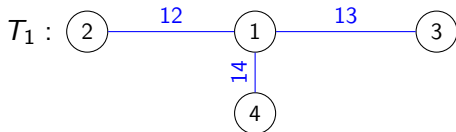
$$\cdot c_{12} \cdot c_{13} \cdot c_{14}$$

$$\cdot c_{23|1} \cdot c_{24|1}$$

$$\cdot c_{13|24}$$

Condição de proximidade

Se dois vértices estão conectados por uma aresta em T_ℓ , $\ell > 2$, as correspondentes arestas em $T_{\ell-1}$ devem dividir um vértice em comum.



- 1 Cópulas
 - Motivação
 - Uma breve introdução
 - Construção pair-copula
 - Regular vine (R-vine)

- 2 Metodologia e aplicações
 - Univariado
 - Multivariado

- 3 Considerações finais

Univariado

Stoch Environ Res Risk Assess
DOI 10.1007/s00477-017-1411-2



ORIGINAL PAPER

PAR(p)-vine copula based model for stochastic streamflow scenario generation

Guilherme Pereira¹ · Álvaro Veiga¹

© Springer-Verlag Berlin Heidelberg 2017

Abstract Synthetic streamflow data is vital for the energy sector, as it feeds stochastic optimisation models that determine operational policies. Considered scenarios should differ from each other, but be the same from a

1 Introduction

A unique characteristic of the Brazilian Electricity Sector is that most energy is generated by hydroelectric power

Contribuições e características do modelo:

- Alternativa ao modelo PAR(p).
- Efeitos não-lineares podem ser modelados.
- Cenários serão sempre positivos.
- Flexibilidade para modelar as distribuições marginais.

Univariado

Metodologia proposta - Ideia geral

- Para cada mês, uma dist. marginal é estimada: $u_t = F^m(y_t)$.
- Para cada mês, estimamos uma cópula d -dimensional.
- Para um mês qualquer, temos:

$$F^m(y_{t-d+1}, \dots, y_{t-1}, y_t) = C^m(u_{t-d+1}, \dots, u_{t-1}, u_t).$$

- Dimensão e cópulas variam de acordo com o período.

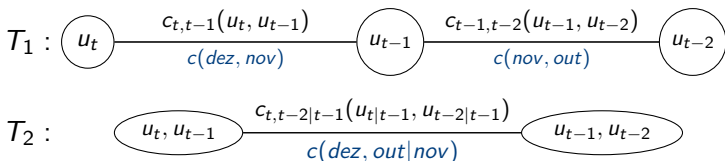


Figura: Exemplo de cóp. vine para um processo AR(2) - dezembro.

Univariado: estimação

Banco de dados e distribuições marginais

- Séries de vazões mensais de Jan-1931 até Dez-2015 - (ONS).
- Manso, Quebra-queixo e Garibaldi.
- Distribuições marginais mensais: **Gama**.

Tabela: Número de cópulas selecionadas (via BIC)

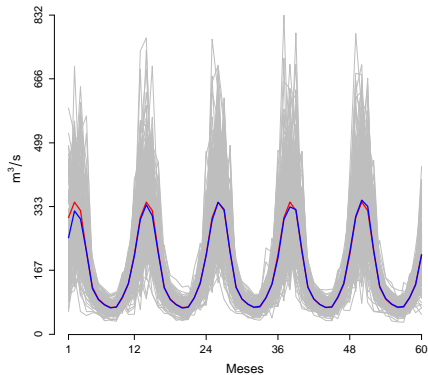
	Indep.	Normal	t-Stud.	Clayton	Gumbel	Frank
Manso	2	3	2	7	3	2
Quebra-queixo	4	2	0	9	0	2
Garibaldi	0	1	0	10	0	2

Tabela: Dimensões selecionadas para os meses

	Jan.	Fev.	Mar.	Abr.	Mai.	Jun.	Jul.	Ago.	Set.	Out.	Nov.	Dez.
Man.	2	2	3	2	4	4	2	3	2	2	2	2
Q.-q.	2	2	2	4	2	2	2	2	2	4	3	2
Gar.	2	2	2	3	2	2	2	2	2	2	2	2

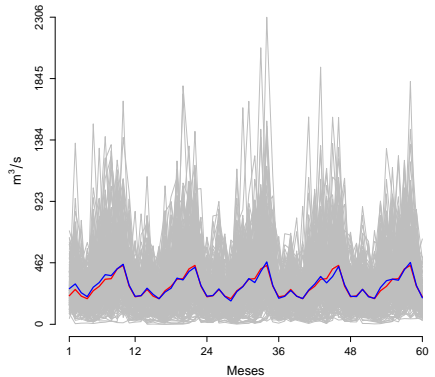
Univariado: simulação dos cenários

■ 200 cenários com 60 meses para cada série.



■ Cenários ■ Média dos cenários ■ Média histórica

(a) Manso

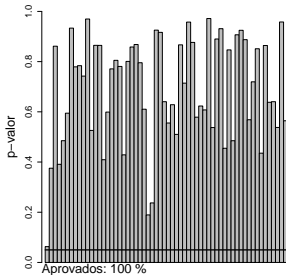


■ Cenários ■ Média dos cenários ■ Média histórica

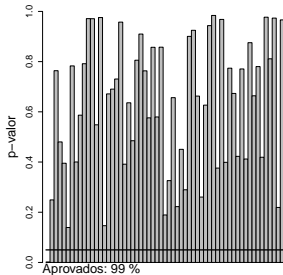
(b) Garibaldi

Univariado: análise dos cenários

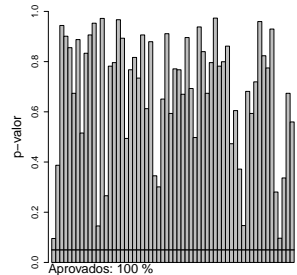
■ Testes t, Levene e Kolmogorov-Smirnov.



(a) t.



(b) Levene.



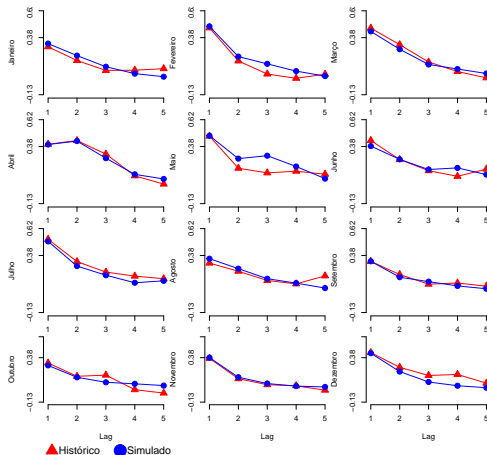
(b) K-S.

Tabela: Taxa de aprovação

	t	Levene	K-S
Manso	99%	99%	99%
Quebra-queixo	100%	99%	100%
Garibaldi	99%	100%	95%

Univariado: análise da estrutura de dep. temporal

- Comparação da auto-dependência de *lags* 1,...,5: Garibaldi.
- Simulamos 200 cenários com 100 anos cada.



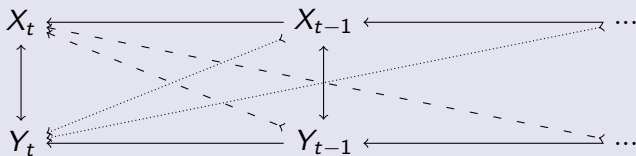
- 1 Cópulas
 - Motivação
 - Uma breve introdução
 - Construção pair-copula
 - Regular vine (R-vine)

- 2 Metodologia e aplicações
 - Univariado
 - **Multivariado**

- 3 Considerações finais

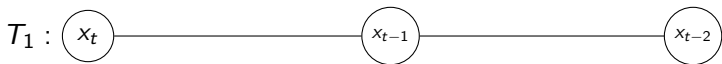
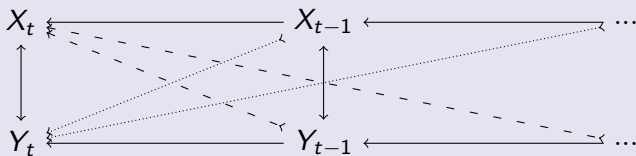
Multivariado

Modelo espaço-temporal



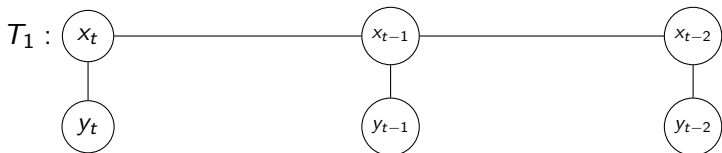
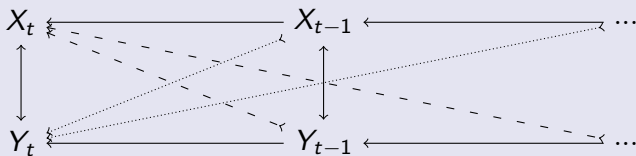
Multivariado

Modelo espaço-temporal



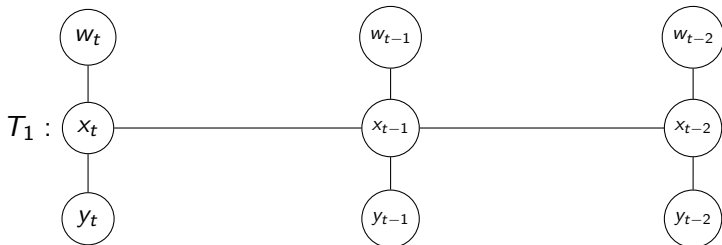
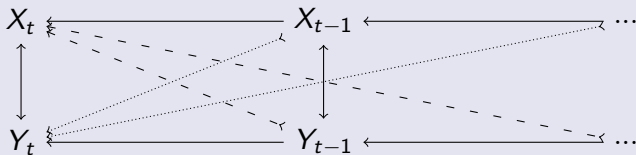
Multivariado

Modelo espaço-temporal



Multivariado

Modelo espaço-temporal



Multivariado

Metodologia proposta - Ideia geral

- Para cada mês e série uma dist. marginal é estimada.
- Para cada mês estimamos uma cópula de dimensão igual a $(p + 1)n$.
 - $p = 1$ (ordem autorregressiva)
 - $n =$ número de séries

Contribuições e características do modelo:

- Modelo não-linear para séries periódicas multivariadas de vazões.
- Alternativa ao modelo VAR(p) periódico.

Multivariado: estimação

Banco de dados e distribuições marginais

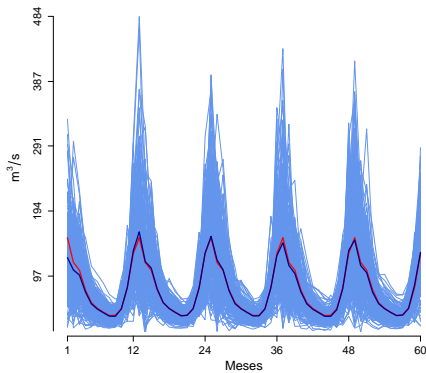
- 4 séries de vazões mensais - (ONS).
- Distribuições marginais mensais: **Gama**.

Tabela: Número de cópulas selecionadas (via BIC)

	Indep.	Normal	t-Stud.	Clayton	Gumbel	Frank
Total	124	65	75	17	32	23

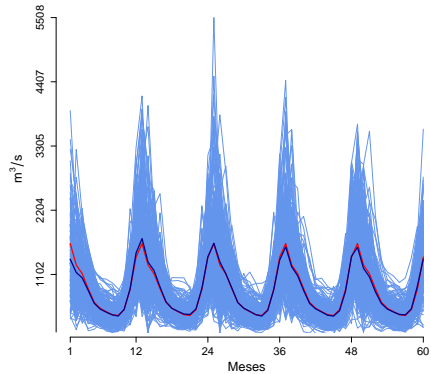
Multivariado: simulação dos cenários

■ Simulamos 200 cenários com 60 meses cada.



■ Cenários ■ Média dos cenários ■ Média histórica

(a) Vazão 1.

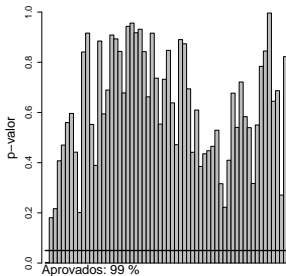


■ Cenários ■ Média dos cenários ■ Média histórica

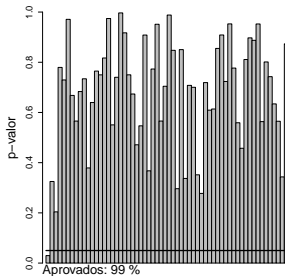
(b) Vazão 4.

Multivariado: análise dos cenários

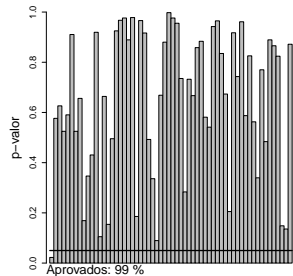
- Testes t, Levene e Kolmogorov-Smirnov.



(a) t.



(b) Levene.



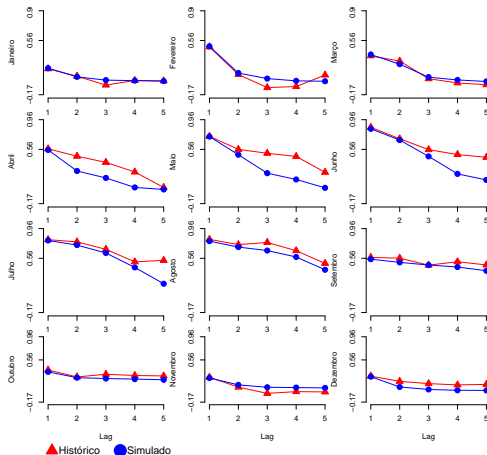
(b) K-S.

Tabela: Taxa de aprovação

	t	Levene	K-S
Vazão 1	99%	99%	99%
Vazão 2	99%	100%	99%
Vazão 3	99%	89%	99%
Vazão 4	99%	89%	99%

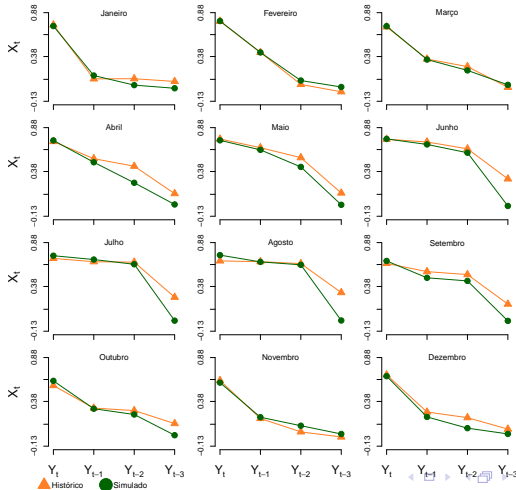
Multivariado: análise da estrutura de dep. temporal

- Comparação da auto-dependência de *lags* 1,...,5: Vazão 1.
- 200 cenários de 100 anos cada.



Multivariado: análise da dependência espacial

- Comparação da dependência espacial: Vazão 1 \leftarrow Vazão 4.
- 200 cenários com 100 anos cada.



Considerações finais

Principais resultados:

- Modelo não-linear para séries periódicas de vazões.
- Alternativas aos modelos PAR(p) e VAR(p) periódico.
- Efeitos não-lineares podem ser modelados.
- Cenários serão sempre positivos.
- Flexibilidade para modelar as distribuições marginais.
- Modelos são capazes de replicar nos cenários as estruturas de dependência temporal e espacial observadas nas séries históricas.

Considerações finais

Alguns trabalhos já publicados:

- G. Pereira, A. Veiga, (2017), *PAR(P)-vine copula based model for stochastic streamflow scenario generation*, **Stochastic Environmental Research and Risk Assessment**.
<https://doi.org/10.1007/s00477-017-1411-2>
- G. Pereira, C. Czado, T. Erhardt, A. Veiga, (2017), *A periodic spatial vine copula model for multi-site streamflow simulation*, **Electric Power Systems Research**, v. 152, p. 9-17.
<https://doi.org/10.1016/j.epsr.2017.06.017>
- G. Pereira, C. Czado, T. Erhardt, A. Veiga, (2016), *Spatial R-vine copula for streamflow scenario simulation*, **Power Systems Computation Conference (PSCC)**, doi: 10.1109/PSCC.2016.7540939.
<http://ieeexplore.ieee.org/document/7540939>.



fgvenergia.fgv.br

Guilherme Pereira
FGV Energia
guilherme.pereira@fgv.br

Referências

- Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependences. *Insurance: Mathematics and Economics*, 44(2):182–198.
- Bedford, T. and 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. and Cooke, R. M. (2002). Vines: A new graphical model for dependent random variables. *Annals of Statistics*, 30(4):1031–1068.
- Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris*, 8:229–231.

Definição - Vine regular

Uma *regular vine* (R-vine) d -dimensional $\mathcal{V} = (T_1, T_2, \dots, T_{d-1})$ é uma sequência $d-1$ árvores que respeita as seguintes condições:

- i. T_1 é uma árvore com d vértices representando as variáveis X_1, \dots, X_d e $d - 1$ arestas.
- ii. T_ℓ possui $d - j + 1$ vértices e $d - j$ arestas.
- iii. Arestas em T_ℓ são vértices em $T_{\ell+1}$
- iv. Dois vértices em T_ℓ só podem ser conectados se as correspondentes arestas em $T_{\ell-1}$ possuem um vértice em comum.

voltar