



Guilherme Armando de Almeida Pereira

**Modelos de cópulas para simulação de cenários
hidrológicos**

Tese de Doutorado

Tese apresentada ao Programa de Pós-graduação em Engenharia Elétrica da PUC-Rio como requisito parcial para obtenção do grau de Doutor em Engenharia Elétrica.

Orientador: Prof. Álvaro de Lima Veiga Filho

Rio de Janeiro
Fevereiro de 2017



Guilherme Armando de Almeida Pereira

Modelos de cópulas para simulação de cenários hidrológicos

Tese apresentada como requisito parcial para obtenção do grau de Doutor pelo Programa de Pós-Graduação em Engenharia Elétrica da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Álvaro de Lima Veiga Filho

Orientador

Departamento de Engenharia Elétrica – PUC-Rio

Prof. Cristiano Augusto C. Fernandes

Departamento de Engenharia Elétrica – PUC-Rio

Prof. Alexandre Street de Aguiar

Departamento de Engenharia Elétrica – PUC-Rio

Prof. Gerson Couto de Oliveira

PSR consultoria

Profa. Beatriz Vaz de Melo Mendes

UFRJ

Prof. Márcio da Silveira Carvalho

Coordenador Setorial do Centro

Técnico Científico – PUC-Rio

Rio de Janeiro, 08 de Fevereiro de 2017

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Guilherme Armando de Almeida Pereira

Graduou-se em Ciências Econômicas pela Universidade Federal de Juiz de Fora. Fez mestrado no Departamento de Engenharia Elétrica da PUC-Rio, especializando-se em de séries temporais. Atualmente finaliza seu doutorado na PUC-Rio em Métodos de Apoio à Decisão. Durante o período de 2014/2015, foi pesquisador visitante (doutorado sanduíche) no Departamento de Matemática da Universidade Técnica de Munique, Alemanha. Suas áreas de interesse incluem cópulas, modelos não lineares, modelos em alta dimensão e modelos para simulação de cenários.

Ficha Catalográfica

Pereira, Guilherme Armando de Almeida

Modelos de cópulas para simulação de cenários hidrológicos / Guilherme Armando de Almeida Pereira; orientador: Álvaro de Lima Veiga Filho. – 2017.

v., 131 f: il. color. ; 30 cm

Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica.

Inclui bibliografia

1. Engenharia Elétrica – Teses. 2. Modelos não lineares;. 3. Cópulas;. 4. Cópulas vine;. 5. Simulação de cenários;. 6. Modelos em alta dimensão.. I. Veiga Filho, Álvaro de Lima. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. III. Título.

CDD: 621.3

Agradecimentos

A Deus pela constante presença.

Aos meus pais, Maria Fátima e Armando Eugênio, pelo apoio, dedicação e pela presença constante em todos os momentos dessa trajetória.

Ao meu irmão Rodolfo Armando, pela sua cumplicidade, apoio e ajuda.

Ao Prof. Álvaro Veiga, pela dedicação, paciência e pelo grande incentivo ao longo de todos esses anos. Sua orientação nos momentos difíceis foi de grande valia e importância para a condução desse trabalho.

A todos os meus amigos da PUC-Rio, em especial, Eduardo Dessupoio, Fernando Cyrino, Hugo Baldioti, Pedro Guilherme, Rafael Morais e Victor Lage pelos momentos agradáveis, descontraídos e felizes que compartilhamos nesta caminhada.

À CAPES e ao CNPq, pelo apoio financeiro.

À PUC-Rio, pelo apoio durante as atividades exercidas no decorrer do curso.

Aos membros da banca por aceitarem o convite, pelas contribuições e comentários.

Last but not least, I would like to express my gratitude to Prof. Claudia Czado for accepting me at Technical University of Munich and introducing me to the world of vine copula. I appreciate all the encouragement I have received from her. I am also grateful to Tobias Erhardt for the support and enriching discussions on spatial copulas.

Resumo

Pereira, Guilherme Armando de Almeida; Veiga Filho, Álvaro de Lima. **Modelos de cópulas para simulação de cenários hidrológicos.** Rio de Janeiro, 2017. 131p. Tese de Doutorado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Muitos dos modelos de simulação de cenários de vazões, necessários para o planejamento e operação de setores elétricos, são construídos sob hipóteses rígidas. Isto pode restringir sua capacidade de representar dependências não-lineares e/ou distribuições não usuais. Cópulas superam estas limitações. Elas possibilitam que o comportamento marginal das variáveis seja modelado separadamente da estrutura de dependência do vetor aleatório. Além do mais, podem representar os mais diversos tipos de associações. Isto posto, esta tese apresenta 3 artigos onde modelos de cópulas são desenvolvidos visando a simulação de cenários de vazões. No primeiro artigo, propomos um modelo periódico de cópulas vine espaciais para simulação multivariada. As principais contribuições são a extensão para o caso periódico dos modelos de cópulas vine espaciais; a drástica redução do número de parâmetros; o desenvolvimento de um modelo não linear multivariado para simulação de cenários de vazões que incorpora a dependência temporal, a dependência espacial, a variação sazonal e o elevado número de usinas (alta dimensão). No segundo artigo, realizamos algumas modificações no modelo periódico espacial proposto que resultam em uma menor complexidade sem perda de performance. No terceiro artigo, propomos uma metodologia baseada em cópulas vine para modelar a dependência temporal de séries periódicas univariadas de vazões. Dentre as contribuições destaca-se a construção de uma versão não-linear dos modelos periódicos autorregressivos (PAR(p)) onde a dependência temporal de qualquer ordem pode ser considerada; a possibilidade da incorporação de efeitos lineares e não-lineares; um modelo que não simula cenários com valores negativos; a flexibilidade para se modelar as distribuições marginais mensais.

Palavras-chave

Modelos não lineares; Cópulas; Cópulas vine; Simulação de cenários; Modelos em alta dimensão.

Abstract

Pereira, Guilherme Armando de Almeida; Veiga Filho, Álvaro de Lima (Advisor). **Copula models for streamflow scenario simulation**. Rio de Janeiro, 2017. 131p. PhD Thesis – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Many streamflow scenario simulation models, which are needed for the planning and operation of energy systems, are built on rigid assumptions. This may limit their ability to represent nonlinear dependencies and/or nonstandard distribution functions. Copulas overcome these drawbacks and represent a flexible tool for modeling multivariate distributions. They enable the modeling of the marginal behavior of variables separately from the dependence structure of a random vector. Moreover, they can represent any type of association. This thesis is composed of three working papers, wherein copula-based models are proposed, objectifying the simulation of streamflow scenarios. In the first working paper, a periodic spatial copula model is proposed to simulate multivariate streamflow scenarios. The main contributions include periodic extension of the spatial vine copulas; a distinct reduction in the number of parameters; and the development of a multivariate nonlinear model for streamflow scenario generation that incorporates time dependence, spatial dependence, and seasonal variation, and accounts for the dimensionality of the problem (high number of hydroelectric power plants). In the second working paper, some modifications are made to the periodic spatial model, resulting in lower complexity without the loss of performance. In the third working paper, a methodology based on the vine copula is proposed to model the temporal dependence structures in a univariate periodic streamflow time series. Among the contributions, the construction of a nonlinear version of the periodic autoregressive model (PAR(p)) is highlighted. The possibility of modeling linear and nonlinear effects and the flexibility of modeling the monthly marginal distributions are highlighted as well. This model does not simulate negative values.

Keywords

Non-linear models; Copula; Vine copula; Scenario simulation; High dimensional dependence models.

Sumário

1	Introdução	12
2	Algumas ferramentas matemáticas e cópulas	17
2.1	Cópula	17
2.1.1	Definição	17
2.1.2	Construção pair-cópula	19
2.2	Medidas de dependência	22
2.2.1	Correlação linear	23
2.2.2	Tau de Kendall	24
2.2.3	Dependência de cauda	26
2.3	Algumas cópulas bivariadas	26
2.4	Transformada de Fisher	34
2.5	Modelo de séries temporais - PAR (p)	34
2.5.1	Estimação	35
2.5.2	Simulação de cenários	37
2.6	Conceitos básicos sobre teoria dos grafos	37
3	Vines regulares	40
3.1	Distribuição Rvine	40
3.2	Matrizes vine regular	46
3.3	Famílias de cópulas e seus parâmetros em forma matricial	49
3.4	Simulação	52
4	R-vine espaciais para simulação de cenários hidrológicos	56
4.1	Motivação	56
4.2	Cópulas vine espaciais	58
4.2.1	Estimação do modelo de cópula vine espacial	61
4.3	Modelo periódico de cópula vine espacial	62
4.4	Modelo periódico de cópulas vine aplicados a séries de vazão	63
4.4.1	Transformando variáveis originais em uniformes	63
4.4.2	Definição da estrutura e seleção das pair-cópulas	65
4.4.3	Componentes espaciais	68
4.4.3.1	Distância	68
4.4.3.2	Precipitação	69
4.4.3.3	Topografia dos rios	71
4.4.4	Definindo equação espacial para a T_1	71
4.4.5	Modelo proposto	76
4.5	Resultados	78
4.5.1	Análise da dependência espacial	78
4.5.2	Análise dos cenários simulados	86
4.6	Algumas alternativas para o modelo desenvolvido.	96
4.7	Conclusões do capítulo	102
5	Cópulas periódicas para simulação de cenários hidrológicos univariados	104

5.1	Motivação	104
5.2	Modelo periódico de cópula vine para séries de vazões	108
5.2.1	Simulação de cenários	108
5.3	Estudo de caso	109
5.3.1	Análise do modelo estimado	112
5.3.2	Simulação dos cenários	115
5.4	Conclusões do capítulo	120
	Referências bibliográficas	121
A	Lista das usinas hidrelétricas utilizadas.	127
B	Figuras suplementares: árvore T_1 estimada no Capítulo 4.	128

Lista de figuras

Figura 2.1 -	Exemplo de pares concordantes e discordantes.	24
Figura 2.2 -	Gráfico de pontos e contorno: cópula Gaussiana.	27
Figura 2.3 -	Gráfico de pontos e contorno: cópula t.	29
Figura 2.4 -	Gráfico de pontos e contorno: cópula Clayton.	30
Figura 2.5 -	Gráfico de pontos e contorno: cópula Gumbel.	31
Figura 2.6 -	Gráfico de pontos e contorno: cópula Frank.	32
Figura 2.7 -	Versões rotacionadas das cópulas Clayton e Gumbel.	33
Figura 2.8 -	Exemplo de grafo com 6 vértices e 6 arestas.	38
Figura 2.9 -	Exemplo de grafo com vértice isolado.	39
Figura 2.10 -	Exemplo de uma árvore com 9 vértices 8 arestas.	39
Figura 3.1 -	Vine regular com 4 dimensões.	41
Figura 3.2 -	Exemplos de C-vine e D-vine.	42
Figura 3.3 -	R-vine quadridimensional utilizando $\mathcal{C}_{e,a}, \mathcal{C}_{e,b} \mathcal{D}_e$.	44
Figura 3.4 -	Vine regular com 5 variáveis.	50
Figura 3.5 -	R-vine com 3 variáveis e <i>h-functions</i> .	54
Figura 4.1 -	Exemplo de associação entre $\ln(d_{ij})$ e $g_z(\hat{\tau}_{ij})$.	59
Figura 4.2 -	Exemplo de reparametrização para a árvore T_1 .	60
Figura 4.3 -	Teste de Anderson-Darling para cada uma das 39 usinas.	64
Figura 4.4 -	τ de Kendall por árvore (T_ℓ).	66
Figura 4.5 -	Graus de liberdade das cópulas t contidas na T_1 .	67
Figura 4.6 -	Representação da T_1 sobre o mapa brasileiro.	67
Figura 4.7 -	Associação entre $\log(\text{Dist})$ e o $g_z(\tau_{ij,m}^{hidro})$.	69
Figura 4.8 -	Usinas hidrelétricas e estações meteorológicas.	70
Figura 4.9 -	Associação entre $g_z(\tau_{ij,m}^{hidro})$ e o $g_z(\tau_{ij}^{chuva})$.	70
Figura 4.10 -	$\text{Log}(\text{Dist})$ e $g_z(\tau_{ij,jan}^{hidro})$ para os pares de usinas da árvore T_1 .	71
Figura 4.11 -	$G_z(\tau_{ij,jan}^{hidro})$ e $g_z(\tau_{ij}^{chuva})$ para os pares de usinas da árvore T_1 .	72
Figura 4.12 -	Análise das variáveis espaciais para T_2 .	73
Figura 4.13 -	Análise das variáveis espaciais para T_3 .	74
Figura 4.14 -	Análise das variáveis espaciais para T_4 .	75
Figura 4.15 -	Matrizes $\hat{\tau}_m^{emp}$ e $\bar{\tau}_m^{sim}$: janeiro e fevereiro.	79
Figura 4.16 -	Matrizes $\hat{\tau}_m^{emp}$ e $\bar{\tau}_m^{sim}$: março, abril e maio.	80
Figura 4.17 -	Matrizes $\hat{\tau}_m^{emp}$ e $\bar{\tau}_m^{sim}$: junho, julho e agosto.	81
Figura 4.18 -	Matrizes $\hat{\tau}_m^{emp}$ e $\bar{\tau}_m^{sim}$: setembro, outubro e novembro.	82
Figura 4.19 -	Matrizes $\hat{\tau}_m^{emp}$ e $\bar{\tau}_m^{sim}$: dezembro.	83
Figura 4.20 -	Absoluto da diferença entre $\hat{\tau}_m^{emp}$ e $\bar{\tau}_m^{sim}$ para todos os meses.	83
Figura 4.21 -	Absoluto da diferença ($ \hat{\tau}_m^{emp} - \bar{\tau}_m^{sim} $): pares presentes na T_1 .	84
Figura 4.22 -	Varição sazonal espacial para alguns pares de usinas - I.	85
Figura 4.23 -	Varição sazonal espacial para alguns pares de usinas - II.	86
Figura 4.24 -	Percentual de cenários válidos.	87
Figura 4.25 -	Cenários simulados para as usinas 1, 2, 5, 9, 12 e 16.	88
Figura 4.26 -	Cenários simulados para as usinas 20, 24, 28, 30, 34 e 39.	89
Figura 4.27 -	Teste t: usinas 1 e 5.	90
Figura 4.28 -	Teste t: usinas 9, 16, 20, 28, 34 e 39.	91

Figura 4.29 - Levene: usinas 1 e 5.	92
Figura 4.30 - Levene: usinas 9, 16, 20, 28, 34 e 39.	93
Figura 4.31 - K-S: usinas 1 e 5.	94
Figura 4.32 - K-S: usinas 9, 16, 20, 28, 34 e 39.	95
Figura 4.33 - Modelo alternativo: matrizes $\hat{\tau}_m^{emp}$ e $\bar{\tau}_m^{sim}$ para janeiro e junho.	99
Figura 4.34 - Modelo alternativo: variação sazonal espacial.	100
Figura 4.35 - Modelo alternativo: cenários para as usinas 1, 11, 22 e 25.	101
Figura 5.1 - Exemplo de uma D-vine com 4 variáveis.	106
Figura 5.2 - Série de vazão do rio Manso.	109
Figura 5.3 - Teste de Anderson-Darling para cada um dos 12 meses.	110
Figura 5.4 - Vazão do rio Manso e vazão ajustada pelo modelo.	112
Figura 5.5 - Auto-dependência dos resíduos.	113
Figura 5.6 - Auto-dependência dos resíduos de cada mês.	114
Figura 5.7 - Cenários simulados.	115
Figura 5.8 - Análise dos cenários simulados.	116
Figura 5.9 - Comparação da auto-dependência de <i>lags</i> 1, 2, 3, 4 e 5.	117
Figura 5.10 - Box-plot para auto-dependência de <i>lags</i> 1, 2, 3, 4 e 5.	119
Figura B.1 - A árvore T_1 sobre o mapa brasileiro.	128
Figura B.2 - Ilustração da árvore T_1 sobre a Bacia 1 (Rio Amazonas).	129
Figura B.3 - Parte da árvore T_1 sobre a Bacia 2 (Rio Tocantins).	129
Figura B.4 - Parte de T_1 sobre a Bacia 3 (Atlântico, Norte/Nordeste).	130
Figura B.5 - Parte de T_1 sobre a Bacia 5 (Atlântico, Leste).	130
Figura B.6 - Parte da árvore T_1 sobre a Bacia 6 (Rio Paraná).	131

Lista de tabelas

Tabela 3.1 - Conjunto restrição da matriz \mathcal{M}^* .	47
Tabela 3.2 - Elementos de $B_M(j)$ utilizando como exemplo \mathcal{M}^* .	48
Tabela 3.3 - Elementos de $\tilde{B}_M(j)$ utilizando como exemplo \mathcal{M}^* .	48
Tabela 3.4 - Elementos $(m_{k,i}, \{m_{k+1,i}, \dots, m_{d,i}\})$ da \mathcal{M}^* .	49
Tabela 4.1 - Famílias de cópulas selecionadas para $T_\ell, \ell = 1, \dots, 10$.	66
Tabela 4.2 - Valores iniciais e finais do modelo estimado.	77
Tabela 4.3 - Log-verossimilhança por árvore.	77
Tabela 4.4 - Absoluto da diferença entre $\hat{\tau}_m^{emp}$ e $\bar{\tau}_m^{sim}$.	83
Tabela 4.5 - Absoluto da diferença $(\hat{\tau}_m^{emp} - \bar{\tau}_m^{sim})$: pares presentes na T_1 .	84
Tabela 4.6 - Análise dos cenários: Teste t, Levene e Kolmogorov-Smirnov.	96
Tabela 4.7 - Regressão estimada com as variáveis espaciais.	97
Tabela 4.8 - Parâmetros da equação espacial estimada.	98
Tabela 4.9 - Modelo alternativo: log-verossimilhança por árvore.	98
Tabela 4.10 - Modelo alternativo: $ \hat{\tau}_m^{emp} - \bar{\tau}_m^{sim} $ para os pares de usinas da T_1 .	99
Tabela 4.11 - Modelo alternativo: absoluto da diferença entre $\hat{\tau}_m^{emp}$ e $\bar{\tau}_m^{sim}$.	100
Tabela 4.12 - Modelo alternativo: análise dos cenários.	102
Tabela 5.1 - Número de cópulas selecionadas.	112
Tabela 5.2 - Dimensões selecionadas para cada mês.	112
Tabela 5.3 - Auto-dependência mensal: valores históricos e simulados.	118

1

Introdução

No Brasil, grande parte da energia gerada provém de usinas hidrelétricas. Esta característica peculiar torna o setor elétrico fortemente dependente dos regimes hidrológicos e de chuvas.

Devido ao elevado grau de incerteza associado a tais regimes, o planejamento e a operação do sistema é uma tarefa bastante complexa. Sua solução é obtida via modelos de otimização baseados em programação estocástica. Grande parte destes, são alimentados por cenários de vazões que, de certa forma, representam a incerteza associada às fontes intermitentes. Deste modo, é clara a importância dos modelos de simulação de cenários para o setor elétrico.

A acurácia de um modelo de simulação de cenários pode ser aferida através de sua habilidade em gerar dados sintéticos de vazões preservando determinadas características históricas observadas.

Em sua maioria, tais modelos de simulação são baseados nos modelos autorregressivos de Box e Jenkins [1]. Como exemplo, podemos citar as referências [2–7]. Tais modelos são construídos sob hipóteses rígidas no que diz respeito ao tipo de dependência temporal entre vazões adjacentes ou à forma das distribuições marginais ou das distribuições conjuntas, [8]. Sharma & O'Neill [8] e Zachariah & Reddy [9] elencam algumas desvantagens dessa classe de modelos, tais como: (i) Dados de vazões geralmente não seguem uma distribuição normal. Isto sugere que a distribuição Gaussiana, usualmente implícita na estrutura desses modelos, pode não ser adequada; (ii) A hipótese de normalidade limita a capacidade do modelo em representar formas não usuais de distribuições de probabilidade; (iii) Estas formulações podem não modelar apropriadamente estruturas de dependência não lineares; (iv) Existe a possibilidade de geração de cenários que apresentem valores negativos (não válidos), uma vez que a distribuição Gaussiana é definida no intervalo $(-\infty, \infty)$. Algumas transformações nas séries originais podem ser aplicadas para garantir que os cenários sejam positivos. Entretanto, estas podem incorporar algum tipo de viés, [10, 11].

Sabe-se que com o uso de cópulas, todas as desvantagens supracitadas podem ser contornadas. Desde o trabalho pioneiro de Sklar [12], onde foi de-

monstrado 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 assimetria, dependência de cauda, etc.

O desenvolvimento das cópulas ocorreu principalmente no âmbito bidimensional. Existe uma vasta coleção de cópulas bivariadas o que resulta em grande flexibilidade no que diz respeito as estruturas de dependência. Por outro lado, no caso d -dimensional ($d > 2$), o número de cópulas é reduzido. Além disso, as poucas cópulas existentes possuem restrições em suas formulações que limitam sua capacidade de modelar diferentes estruturas de dependência presentes dentro de um conjunto de dados.

A construção pair-cópula (*pair copula construction* - PCC; [13]) é uma alternativa para a construção de cópulas d -dimensionais que incorpora toda a flexibilidade das cópulas bivariadas. Esta consiste em decompor uma densidade multivariada em um conjunto de cópulas bivariadas incondicionais e condicionais e densidades marginais univariadas. Esta construção, que pode ser organizada em um modelo gráfico proposto por Bedford & Cooke [14, 15], também é conhecida na literatura como cópulas vine ou vines regulares (R-vines).

Modelos de simulação de cenários de vazões podem ser classificados em duas categorias, *multi-site simulation* e *single-site simulation*. Na primeira, várias séries são modeladas levando-se em consideração tanto a dependência temporal quanto a interdependência entre as usinas. No segundo caso, apenas a dependência temporal é considerada. Existem alguns trabalhos empregando cópulas para a simulação de cenários de vazões, [9, 11, 16–22]. Todavia, seu potencial ainda não foi completamente explorado. Assim, o objetivo geral desta tese é propor modelos/metodologias baseadas em cópulas para a simulação de cenários de vazões em ambas categorias.

Em termos de simulação multivariada (*multi-site simulation*), Hao & Sing [19] propuseram um método baseado na teoria de entropia para obtenção da cópula entre diferentes usinas. Os autores avaliaram a metodologia proposta em um conjunto de 3 usinas. Os resultados indicaram que o modelo é capaz de capturar adequadamente a dependência entre usinas de diferentes localidades. Chen et al. [22] também propuseram uma metodologia para simulação de vazões entre diversas localidades. Sua metodologia é baseada na construção de cópulas tridimensionais para modelar a dependência temporal e espacial entre as vazões de duas localidades. Sua abordagem modela pares de localidades

a cada iteração, o que pode resultar em uma não captura adequada da dependência espacial caso o número de regiões seja elevado. Os autores avaliaram esta metodologia na vazão mensal de 3 localidades e nas vazões diárias de 6 localidades. Os resultados indicaram que a metodologia proposta foi capaz simular cenários respeitando características observadas nas séries originais.

Isto posto, no Capítulo 4 desenvolvemos um modelo multivariado não linear baseado nas cópulas vine para geração simultânea de cenários mensais para diversas usinas hidrelétricas. Especialmente no caso do SEB onde o número de usinas é elevado, o modelo proposto deve ser capaz de lidar com o elevado número de usinas.

Embora bastante flexíveis para a construção de cópulas em altas dimensões, o número de parâmetros de uma cópula vine cresce quadraticamente com o número de variáveis. Como alternativa, Erhardt et al. [23] introduziram os modelos de cópulas vine espaciais. Estes exploram informações geográficas com o intuito de reduzir o número de parâmetros a serem estimados. Assim é possível construir um modelo com muitas variáveis estimando-se uma quantidade reduzida de parâmetros. Ademais, é possível aumentar o número de variáveis sem a necessidade de se elevar a quantidade de parâmetros.

Desta forma, no Capítulo 4 propomos um modelo de cópulas vine espaciais para simulação de cenários de vazões mensais para diversas usinas hidrelétricas. Tal modelo considera a dependência temporal, a dependência espacial entre as usinas e a forte variação sazonal. Para acomodar esta variação periódica, propomos uma extensão do modelo original de Erhardt et al. [23].

Dentre as contribuições deste capítulo, destaca-se a extensão das cópulas vine espaciais para o caso periódico. O modelo resultante pode ser entendido como um modelo de cópulas vine espaciais com parâmetros variando periodicamente. Além disso, apresentamos um conjunto de variáveis que descrevem razoavelmente a dependência espacial entre as usinas: (i) distância entre usinas; (ii) grau de associação entre séries de precipitações; (iii) topografia dos rios. A reparametrização com estas variáveis reduz consideravelmente o número de parâmetros a serem estimados. Além disso, o modelo proposto incorpora a dependência temporal, a dependência espacial, o forte componente periódico e o elevado número de usinas. Os resultados apontam que o modelo estimado captura as principais associações existentes no histórico. Com isso, a variação sazonal entre usinas é razoavelmente aproximada. Por fim, a análise dos cenários simulados indica que as séries sintéticas geradas respeitam características históricas presentes nas séries originais.

No Capítulo 5 a simulação de cenários de vazões univariadas (*single-site*

simulation) é abordada. É de conhecimento comum que cópulas também podem ser empregadas para modelar a dependência temporal dentro de uma série temporal. Como exemplos, pode-se citar as representações através de cópulas dos modelos $AR(p)$, $MA(q)$ e $ARMA(p,q)$ apresentadas em Joe [24]. Mendes & Aíube [25] modelaram o primeiro e segundo momento de séries financeiras. Este modelo pode ser entendido como uma alternativa ao $AR(1)$ - $ARCH(1)$. Mendes & Accioly [26] propuseram uma estimação robusta para C-vines e as utilizaram para previsão de séries de volatilidade realizada. Brechmann & Czado [27] propuseram o COPAR (*Copula autoregressive model*), o que pode ser visto como uma alternativa ao VAR.

Em se tratando de modelos de cópulas para geração univariada de cenários de vazões, Lee & Salas [17] foram os pioneiros. Estes utilizaram cópulas bivariadas e a cópula Gaussiana trivariada para simulação e análise da série anual do Rio Nilo. Hao & Singh [18] 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 & Reddy [9] utilizaram a cópula Gumbel bivariada para a simulação de cenários de vazões mensais na barragem de Hirakud, Índia. Kong et al. [21] 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. [11] estimaram a distribuição conjunta entre dois meses consecutivos condicionadas a outras variáveis. Jeong & Lee [20] aliaram a abordagem proposta por Lee & Salas [17] ao modelo PDAR (*periodic discrete autoregressive model*) para a simulação de vazões mensais de rios intermitentes.

Como observado, os trabalhos para simulação univariada de cenários hidrológicos se preocuparam apenas com a dependência temporal de lag 1. Muitos destes, podem ser entendidos como uma versão não linear do tradicional modelo periódico autorregressivo Gaussiano de ordem 1, $PAR(1)$.

Dessa forma, no Capítulo 5 propomos uma metodologia baseada em cópulas vine para tratar a dependência temporal de ordens superiores a 1 em séries hidrológicas periódicas. O modelo resultante é uma alternativa ao tradicional modelo periódico autorregressivos de ordem p ($PAR(p)$, [5]) e pode ser visto como uma versão não linear do mesmo. Além do mais, a abordagem proposta contorna algumas limitações dos modelos lineares Gaussianos. Ela não simula cenários negativos, é flexível o suficiente para modelar efeitos não-lineares e pode modelar qualquer comportamento das distribuições marginais. A análise dos cenários simulados indica que o modelo é capaz de captar estruturas de dependência de ordem superior a 1. Por fim, testes estatísticos realizados demonstram que esta metodologia é capaz de simular cenários com

as mesmas características observadas no histórico.

Em suma, esta tese está organizada da seguinte maneira. O Capítulo 2 apresenta todos os conceitos necessários para o desenvolvimento deste trabalho. O Capítulo 3 introduz as cópulas vine que serão amplamente exploradas ao longo deste texto. No Capítulo 4, propomos um modelo de cópulas para a simulação multivariada de cenários de vazões para diversas usinas. Por fim, no Capítulo 5 propomos uma metodologia para simulação univariada de série hidrológicas.

2

Algumas ferramentas matemáticas e cópulas

Este capítulo introduz as cópulas e alguns conceitos pertinentes. Adicionalmente, também são incluídas algumas ferramentas matemáticas e modelos utilizados ao longo dessa tese.

2.1

Cópula

2.1.1

Definição

Iniciaremos esta exposição definindo a transformação integral da probabilidade (*probability integral transform*). Se a variável aleatória X_1 tem função de distribuição acumulada contínua (*f.d.a*) F_1 , então

$$U = F_1(X_1) \sim Unif(0, 1). \quad (2-1)$$

Por outro lado, denotando F_1^{-1} como a função inversa (função quantil) da *f.d.a* F_1 , temos que

$$X_1 = F_1^{-1}(U) \sim F_1. \quad (2-2)$$

Esta transformação possui papel central na teoria de cópulas, uma vez que cópulas são distribuições multivariadas com marginais uniformes.

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 $[0, 1]^d$, $C : [0, 1]^d \rightarrow [0, 1]$, com marginais uniformemente distribuídas.

O teorema de Sklar [12] desempenha papel fundamental na teoria de cópulas. Seja $\mathbf{X} = (X_1, \dots, X_d)$ 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 $(x_1, \dots, x_d) \in \mathfrak{R}^d$,

$$F(\mathbf{x}) = F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)), \quad \mathbf{x} \in \mathfrak{R}^d. \quad (2-3)$$

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.

Por outro lado, é possível representar a cópula C em termos da distribuição conjunta F e das distribuições marginais inversas,

$$C(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)). \quad (2-4)$$

Em termos de densidades, a função $f(x_1, \dots, x_d)$ correspondente a F pode ser escrita em termos de densidade cópula ao se calcular a derivada parcial de 2-3. Dessa forma, temos

$$f(\mathbf{x}) = c_{1,\dots,d}(F_1(x_1), \dots, F_d(x_d)) \cdot \prod_{i=1}^d f_i(x_i), \quad (2-5)$$

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

$$c(u_1, \dots, u_d) = \frac{\partial^d C(u_1, \dots, u_d)}{\partial u_1 \dots \partial u_d}. \quad (2-6)$$

Por fim, a densidade $c_{1,\dots,d}(\cdot)$ pode ser expressa em termos da função de densidade $f(x_1, \dots, x_d)$ e das respectivas marginais como

$$c_{1,\dots,d}(u_1, \dots, u_d) = \frac{f(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))}{f_1(F_1^{-1}(u_1)) \cdot \dots \cdot f_d(F_d^{-1}(u_d))}. \quad (2-7)$$

Segue abaixo três cópulas famosas por suas propriedades. Estas são as cópulas de Fréchet (dependência perfeita positiva/negativa) e a cópula de independência (cópula produto). Os exemplos são dados para o caso bivariado.

Cópula de independência (produto)

No caso de duas variáveis aleatórias independentes X_1 e X_2 , a distribuição conjunta é igual ao produto das distribuições marginais, ou seja, $F(x_1, x_2) = F_1(x_1)F_2(x_2) = C(F_1(x_1)F_2(x_2))$. Com isso, podemos dizer que a cópula produto, usualmente representada por C_{\perp} , é

$$C_{\perp}(u_1, u_2) = u_1 \cdot u_2. \quad (2-8)$$

Cópulas de Fréchet - C_U e C_L

As cópulas de Fréchet são dadas por

$$C_L(u_1, u_2) = \max(u_1 + u_2 - 1, 0), \quad (2-9)$$

e

$$C_U(u_1, u_2) = \min(u_1, u_2). \quad (2-10)$$

Estas cópulas descrevem dependência perfeita negativa e positiva, respectivamente. Para qualquer cópula C , temos que

$$C_L \leq C \leq C_U. \quad (2-11)$$

Por este motivo, as cópulas C_U e C_L também são conhecidas como limites de Fréchet (*Fréchet bounds*). [28]. Mais detalhes sobre limites de Fréchet podem ser encontradas nas referências [29, 30].

2.1.2

Construção pair-cópula

Segundo Aas et al. [13], construir cópulas em alta dimensão (número elevado de variáveis) é uma tarefa difícil. Embora as cópulas permitam uma representação flexível das diferentes estruturas de dependência, seu desenvolvimento ocorreu principalmente no âmbito das cópulas bidimensionais. O número de cópulas bivariadas é elevado, entretanto, o número de cópulas d -dimensionais é limitado. Além disso, as cópulas d -dimensionais existentes impõem algum tipo de restrição tornando-as menos atrativas. Por exemplo, algumas cópulas Arquimedianas multivariadas possuem apenas um ou dois parâmetros responsáveis por governar a estrutura de dependência entre todos os pares de variáveis (veja [31]). A cópula t d -dimensional possui um único parâmetro (graus de liberdade - ν) que determina a dependência de cauda para todos os pares de variáveis.

A construção pair-cópula (*pair-copula construction* - PCC) é um método para a construção de cópulas em altas dimensões que supera as desvantagens supracitadas pois utiliza toda a flexibilidade das cópulas bivariadas. De modo geral, consiste em reescrever uma densidade multivariada em um conjunto de cópulas bivariadas denominadas pair-cópulas. Esta abordagem foi inicialmente proposta por Joe [32] e posteriormente desenvolvida em Bedford & Cooke [14, 15]. Entretanto, somente quando aspectos inferenciais foram abordados por Aas et al. [13] é que pesquisadores deram a devida atenção a esta poderosa ferramenta. Esta seção segue o trabalho de Aas et al. [13].

Para introduzirmos formalmente a PCC, considere d variáveis aleatórias $\mathbf{X} = (X_1, \dots, X_d)$ com a respectiva função de densidade conjunta $f(x_1, \dots, x_d)$. Esta última, pode ser decomposta como

$$f(x_1, \dots, x_d) = f_1(x_1)f_{2|1}(x_2|x_1)f_{3|1,2}(x_3|x_1, x_2) \dots f_{d|1, \dots, d-1}(x_d|x_1, \dots, x_{d-1}). \quad (2-12)$$

Mantida essa estrutura, outras possíveis decomposições podem ser obtidas simplesmente ao se renomear as variáveis.

Inicialmente iremos considerar, sem perda de generalidade, o caso com 4 variáveis e, posteriormente apresentaremos o caso geral d -dimensional. Desta forma, a equação (2-12) passa a ser escrita como

$$f(x_1, x_2, x_3, x_4) = f(x_1)f_{2|1}(x_2|x_1)f_{3|1,2}(x_3|x_1, x_2)f_{4|1,2,3}(x_4|x_1, x_2, x_3). \quad (2-13)$$

A ideia principal da construção pair-cópula é expressar cada distribuição condicional presente na equação (2-13) através de cópulas bivariadas (condicionais e incondicionais) e funções de densidade univariadas (distribuições marginais). Assim sendo, $f_{2|1}(x_2|x_1)$ pode ser reescrita como

$$\begin{aligned} f_{2|1}(x_2|x_1) &= \frac{f_{12}(x_1, x_2)}{f_1(x_1)} \\ &= \frac{c_{12}(F_1(x_1), F_2(x_2))f_1(x_1)f_2(x_2)}{f_1(x_1)} \\ &= c_{12}(F_1(x_1), F_2(x_2))f_2(x_2) \end{aligned} \quad (2-14)$$

onde $f_{12}(x_1, x_2) = c_{12}(F_1(x_1), F_2(x_2))f_1(x_1)f_2(x_2)$ surge da definição estabelecida na equação (2-5).

O termo $f_{3|1,2}(x_3|x_1, x_2)$ pode ser decomposto em

$$\begin{aligned} f_{3|1,2}(x_3|x_1, x_2) &= \frac{f_{1,2,3}(x_1, x_2, x_3)}{f_{1,2}(x_1, x_2)} \\ &= \frac{f_{3,2|1}(x_3, x_2|x_1)}{f_{1,2}(x_1, x_2)} \frac{f_1(x_1)}{f_1(x_1)} \\ &= \frac{f_{3,2|1}(x_3, x_2|x_1)}{f_{2|1}(x_2|x_1)} \frac{1}{f_1(x_1)} \\ &= \frac{c_{3,2|1}(F_3(x_3|x_1), F_{2|1}(x_2|x_1))f_{3|1}(x_3|x_1)f_{2|1}(x_2|x_1)}{f_{2|1}(x_2|x_1)} \frac{1}{f_1(x_1)} \\ &= c_{3,2|1}(F_3(x_3|x_1), F_{2|1}(x_2|x_1))f_{3|1}(x_3|x_1) \end{aligned} \quad (2-15)$$

De maneira similar à 2-14, temos que

$$f_{3|1}(x_3|x_1) = c_{31}(F_3(x_3), F_1(x_1))f_3(x_3). \quad (2-16)$$

Inserindo 2-16 em 2-15 temos então que

$$f_{3|1,2}(x_3|x_1, x_2) = c_{3,2|1}(F_3(x_3|x_1), F_{2|1}(x_2|x_1))c_{31}(F_3(x_3), F_1(x_1))f_3(x_3). \quad (2-17)$$

Neste momento, vale destacar que existem outras possíveis decomposições válidas para o termo $f_{3|1,2}(x_3|x_1, x_2)$. Embora esta conclusão seja óbvia, ela nos indica que a construção de uma função de densidade multivariada através de pair-cópulas não é única. De fato, o número de possíveis combinações aumenta conforme o aumento do número de variáveis. Por exemplo, quando $d = 3$, existem 3 possíveis decomposições. Quando $d = 4$, existem 24 decomposições. Este número cresce rapidamente para 23040 possíveis combinações quando o número de variáveis é de apenas 6, [33].

De volta a equação (2-13), ainda precisamos decompor o termo $f_{4|1,2,3}(x_4|x_1, x_2, x_3)$ para obtermos uma construção pair-cópula completa. Assim sendo, temos que

$$\begin{aligned}
 f_{4|1,2,3}(x_4|x_1, x_2, x_3) &= \frac{f_{1,2,3,4}(x_1, x_2, x_3, x_4)}{f_{1,2,3}(x_1, x_2, x_3)} \\
 &= \frac{f_{3,4|1,2}(x_3, x_4|x_1, x_2)f_{1,2}(x_1, x_2)}{f_{1,2,3}(x_1, x_2, x_3)} \\
 &= \frac{f_{3,4|1,2}(x_3, x_4|x_1, x_2)f_{1,2}(x_1, x_2)}{f_{3|1,2}(x_3|x_1, x_2)f_{1,2}(x_1, x_2)} \\
 &= \frac{f_{3,4|1,2}(x_3, x_4|x_1, x_2)}{f_{3|1,2}(x_3|x_1, x_2)} \\
 &= c_{34|1,2}(F_{3|1,2}(x_3|x_1, x_2), F_{4|1,2}(x_4|x_1, x_2))f_{4|12}(x_4|x_1, x_2).
 \end{aligned} \tag{2-18}$$

O termo $f_{4|12}(x_4|x_1, x_2)$ pode ser manipulado como nas equações (2-15), (2-16) e (2-17). Desta forma, obtemos:

$$\begin{aligned}
 f_{4|1,2,3}(x_4|x_1, x_2, x_3) &= c_{34|1,2}(F_{3|1,2}(x_3|x_1, x_2), F_{4|1,2}(x_4|x_1, x_2)) \\
 &\quad \cdot c_{24|1}(F_{2|1}(x_2|x_1), F_{4|1}(x_4|x_1)) \\
 &\quad \cdot c_{14}(F_1(x_1), F_4(x_4))f_4(x_4)
 \end{aligned} \tag{2-19}$$

Por fim, reunindo as equações (2-14), (2-17) e (2-19), a densidade $f(x_1, x_2, x_3, x_4)$ é finalmente reescrita como

$$\begin{aligned}
 f(x_1, x_2, x_3, x_4) &= f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot f_4(x_4) \\
 &\quad \cdot c_{12}(F_1(x_1), F_2(x_2)) \cdot c_{13}(F_1(x_1), F_3(x_3)) \cdot c_{14}(F_1(x_1), F_4(x_4)) \\
 &\quad \cdot c_{32|1}(F_3(x_3|x_1), F_{2|1}(x_2|x_1)) \cdot c_{24|1}(F_{2|1}(x_2|x_1), F_{4|1}(x_4|x_1)) \\
 &\quad \cdot c_{34|1,2}(F_{3|1,2}(x_3|x_1, x_2), F_{4|1,2}(x_4|x_1, x_2)).
 \end{aligned} \tag{2-20}$$

O exemplo acima demonstra que cada termo da equação (2-13) pode ser reescrito através de um produto de cópulas bivariadas (incondicionais

e condicionais) e densidades univariadas. Vale ressaltar que esta construção assume que a densidade cópula depende da variável condicionante apenas através de seus argumentos, e.g., no caso da cópula $c_{32|1}$, assumimos que $c_{32|1}(F_3(x_3|x_1), F_{2|1}(x_2|x_1)|x_1) = c_{32|1}(F_3(x_3|x_1), F_{2|1}(x_2|x_1))$. Esta hipótese é conhecida na literatura como *simplifying assumption*. Alguns poucos trabalhos investigaram esta hipótese, dentre eles citamos as referências [34–36].

A construção apresentada sugere um procedimento iterativo para a construção pair-cópula. De fato, cada termo em 2-12 pode ser obtido através da seguinte relação (veja [13]):

$$f(x|\mathbf{v}) = c_{xv_j|\mathbf{v}_{-j}}(F(x|\mathbf{v}_{-j}), F(v_j|\mathbf{v}_{-j})) \cdot f(x|\mathbf{v}_{-j}), \quad (2-21)$$

para um vetor d -dimensional \mathbf{v} . Neste caso, v_j é o j -ésimo componente de \mathbf{v} enquanto que \mathbf{v}_{-j} é um vetor igual a \mathbf{v} excluído o j -ésimo componente.

Com relação as funções de distribuições condicionais, Joe [32] demonstrou que

$$F(x|\mathbf{v}) = \frac{\partial C_{x,v_j|\mathbf{v}_{-j}}(F(x|\mathbf{v}_{-j}), F(v_j|\mathbf{v}_{-j}))}{\partial F(v_j|\mathbf{v}_{-j})} \quad (2-22)$$

onde $C_{ij|k}$ é a função de distribuição de uma cópula bivariada. Quando v é escalar,

$$F(x|v) = \frac{\partial C_{x,v}(F_x(x), F_v(v))}{\partial F_v(v)}. \quad (2-23)$$

Quando x e v são uniformes, a equação (2-23) é conhecida na literatura como *h-function*. Desta maneira, a seguinte notação é usualmente empregada:

$$h(x; v, \Theta) = F(x|v) = \frac{\partial C_{x,v}(x, v, \Theta)}{\partial v}. \quad (2-24)$$

Nesta função, o segundo argumento representa a variável condicionante enquanto que Θ representa o conjunto de parâmetros da cópula bivariada $C_{x,v}$.

2.2

Medidas de dependência

A estrutura de dependência que governa uma função de distribuição multivariada é representada por uma cópula. Desta forma medidas de associação baseadas em cópulas são de extrema importância.

Aqui serão introduzidas algumas medidas de dependência. Discutiremos brevemente a correlação de Pearson e duas outras medidas recorrentes na literatura de cópulas; o τ de Kendall's e a dependência de cauda. As principais referências para esta seção são os trabalhos de Kurowicka & Cooke [28], McNeil et al. [29] e Czado [37].

2.2.1

Correlação linear

A correlação de Pearson, ou simplesmente, coeficiente de correlação ρ entre duas variáveis aleatórias X_1 e X_2 , com valores esperados $E(X_1)$ e $E(X_2)$ e variâncias finitas $\sigma_{X_1}^2$ e $\sigma_{X_2}^2$ é dado por

$$\rho(X_1, X_2) = \frac{E(X_1 X_2) - E(X_1)E(X_2)}{\sigma_{X_1} \sigma_{X_2}} = \frac{Cov(X_1, X_2)}{\sigma_{X_1} \sigma_{X_2}}. \quad (2-25)$$

Dado N pares de observações (x_{1i}, x_{2i}) do vetor aleatório (X_1, X_2) , a correlação amostral $\hat{\rho}$ pode ser calculada como

$$\hat{\rho}(X_1, X_2) = \frac{\sum_{i=1}^N (x_{1i} - \bar{X}_1)(x_{2i} - \bar{X}_2)}{\sqrt{\sum_{i=1}^N (x_{1i} - \bar{X}_1)^2} \sqrt{\sum_{i=1}^N (x_{2i} - \bar{X}_2)^2}} \quad (2-26)$$

onde $\bar{X}_1 = \frac{1}{N} \sum_{i=1}^N x_{1i}$ e $\bar{X}_2 = \frac{1}{N} \sum_{i=1}^N x_{2i}$. [28].

A correlação linear é uma medida de dependência bastante popular devido a sua fácil estimação e interpretação. Ela resume num único valor toda a informação sobre o grau de associação entre duas variáveis. É importante notar que esta medida apenas faz sentido no contexto da distribuição normal multivariada, ou de forma mais genérica, distribuições elípticas (veja [29]), pois descreve (parametriza) toda a estrutura de dependência entre as variáveis. Por outro lado em contextos não Gaussianos, esta perde completamente seu significado, uma vez que não parametriza distribuições não elípticas. O simples conhecimento das distribuições marginais e das correlações de um vetor aleatório não é suficiente para determinar a respectiva distribuição conjunta, [29]. McNeil et al. [29] e Kurowicka & Cooke [28] citam algumas desvantagens da correlação linear, dentre as quais destacamos:

- (i) A correlação não é invariante a transformações não lineares estritamente crescentes. No caso de duas variáveis aleatórias, temos em geral que $\rho(T(X_1), T(X_2)) \neq \rho(X_1, X_2)$.
- (ii) A correlação só existe quando as variâncias de X_1 e X_2 são finitas. Como exemplo, a distribuição de Cauchy não possui segundo momento definido.
- (iii) A correlação sofre influência das distribuições marginais de X_1 e X_2 .

Uma discussão mais aprofundada sobre os riscos do uso indevido da correlação linear na construção de modelos multivariados pode ser encontrada em [29].

Isto posto, medidas de dependência com propriedades matemáticas mais atrativas são necessárias quando estamos em um contexto mais amplo de dependência. Além disso, muitas vezes desejamos construir modelos a partir das

distribuições marginais e como já mencionado, o conhecimento das marginais e da correlação não é suficiente para determinar a distribuição conjunta. Assim sendo, medidas de dependência baseadas em cópulas são de extrema importância.

2.2.2 Tau de Kendall

Correlações de *rank* (*rank correlations*) são medidas de dependência com propriedades mais atrativas do que a correlação linear. Exemplos de *rank correlations* são o τ de Kendall e o ρ de Spearman. Do ponto de vista teórico, essas medidas de dependência são independentes das distribuições marginais, invariantes a transformações não lineares estritamente crescentes e sempre estarão definidas. Elas assumem valores no intervalo $[-1, 1]$ e, além disso, podem ser expressas em termos da cópula.

Do ponto de vista empírico, estimadores de *rank correlations* podem ser calculados através dos *ranks* de um conjunto de dados. Precisa-se apenas conhecer a ordem das observações na amostra, sendo seus valores irrelevantes.

Estes estimadores utilizam o conceito de pares concordantes e pares discordantes. A Figura 2.1 exemplifica estes dois conceitos. Nesta, apresentamos observações (x_{1i}, x_{2i}) , $i = 1, \dots, 4$, oriundas do par de variáveis aleatórias (X_1, X_2) . Os pontos $(0.75, 0.75)$ e $(0.45, 0.55)$ são ditos pares concordantes pois $x_{1i} > x_{1j}$ e $x_{2i} > x_{2j}$. Por outro lado, os pontos $(0.1, 0.45)$ e $(0.25, 0.1)$ são denominados pares discordantes pois $x_{1i} < x_{1j}$ e $x_{2i} > x_{2j}$. De maneira intuitiva podemos afirmar que dois pares são concordantes se a ordem das duas observações da variável X_1 é a mesma que a ordem de duas observações da variável X_2 . Caso a ordem seja inversa, temos um par discordante.

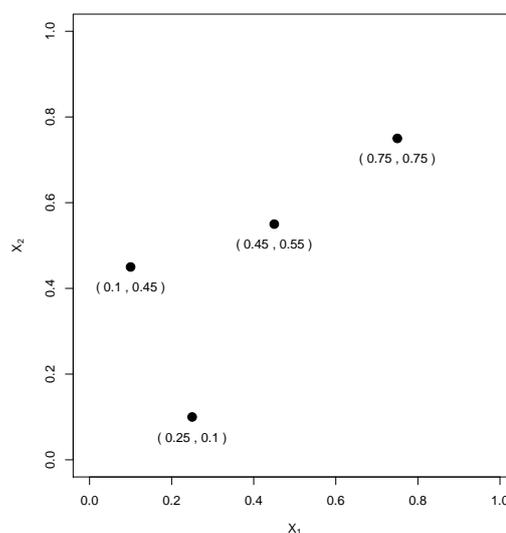


Figura 2.1: Exemplo de pares concordantes e discordantes.

O τ de Kendall é definido como a probabilidade de pares concordantes menos a probabilidade de pares discordantes. Formalmente, seja (X_{11}, X_{21}) e (X_{12}, X_{22}) dois vetores aleatórios independentes oriundos da mesma distribuição conjunta $F_{12}(X_1, X_2)$. Dois pares são ditos concordantes se $(X_{11} - X_{12})(X_{21} - X_{22}) > 0$ e discordantes se $(X_{11} - X_{12})(X_{21} - X_{22}) < 0$. O τ de Kendall é dado por

$$\tau = \mathbf{P}[(X_{11} - X_{12})(X_{21} - X_{22}) > 0] - \mathbf{P}[(X_{11} - X_{12})(X_{21} - X_{22}) < 0]. \quad (2-27)$$

Para estimar o τ de Kendall a partir de um conjunto com N pares de observações $(x_{1i}, x_{2i}), i = 1, \dots, n$, devemos considerar todos os $N(N - 1)/2$ pares diferentes (x_{1i}, x_{2i}) e $(x_{1j}, x_{2j}), i < j$ e classificá-los como:

- pares concordantes: se $x_{1i} < x_{1j}$ e $x_{2i} < x_{2j}$ ou $x_{1i} > x_{1j}$ e $x_{2i} > x_{2j}$,
- pares discordantes: se $x_{1i} < x_{1j}$ e $x_{2i} > x_{2j}$ ou $x_{1i} > x_{1j}$ e $x_{2i} < x_{2j}$,
- pares extra- x_1 : se $x_{1i} = x_{1j}$,
- pares extra- x_2 : se $x_{2i} = x_{2j}$.

Definindo N_{con} como número de pares concordantes, N_{dis} como número de pares discordantes, N_x como número de pares extra- x e N_y como o número de pares extra- y , o estimador amostral do τ de Kendall de acordo com [28] é dado por:

$$\tau = \frac{N_{con} - N_{dis}}{\sqrt{N_{con} + N_{dis} + N_y} \sqrt{N_{con} + N_{dis} + N_x}}. \quad (2-28)$$

Como dito anteriormente, o τ de Kendall pode ser expresso em função da cópula. A equação (2-27), pode ser reescrita como (veja [29, 37]):

$$\begin{aligned} \tau &= \mathbf{P}[(X_{11} - X_{12})(X_{21} - X_{22}) > 0] - \mathbf{P}[(X_{11} - X_{12})(X_{21} - X_{22}) < 0] \\ &= 2\mathbf{P}[(X_{11} - X_{12})(X_{21} - X_{22}) > 0] - 1 \\ &= 4\mathbf{P}[X_{11} < X_{12}, X_{21} < X_{22}] - 1 \\ &= 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1 \end{aligned} \quad (2-29)$$

A equação (2-29) estabelece uma relação entre o τ de Kendall e o parâmetro da cópula. Para as principais cópulas bivariadas esta relação é uma função conhecida. Assim, pode-se utilizá-la como um estimador para os parâmetros das cópulas. Esta correspondência também possui papel fundamental nos modelos de cópulas vine espaciais que serão abordados no Capítulo 4.

2.2.3

Dependência de cauda

A dependência de cauda (*coefficients of tail dependence*) fornece informações acerca da dependência nos extremos de uma distribuição bivariada. [29].

A dependência de cauda superior mede a probabilidade da variável X_2 exceder determinado quantil q dado que a variável X_1 já excedeu seu q -ésimo quantil. Formalmente, se X_1 e X_2 são variáveis aleatórias com distribuições marginais contínuas dadas por F_1 e F_2 , o índice de cauda superior (λ_u) é

$$\lambda_u := \lim_{q \rightarrow 1^-} P\left(X_2 > F_2^{-1}(q) | X_1 > F_1^{-1}(q)\right) = \lim_{q \rightarrow 1^-} \frac{1 - 2q + C(q, q)}{1 - q}, \quad (2-30)$$

se o limite $\lambda_u \in [0, 1]$ existir. Se $\lambda_u \in (0, 1]$, dizemos que X_1 e X_2 possuem dependência de cauda superior. A ausência de dependência de cauda é caracterizada por $\lambda_u = 0$.

De maneira análoga, a dependência de cauda inferior é dada por

$$\lambda_l := \lambda_l(X_1, X_2) = \lim_{q \rightarrow 0^+} P\left(X_2 \leq F_2^{-1}(q) | X_1 \leq F_1^{-1}(q)\right) = \lim_{q \rightarrow 0^+} \frac{C(q, q)}{q}, \quad (2-31)$$

se o limite $\lambda_u \in [0, 1]$ existir. [31, 37]

2.3

Algumas cópulas bivariadas

O objetivo das próximas duas subseções é ilustrar as principais propriedades das cópulas bivariadas empregadas ao longo desta tese. Como principais referências utilizou-se os textos de Aas et al. [13] e Joe [24]. Uma extensa lista de cópulas e suas propriedades são elencadas em Joe [24]. Além disso, equações para as cópulas condicionais ($w = C_{2|1}(u_2|u_1)$) e as respectivas funções inversas ($u_2 = C_{2|1}^{-1}(w|u_1)$) estão presentes neste livro. Aas et al. [13] listam algumas cópulas bivariadas e derivam as respectivas cópulas condicionais.

Cópula Gaussiana

A cópula Gaussiana ou cópula normal pode ser definida como

$$C(u_1, u_2; \rho) = \Phi_2\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho\right), \quad 0 < u_1, u_2 < 1 \\ \rho \in [-1, 1] \quad (2-32)$$

onde $\Phi_2(\cdot, \cdot, \rho)$ é a função de distribuição acumulada de uma normal padrão bivariada e Φ^{-1} é a função de distribuição acumulada inversa de uma normal padrão. A densidade da cópula Gaussiana é dada por

$$c(u_1, u_2; \rho) = \frac{1}{\sqrt{1 - \rho^2}} \exp\left(-\frac{\rho^2(x_1^2 - x_2^2) - 2\rho x_1 x_2}{2(1 - \rho^2)}\right), \quad (2-33)$$

onde $x_1 = \Phi^{-1}(u_1)$, e $x_2 = \Phi^{-1}(u_2)$.

A distribuição condicional é:

$$C_{2|1}(u_2|u_1; \rho) = \Phi\left(\frac{\Phi^{-1}(u_2) - \rho\Phi^{-1}(u_1)}{\sqrt{1 - \rho^2}}\right) \quad (2-34)$$

A relação entre o parâmetro da cópula e o τ de Kendall, obtida ao se resolver a equação (2-29), é

$$\tau = \frac{2}{\pi} \arcsin(\rho). \quad (2-35)$$

A Figura 2.2 apresenta as curvas de níveis da cópula Gaussiana com diferentes graus de associação. Quando $\tau = 1$, a cópula Gaussiana apresenta dependência positiva perfeita. Se $\tau = -1$ a mesma apresenta dependência negativa perfeita. A cópula Gaussiana não apresenta dependência de cauda, ou seja, $\lambda_l = \lambda_u = 0$.

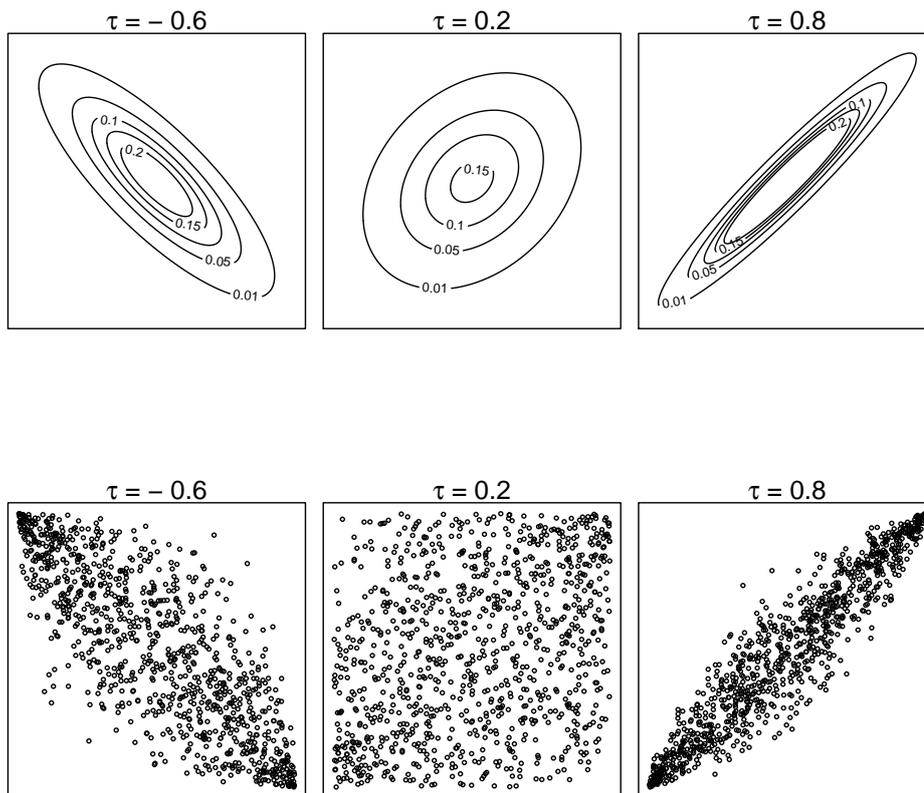


Figura 2.2: Gráfico de pontos e contorno: cópula Gaussiana.

Cópula t

A cópula t de Student com parâmetros $\rho \in (-1, 1)$ e $\nu > 2$ pode ser definida como:

$$C(u_1, u_2; \rho, \nu) = T_{\rho, \nu}(T_\nu^{-1}(u_1), T_\nu^{-1}(u_2)), \quad 0 \leq u_1, u_2 \leq 1, \quad (2-36)$$

onde $T_{\rho, \nu}$ é função de distribuição acumulada t de Student bivariada com $\nu > 2$ graus de liberdade e correlação igual a ρ . T_ν é a função de distribuição acumulada univariada t de Student e T_ν^{-1} é a respectiva função inversa.

Sua densidade é dada por

$$c(u_1, u_2; \rho, \nu) = \frac{1}{\sqrt{1-\rho^2}} \frac{\Gamma(\frac{\nu+2}{2})\Gamma(\frac{\nu}{2})}{[\Gamma(\frac{\nu+1}{2})]^2} \frac{\left(1 + \frac{x_1^2 + x_2^2 - 2\rho x_1 x_2}{(1-\rho^2)\nu}\right)^{-(\nu+2)/2}}{\left(1 + \frac{x_1^2}{\nu}\right)^{-(\nu+1)/2} \left(1 + \frac{x_2^2}{\nu}\right)^{-(\nu+1)/2}} \quad (2-37)$$

onde $x_1 = T_\nu^{-1}(u_1)$ e $x_2 = T_\nu^{-1}(u_2)$.

A cópula condicional é dada por:

$$C_{2|1}(u_2|u_1; \rho, \nu) = T_{\nu+1} \left(\frac{T_\nu^{-1}(u_2) - \rho T_\nu^{-1}(u_1)}{\sqrt{(1-\rho^2)(\nu + [T_\nu^{-1}(u_1)]^2)/(\nu+1)}} \right) \quad (2-38)$$

O τ de Kendall expresso em termos do parâmetro ρ é dado por

$$\tau = \frac{2}{\pi} \arcsin(\rho). \quad (2-39)$$

A dependência de cauda superior e inferior é dada por

$$\lambda_l = \lambda_u = 2T_{\nu+1} \left(-\sqrt{\nu+1} \frac{\sqrt{1-\rho}}{\sqrt{1+\rho}} \right). \quad (2-40)$$

Assim como a cópula Gaussiana, a dependência negativa perfeita e a dependência positiva perfeita são obtidas quando $\tau = -1$ e $\tau = 1$. A cópula Gaussiana pode ser entendida como um caso extremo da cópula t quando $\nu \rightarrow \infty$. Na Figura 2.3 é possível visualizar as curvas de nível para cópulas t com diferentes parâmetros.

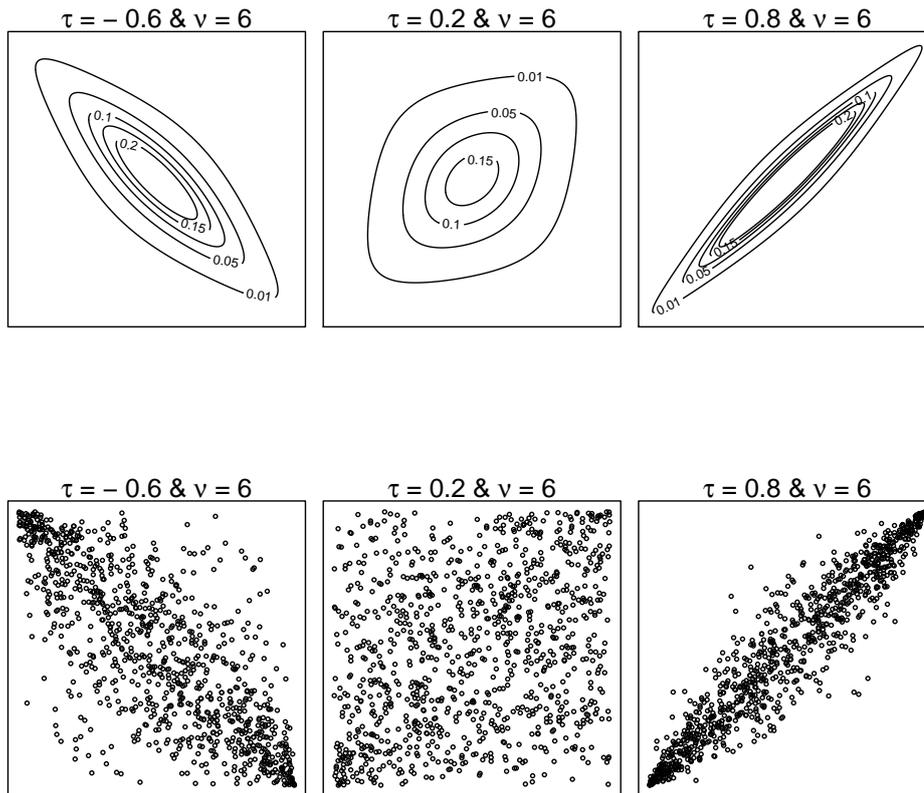


Figura 2.3: Gráfico de pontos e contorno: cópula t.

Cópula Clayton

A cópula Clayton bivariada com parâmetro $\delta \in [0, \infty)$ é dada por

$$C(u_1, u_2; \delta) = (u_1^{-\delta} + u_2^{-\delta} - 1)^{-\frac{1}{\delta}}, \quad 0 \leq u_1, u_2 \leq 1, \quad (2-41)$$

enquanto sua densidade é definida como

$$c(u_2, u_1; \delta) = (1 + \delta)(u_1 u_2)^{-1-\delta} (u_1^{-\delta} + u_2^{-\delta} - 1)^{-\frac{1}{\delta}-2}. \quad (2-42)$$

A cópula condicional é:

$$C_{2|1}(u_2|u_1; \delta) = [1 + u_1(u_2^{-\delta} - 1)]^{-1-1/\delta} \quad (2-43)$$

O τ de Kendall expresso em termos do parâmetro da cópula é:

$$\tau = \frac{\delta}{\delta + 2}. \quad (2-44)$$

Esta cópula possui somente dependência de cauda inferior, $\lambda_l = 2^{-1/\delta}$. Diferentemente das cópulas Gaussiana e t, apenas associações positivas podem

ser modeladas. Quando $\delta \rightarrow 0^+$ a cópula Clayton se aproxima da cópula de independência. Por outro lado, quando $\delta \rightarrow \infty$ esta cópula apresenta dependência positiva perfeita. Na Figura 2.4 é possível observar o formato da cópula Clayton com diferentes graus de associação.

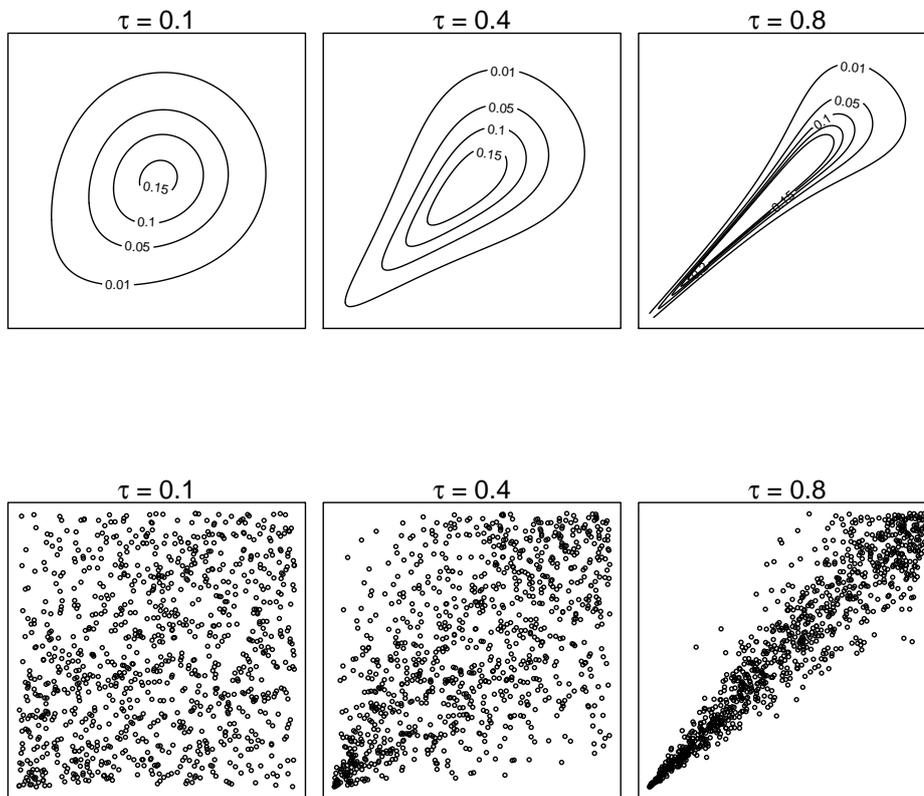


Figura 2.4: Gráfico de pontos e contorno: cópula Clayton.

Cópula Gumbel

A cópula Gumbel com o parâmetro $\delta \in [1, \infty)$ é igual a

$$C(u_1, u_2; \delta) = \exp\{-[(-\log u_1)^\delta + (-\log u_2)^\delta]^{1/\delta}\}, \quad 0 \leq u_1, u_2 \leq 1 \quad (2-45)$$

A função densidade cópula é expressa por

$$\begin{aligned} c(u_1, u_2; \delta) &= C(u_1, u_2; \delta)(u_1 u_2)^{-1} ((-\log u_1)^\delta + (-\log u_2)^\delta)^{-2+2/\delta} \\ &\quad \times (\log u_1 \log u_2)^{\delta-1} \\ &\quad \times \{1 + (\delta - 1)((-\log u_1)^\delta + (-\log u_2)^\delta)^{-1/\delta}\} \end{aligned} \quad (2-46)$$

enquanto que a distribuição condicional é dada por

$$C_{2|1}(u_2|u_1; \delta) = u_1^{-1} \exp\{-[-\log u_1^\delta - \log u_2^\delta]^{1/\delta}\} [1 + (-\log u_2 / -\log u_1)^\delta]^{1/\delta-1}. \quad (2-47)$$

A função que estabelece a relação entre o parâmetro da cópula e τ de Kendall é

$$\tau = \frac{\delta - 1}{\delta}. \quad (2-48)$$

A cópula Gumbel possui apenas dependência de cauda superior ($\lambda_u = 2 - 2^{1/\delta}$). Além do mais, apenas associações positivas entre variáveis são passíveis de serem modeladas. A independência é obtida quando $\delta = 1$. Por outro lado, $\delta \rightarrow \infty$ implica em dependência perfeita. A Figura 2.5 apresenta a cópula Gumbel para diferentes valores do τ de Kendall.

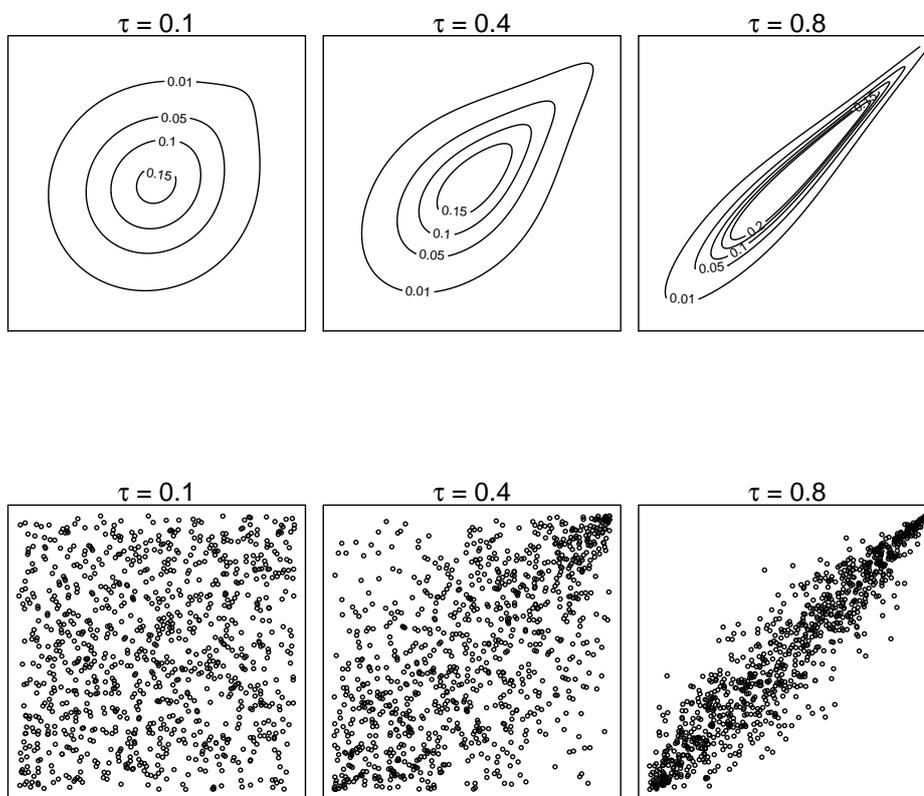


Figura 2.5: Gráfico de pontos e contorno: cópula Gumbel.

Cópula Frank

A cópula Frank é definida como

$$C(u_1, u_2; \delta) = -\delta^{-1} \log \left(\frac{1 - e^{-1} - (1 - e^{-\delta u_1})(1 - e^{-\delta u_2})}{1 - e^{-\delta}} \right), \quad 0 \leq u_1, u_2 \leq 1 \quad (2-49)$$

onde $-\infty < \delta < \infty$.

Sua densidade é

$$c(u_1, u_2; \delta) = \frac{\delta(1 - e^{-\delta})e^{-\delta(u_1+u_2)}}{[1 - e^{-\delta} - (1 - e^{-\delta u_1})1 - e^{-\delta u_2}]^2}. \quad (2-50)$$

A distribuição condicional da cópula Frank é dada por

$$C_{2|1}(u_1|u_2; \delta) = e^{-\delta u_1}[(1 - e^{-\delta})(1 - e^{-\delta u_2})^{-1} - (1 - e^{-\delta u_1})]^{-1}. \quad (2-51)$$

O τ de Kendall pode ser expresso em termos do parâmetro da cópula Frank como

$$\tau = 1 + 4\delta^{-1}[D(\delta) - 1] \quad (2-52)$$

onde $D(x) = x^{-1} \int_0^x t(e^t - 1)^{-1} dt$.

A cópula Frank pode representar tanto associações negativas quanto positivas. A dependência negativa perfeita ocorre quando $\delta \rightarrow -\infty$. Por outro lado, $\delta \rightarrow \infty$ implica em dependência positiva perfeita. A independência é obtida quando $\delta \rightarrow 0^+$. A Figura 2.6 ilustra a cópula Frank para diferentes valores do τ de Kendall.

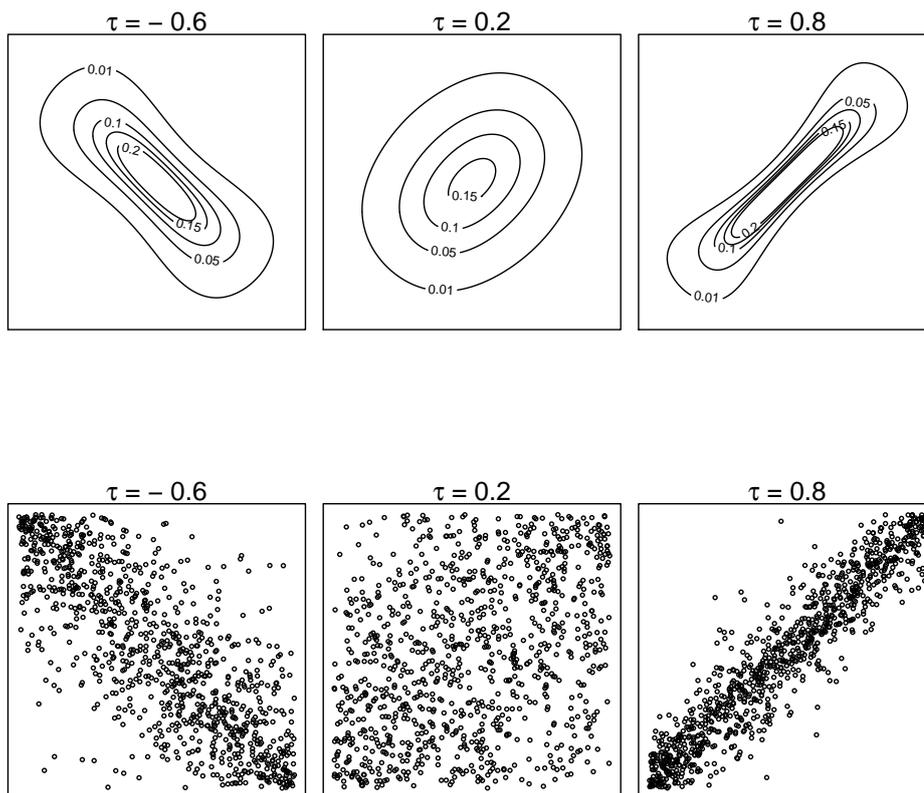


Figura 2.6: Gráfico de pontos e contorno: cópula Frank.

Cópulas e suas versões rotacionadas

Anteriormente foi apresentado algumas famílias que acomodam apenas associação positiva entre variáveis. Num primeiro momento, pode parecer que isto as tornem menos atrativas. Porém, é possível ampliar a gama de dependência que estas podem abranger simplesmente rotacionando-as.

Para estender estas famílias e permitir que descrevam formas mais gerais de dependência devemos rotacionar a densidade cópula no sentido anti-horário em 90° , 180° ou 270° . Matematicamente, estas rotações podem ser definidas como:

- 90° : $c_{90^\circ}(u_1, u_2) := c(1 - u_1, u_2)$
- 180° : $c_{180^\circ}(u_1, u_2) := c(1 - u_1, 1 - u_2)$
- 270° : $c_{270^\circ}(u_1, u_2) := c(u_1, 1 - u_2)$

Na Figura 2.7 é possível observar as curvas de nível das versões rotacionadas das cópulas Clayton e Gumbel em 90° , 180° e 270° . Vale lembrar que estas rotações também alteram a dependência de cauda.

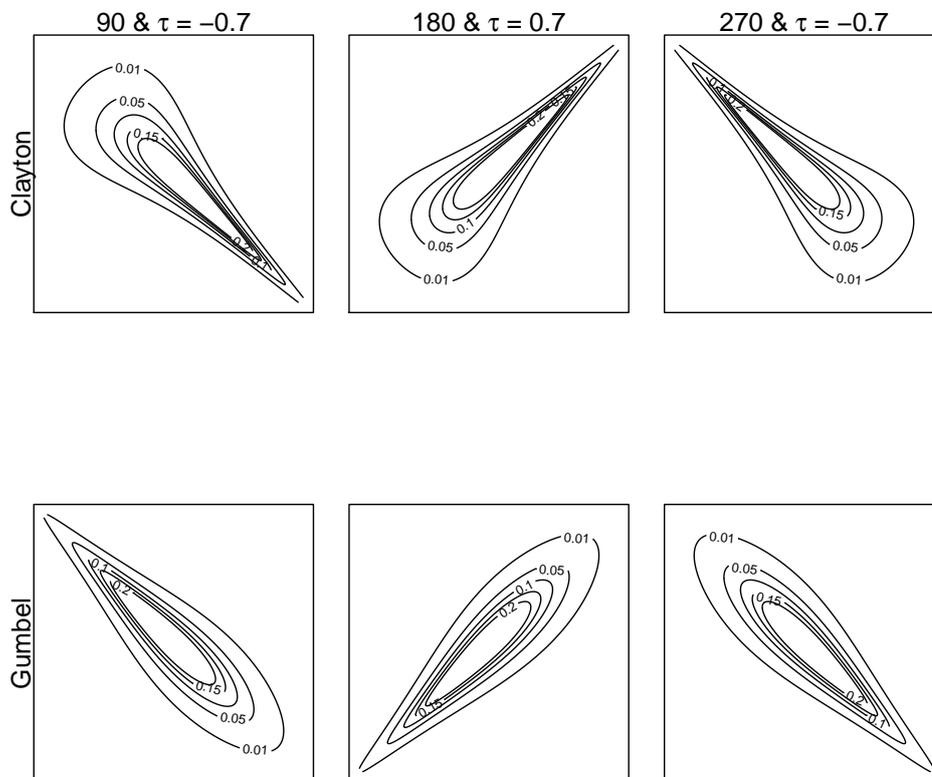


Figura 2.7: Versões rotacionadas das cópulas Clayton e Gumbel.

2.4

Transformada de Fisher

A transformada de Fisher ([38]) serve para transformar variáveis restritas no intervalo $(-1, 1)$ em variáveis definidas no intervalo $(-\infty, \infty)$. Matematicamente é dada por

$$g_z(r) = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right), \quad r \in (-1, 1) \quad (2-53)$$

enquanto que sua inversa é definida como

$$g_z^{-1}(z) = \frac{e^{2z} - 1}{e^{2z} + 1}, \quad z \in \mathbb{R}. \quad (2-54)$$

Estas duas relações desempenham um papel fundamental no modelo de cópulas vine espaciais e serão retomadas no Capítulo 4.

2.5

Modelo de séries temporais - PAR (p)

No Capítulo 4, propomos um sofisticado modelo para simulação simultânea de cenários de vazões para diversas usinas. Tal formulação modela inicialmente a dependência temporal e posteriormente, a relação entre usinas através de cópulas. Neste sentido, um dos componentes do modelo proposto é um modelo de séries temporais que represente adequadamente a estrutura temporal de cada série de vazão. Para tal tarefa, o modelo PAR(**p**) - *periódico autorregressivo* é empregado.

De acordo com Hipel & McLeod [39], algumas séries hidrológicas exibem estruturas de autocorrelação que não dependem somente do *lag* entre as observações, mas também do período onde estas se encontram. Além disso, Salas et al. [5] argumentam que estruturas probabilísticas de tais séries, e.g. média e variância, variam de acordo com o período do ano. Nestes casos, o emprego de modelos periódicos é recomendado, sendo um dos mais proeminentes o modelo PAR(**p**). No âmbito do Setor Elétrico Brasileiro (SEB), este modelo é amplamente utilizado. Dentre alguns trabalhos, mencionamos as referências [4, 40, 41].

Antes de apresentarmos o modelo, definiremos a notação utilizada no modelo PAR(**p**). Considere y_t , $t = 1, \dots, T$, observações consecutivas de uma série temporal mensal com s períodos ($s = 12$). Sem perda de generalidade, vamos considerar que $T/s = n$ é um número inteiro, ou seja, existem n anos completos de observações disponíveis no histórico. No caso de séries mensais, o índice temporal t pode ser reescrito em função dos anos e dos meses. Dessa forma, temos que $t = t(r, m) = (r-1)s + m$, onde $r = 1, \dots, n$ e $m = 1, \dots, 12$. Neste caso específico, r representa os anos e m simboliza os meses. [42].

O modelo PAR(\mathbf{p}) admite que seus parâmetros mudem de acordo com o período (meses, no caso de séries mensais). De maneira geral, para cada mês um modelo autorregressivo é ajustado onde tanto a ordem do modelo quanto os parâmetros variam a cada mês. Mais detalhes podem ser encontrados nas referências [5, 39]. Matematicamente, o modelo PAR(\mathbf{p}) de ordem \mathbf{p} ($\mathbf{p} = [p_1, \dots, p_{12}]$) é dado por

$$\left(\frac{y_t - \mu_m}{\sigma_m}\right) = \sum_{i=1}^{p_m} \phi_i^{(m)} \left(\frac{y_{t-i} - \mu_{m-i}}{\sigma_{m-i}}\right) + \varepsilon_t \quad (2-55)$$

onde m é o período ($m = 1, \dots, s$; $s = 12$, no caso de séries mensais), μ_m é a média mensal, σ_m é o desvio padrão mensal, $\phi_i^{(m)}$, é o i -ésimo parâmetro autorregressivo do período m , p_m é a ordem do período m e ε_t é uma série de ruídos independentes com média zero e variância igual à $\sigma_{\varepsilon, m}^2$.

2.5.1

Identificação das ordens p_m e estimação dos parâmetros autorregressivos

A estimação dos parâmetros autorregressivos para determinado período m pode ser conduzida através da versão periódica do sistema de equações Yule-Walker.

Para isto, defina a autocorrelação periódica de lag k para o período m como

$$\rho_k^{(m)} = E \left[\left(\frac{y_t - \mu_m}{\sigma_m}\right) \left(\frac{y_{t-k} - \mu_{m-k}}{\sigma_{m-k}}\right) \right], \quad k = 1, 2, \dots \quad (2-56)$$

A função de autocorrelação periódica teórica para determinado período m é obtida ao multiplicarmos a equação (2-55) por $\left(\frac{y_{t-k} - \mu_{m-k}}{\sigma_{m-k}}\right)$ e calcularmos o valor esperado, ou seja,

$$E \left[\left(\frac{y_t - \mu_m}{\sigma_m}\right) \left(\frac{y_{t-k} - \mu_{m-k}}{\sigma_{m-k}}\right) \right] = E \left[\sum_{i=1}^{p_m} \phi_i^{(m)} \left(\frac{y_{t-i} - \mu_{m-i}}{\sigma_{m-i}}\right) \left(\frac{y_{t-k} - \mu_{m-k}}{\sigma_{m-k}}\right) \right] + E \left[\varepsilon_t \left(\frac{y_{t-k} - \mu_{m-k}}{\sigma_{m-k}}\right) \right]. \quad (2-57)$$

Para determinado período m e considerando $k = 1, \dots, p_m$, a representação do sistema periódico de equações de Yule-Walker é $\mathbf{R}^{(m)}\Phi^{(m)} = \mathbf{r}^{(m)}$, onde

$$\mathbf{R}^{(m)} = \begin{pmatrix} 1 & \rho_1^{(m-1)} & \rho_2^{(m-1)} & \cdots & \rho_{p_m-1}^{(m-1)} \\ \rho_1^{(m-1)} & 1 & \rho_1^{(m-2)} & \cdots & \rho_{p_m-2}^{(m-2)} \\ \rho_2^{(m-1)} & \rho_1^{(m-2)} & 1 & \cdots & \rho_{p_m-3}^{(m-3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{p_m-1}^{(m-1)} & \rho_{p_m-2}^{(m-2)} & \rho_{p_m-3}^{(m-3)} & \cdots & 1 \end{pmatrix}, \Phi^{(m)} = \begin{pmatrix} \phi_1^{(m)} \\ \phi_2^{(m)} \\ \phi_3^{(m)} \\ \vdots \\ \phi_{p_m}^{(m)} \end{pmatrix},$$

$$\mathbf{r}^{(m)} = \begin{pmatrix} \rho_1^{(m)} \\ \rho_2^{(m)} \\ \rho_3^{(m)} \\ \vdots \\ \rho_{p_m}^{(m)} \end{pmatrix}. \quad (2-58)$$

Se substituirmos na equação (2-58) as autocorrelações teóricas pelas respectivas autocorrelações amostrais, as estimativas dos parâmetros podem ser obtidas ao se resolver

$$\hat{\Phi}^{(m)} = \hat{\mathbf{R}}^{(m)-1} \hat{\mathbf{r}}^m. \quad (2-59)$$

A autocorreção periódica amostral de lag k para o período m é dada por

$$\hat{\rho}_k^{(m)} = \frac{\hat{\gamma}_k^{(m)}}{\hat{\sigma}_m \hat{\sigma}_{m-k}} \quad (2-60)$$

onde

$$\hat{\gamma}_k^{(m)} = \frac{1}{n} \sum_{j=1}^n (y_{s(j-1)+m} - \hat{\mu}_m)(y_{s(j-1)+m-k} - \hat{\mu}_{m-k}). \quad (2-61)$$

A estimativa para a variância do período m é

$$\hat{\gamma}_o^{(m)} = \hat{\sigma}_m^2 = \frac{1}{n} \sum_{j=1}^n (y_{s(j-1)+m} - \hat{\mu}_m)^2, \quad (2-62)$$

enquanto que a média periódica é

$$\hat{\mu}_m = \frac{1}{n} \sum_{j=1}^n y_{s(j-1)+m}. \quad (2-63)$$

A variância dos resíduos $\sigma_{\varepsilon,m}^2$ é estimada pela seguinte equação:

$$\hat{\sigma}_{\varepsilon,m}^2 = 1 - \sum_{i=1}^{p_m} \hat{\phi}_i^{(m)} \hat{\rho}_i^{(m)}. \quad (2-64)$$

Mais detalhes sobre estimação de modelos periódicos podem ser encontradas na referências [5, 40].

Com relação a identificação das ordens ($\mathbf{p} = [p_1, \dots, p_{12}]$), esta pode ser realizada através de critérios de seleção automáticos tais como o BIC (*Bayesian Information Criteria*). O BIC para o modelo PAR(p) é dado por

$$BIC = \sum_{m=1}^s BIC_m, \quad (2-65)$$

onde

$$BIC_m = n \log \hat{\sigma}_{\varepsilon,m}^2 + \log(n)p_m, \quad (2-66)$$

sendo s o número de períodos e n o número de anos. [42].

2.5.2

Simulação de cenários

A existência de modelos de tempo se justifica não só pela análise e previsão de séries temporais como também pela possibilidade de simulação de cenários.

Dado um modelo de séries temporais ajustado, é possível gerar outras séries temporais diferentes uma das outras porém equiprováveis do ponto de vista probabilístico. No Capítulo 4 utilizaremos esta ideia para a simulação de cenários multivariados de vazões.

Novamente, a equação do modelo PAR(p) (eq. 2-55) é

$$\left(\frac{y_t - \mu_m}{\sigma_m} \right) = \sum_{i=1}^{p_m} \phi_i^{(m)} \left(\frac{y_{t-i} - \mu_{m-i}}{\sigma_{m-i}} \right) + \varepsilon_t.$$

O termo ε pode ser reescrito como $\sigma_{\varepsilon,m}\epsilon_t$, onde ϵ_t segue uma distribuição normal com média zero e variância unitária.

Dessa forma a equação (2-55) pode ser reescrita como

$$y_t = \mu_m + \sigma_m \left[\sum_{i=1}^{p_m} \phi_i^{(m)} \left(\frac{y_{t-i} - \mu_{m-i}}{\sigma_{m-i}} \right) \right] + \sigma_m \sigma_{\varepsilon,m} \epsilon_t. \quad (2-67)$$

A equação acima é extremamente útil no contexto de simulação, pois permite que novas realizações da série possam ser gerados. Para tal, basta substituir os parâmetros pelas suas respectivas estimativas, realizar sorteios aleatórios de uma distribuição normal padrão e injetar sequencialmente estes novos resíduos na equação (2-67).

2.6

Conceitos básicos sobre teoria dos grafos

Alguns conceitos básicos acerca da teoria dos grafos são introduzidos neste momento. Estes são fundamentais para o entendimento de cópulas vine. Os conceitos aqui expostos seguem os apresentados em [37, 43, 44].

Seja N um conjunto arbitrário e E é um subconjunto bidimensional composto pelas combinações dos elementos de N ,

$$E \subset \{ \{n_1, n_2\} \mid n_1, n_2 \in N \}. \quad (2-68)$$

Um grafo é um par $G = (N, E)$ onde os elementos de N são denominados vértices enquanto que os elementos do conjunto E são conhecidos como arestas.

A Figura 2.8 ilustra um grafo $G' = (N', E')$ onde

- $N' = \{1, \dots, 6\}$ e
- $E' = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{4, 6\}, \{5, 6\}, \{3, 5\}\}$.

Nesta Figura, os círculos representam os vértices (N) e as linhas que os conectam são as arestas (E). Dois vértices n_i e n_j são conectados se $\{n_i, n_j\} \in E$.

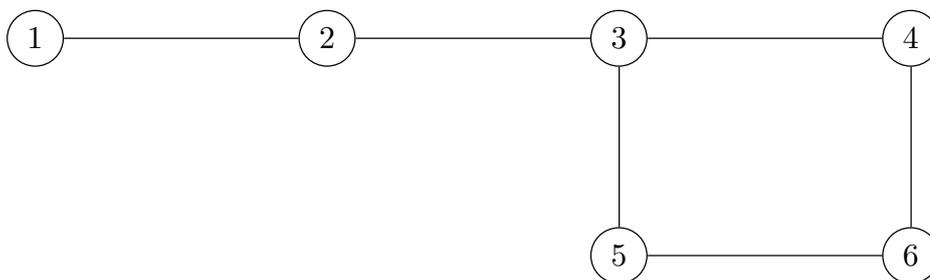


Figura 2.8: Exemplo de grafo com 6 vértices e 6 arestas.

Com o auxílio da Figura 2.8, podemos ilustrar o conceito de caminho (*path*).

De maneira intuitiva, um caminho nada mais é do que o percurso entre dois vértices. Formalmente, este é definido como um grafo $G^* = (N^*, E^*)$ onde $N^* = \{n_0, n_1, \dots, n_k\}$ e $E^* = \{\{n_0, n_1\}, \{n_1, n_2\}, \dots, \{n_{k-1}, n_k\}\}$. O primeiro vértice é denominado vértice inicial enquanto que o último é conhecido como vértice final.

Por exemplo, entre os vértices 1 e 4 existem dois caminhos possíveis. Estes são $1 - 2 - 3 - 4$ e $1 - 2 - 3 - 5 - 6$. Um caso particular surge quando os vértices iniciais e finais são iguais, ou seja, $n_0 = n_k$. Esta situação é conhecida como ciclo. Observe que na Figura 2.8, se iniciarmos a caminhada pelo vértice 3, é possível retornar ao mesmo através do caminho $3 \rightarrow 4 \rightarrow 6 \rightarrow 5 \rightarrow 3$.

Outro conceito importante é o conceito de grafos conexos. Dizemos que dois vértices $a, b \in N, a \neq b$ estão conectados se existe um caminho entre a e b . Um grafo $G = (N, E)$ é dito conexo se a partir de qualquer vértice, todos os outros vértices são alcançáveis. Em outras palavras, dizemos que um grafo é conexo se para dois vértices quaisquer sempre existe um caminho que os liguem.

A Figura 2.8 ilustra esse conceito. Observe que para dois vértices quaisquer, sempre existe um caminho ligando-os. Um contraexemplo, é apresentado na Figura 2.9 onde o vértice 7 não possui nenhuma conexão com os demais.

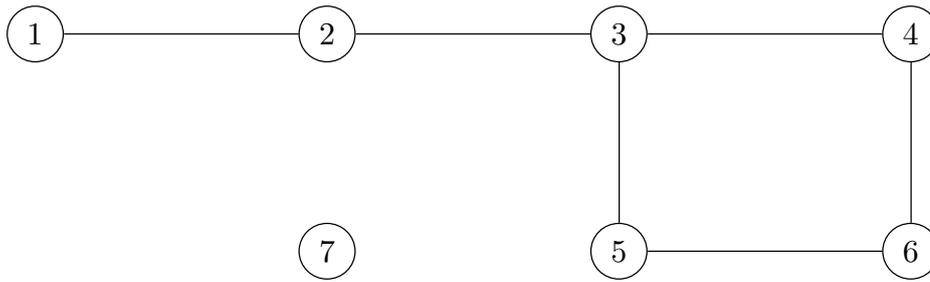


Figura 2.9: Exemplo de grafo com vértice isolado.

Alguns casos particulares de cópulas vines são identificados pelo grau de seus vértices. Este conceito pode ser entendido como o número de arestas que estão conectadas a determinado vértice. Matematicamente, se $n \in N$, então o grau do vértice n é definido com $grau(n) = \#\{e \in E | n \in e\}$. Na Figura 2.9, $grau(7) = 0$, $grau(4) = 2$ e $grau(1) = 1$.

Finalmente, estamos aptos a definir o conceito de árvore. Como ficará claro adiante, toda a representação gráfica de cópulas vine é construída utilizando este conceito.

Uma árvore é definida como um grafo conexo e acíclico. Para o grafo $T = (N, E)$, as seguintes propriedades são válidas:

- (i) T é uma árvore.
- (ii) Existe apenas um caminho entre dois vértices quaisquer.
- (iii) T é um grafo conexo e possui $N - 1$ arestas.

A Figura 2.10 ilustra uma árvore com 9 vértices e 8 arestas. Nesta Figura, as propriedades supracitadas podem ser facilmente verificadas.

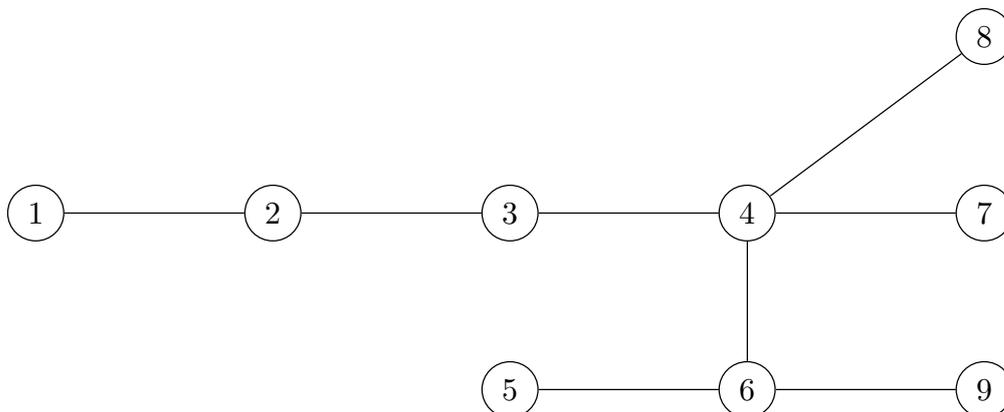


Figura 2.10: Exemplo de uma árvore com 9 vértices 8 arestas.

3

Vines regulares

Este capítulo apresenta as cópulas vine regulares (*regular vine copula - R-vine*). Estes modelos foram desenvolvidos a partir do trabalho de Joe [32]. Posteriormente, Bedford & Cooke [14,15] propuseram um método gráfico (*vine models*) para organizar a construção pair-cópula. Todavia, somente após Aas et al. [13] abordarem aspectos inferenciais é que tais modelos começaram a ser amplamente utilizados.

Uma vine com d variáveis é um conjunto de árvores onde as arestas da árvore ℓ (T_ℓ) são os vértices das árvores $\ell + 1$. Uma vine regular (R-vine) é um caso especial de vine onde duas arestas de T_ℓ estão conetadas em $T_{\ell+1}$ somente se estas arestas possuem um vértice em comum na T_ℓ . [28].

O grande atrativo das vines regulares é a possibilidade de organizar e representar a construção pair-cópula (PCC). Com esta representação, é possível verificar se determinada construção é válida ou não, por exemplo. A junção das PCCs às vines regulares dão origem aos modelos de cópula vine regular. De maneira intuitiva, o que se faz é associar distribuições (incondicionais e condicionais) e cópulas bivariadas (incondicionais e condicionais) oriundas da decomposição apresentada na seção 2.1.2 aos vértices e arestas de um conjunto de árvores.

A apresentação a seguir resume alguns resultados apresentados por Kurowicka & Cooke [28]. Adicionalmente, utilizamos também as referências [37,43,44] para a elaboração desse capítulo.

3.1

Distribuição Rvine

Para que seja possível representar uma construção pair-cópula graficamente, é necessário que façamos algumas definições. Iniciaremos com o conceito de vine regular.

Definição 3.1.1 (Vine Regular) $\mathcal{V} = (T_1, T_2, \dots, T_{d-1})$ é uma vine regular d -dimensional se

- i. T_1 é uma árvore conexa 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 o conjunto de vértices é dado por $N_\ell = E_{\ell-1}$ e o de arestas por E_ℓ .
- iii. para $\ell = 2, \dots, d-1$, e $\{a, b\} \in E_i$ onde $a = \{a_1, a_2\}$ e $b = \{b_1, b_2\}$, a condição $\#(a \cap b) = 1$ deve ser satisfeita (condição de proximidade).

A condição de proximidade diz que se dois vértices estão conectados por uma aresta em T_ℓ , $\ell > 2$, as correspondentes arestas em $T_{\ell-1}$ devem possuir um vértice em comum. Vale lembrar que arestas em T_ℓ serão vértices em $T_{\ell+1}$. Esta condição é fundamental pois estabelece a diferença entre vines e vines regulares. Os dois primeiros pontos da definição acima definem uma vine. Uma vine regular é obtida ao se adicionar a terceira condição supracitada.

A Figura 3.1 ilustra um exemplo de vine regular em 4 dimensões ($d = 4$). Como pode ser observado, esta é composta por 3 árvores ($T_\ell, \ell = 1, 2, 3$), onde cada árvore (T_ℓ) possui $d - \ell + 1$ vértices e $d - \ell$ arestas. Os conjuntos de vértices e arestas que compõem esse exemplo são

$$\begin{aligned}
 N_1 &= \{1, 2, 3, 4\}, \\
 N_2 = E_1 &= \{\{1, 2\}, \{2, 4\}, \{2, 3\}\}, \\
 N_3 = E_2 &= \{\{\{1, 2\}, \{2, 4\}\}, \{\{2, 4\}, \{2, 3\}\}\}, \\
 N_4 = E_3 &= \{\{\{\{1, 2\}, \{2, 4\}\}, \{\{2, 4\}, \{2, 3\}\}\}\}. \tag{3-1}
 \end{aligned}$$

A condição de proximidade é melhor entendida ao se observar a Figura 3.1. Note, por exemplo, que na árvore T_2 , a aresta $\{\{1, 2\}, \{2, 4\}\}$ conecta os vértices $\{1, 2\}$ e $\{2, 4\}$ e que estes, possuem o vértice $\{2\}$ em comum na árvore T_1 .

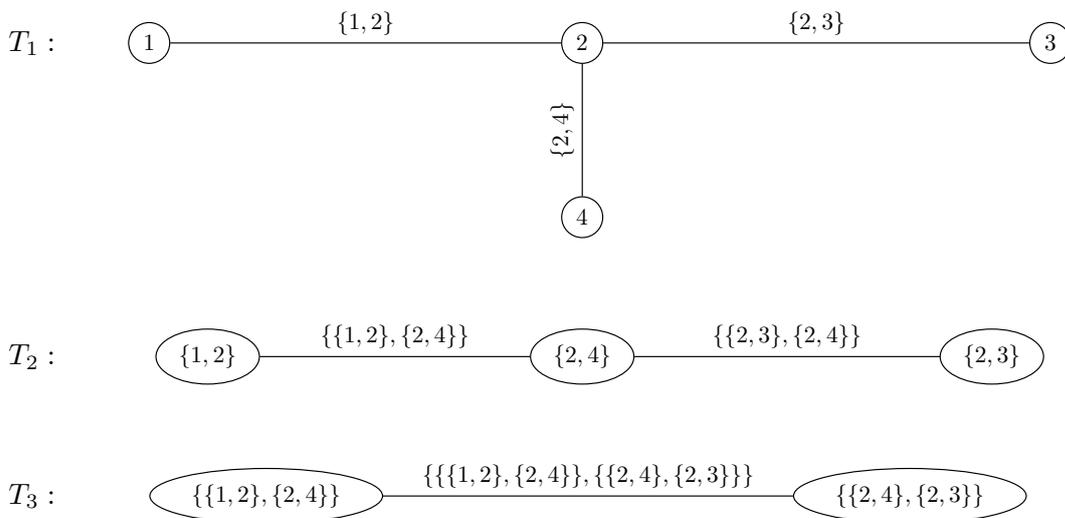


Figura 3.1: Vine regular com 4 dimensões.

As *C-vines* e *D-vines* são dois casos particulares de vines regulares. Uma *C-Vine* ocorre quando cada árvore $(T_\ell, \ell = 1, \dots, d-1)$ possui um vértice com grau $d - \ell$. Por outro lado, uma *D-Vine* existe quando o grau de cada vértice da árvore T_1 é no máximo 2. Na Figura 3.2 são apresentadas as formas gráficas de uma *C-Vine* e uma *D-Vine*.

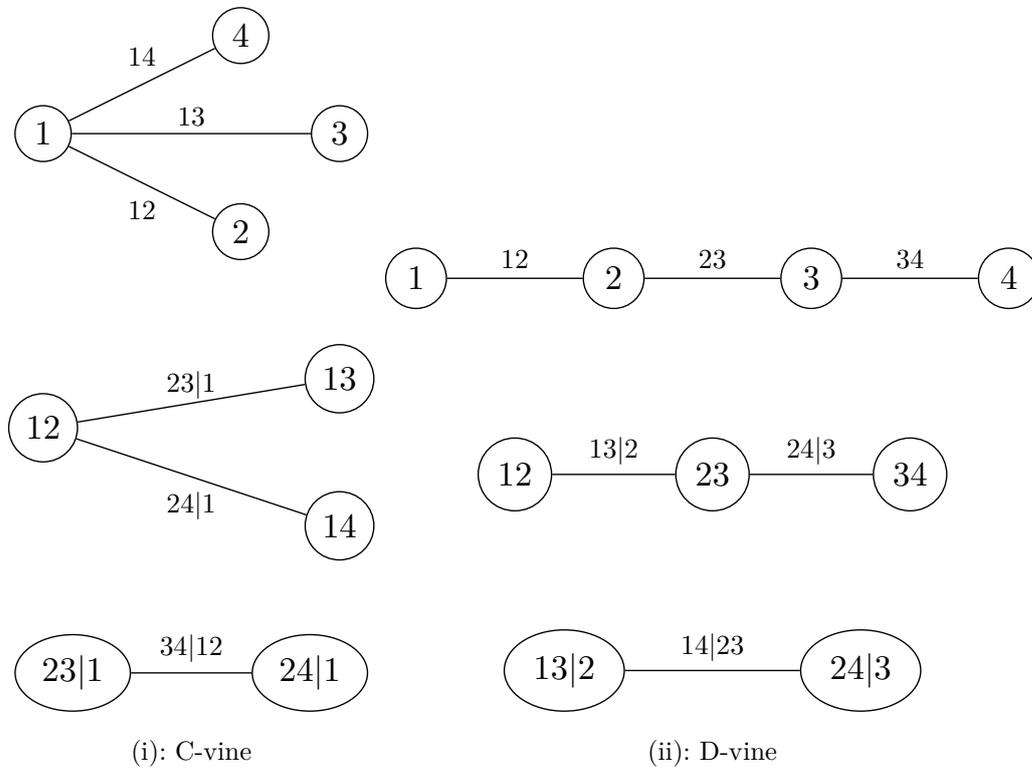


Figura 3.2: Exemplos de C-vine e D-vine.

Vale a pena destacar que até este momento, nenhuma informação sobre cópulas ou variáveis aleatórias foi mencionada. Os conceitos apresentados compõem simplesmente um modelo gráfico. Para organizar a construção pair-cópula (PCC) utilizando as vines regulares, é necessário que façamos algumas definições adicionais. Estas definições estabelecem uma notação particular para as R-vines e tornam o link entre as R-vines e a PCC intuitivo.

Os conceitos de união completa (complete union), conjunto condicionante (*conditioning set*) e conjunto condicionado (*conditioned set*) são associados a cada aresta $e = \{a, b\}$ da vine. De maneira intuitiva, a união completa de uma aresta e é formada pelos vértices em N_1 'alcançáveis' a partir da aresta e . Se dois vértices a e b são conectadas por um aresta e , a interseção da união completa de a com a união completa de b formam o conjunto condicionante da aresta e , enquanto que a diferença simétrica determina o conjunto condicionado. [45].

Definição 3.1.2 (*União completa, conjuntos condicionante, condicionado e restrição*)

Para uma aresta arbitrária $e = \{a, b\} \in E_\ell$, $\ell = 1, \dots, d-1$, temos as seguintes definições:

(i) A união completa da aresta e é definida por

$$\mathcal{U}_e := \{j \in N_1 \mid \exists e_1 \in E_1, \dots, e_{\ell-1} \in E_{\ell-1} : j \in e_1 \in \dots \in e_{\ell-1} \in e\}. \quad (3-2)$$

(ii) O conjunto condicionante da aresta $e = \{a, b\}$ é

$$\mathcal{D}_e := \mathcal{U}_a \cap \mathcal{U}_b. \quad (3-3)$$

(iii) O conjunto condicionado é dado por

$$\mathcal{C}_e := \mathcal{C}_{e,a} \cup \mathcal{C}_{e,b} \quad (3-4)$$

onde

$$\mathcal{C}_{e,a} := \mathcal{U}_a \setminus \mathcal{D}_e \quad e \quad \mathcal{C}_{e,b} := \mathcal{U}_b \setminus \mathcal{D}_e. \quad (3-5)$$

Por fim, o conjunto restrição de uma R-vine \mathcal{V} é

$$CR(\mathcal{V}) := \{\{\mathcal{C}_{e,a}, \mathcal{C}_{e,b} \mid \mathcal{D}_e\} : e = \{a, b\} \in E_\ell, \ell = 1, \dots, d-1\}. \quad (3-6)$$

Para ilustrar os conjuntos estabelecidos na definição 3.1.2, utilizaremos novamente a Figura 3.1 com os vértices e arestas já apresentados na equação (3-1). Como exemplo, utilizaremos a única aresta presente em T_3 ; $e = \{a, b\}$, onde $a = \{\{1, 2\}, \{2, 4\}\}$ e $b = \{\{2, 4\}, \{2, 3\}\}$. A união completa de a e b são

$$\mathcal{U}_a = \{1, 2, 4\} \quad e \quad \mathcal{U}_b = \{2, 3, 4\}. \quad (3-7)$$

O conjunto condicionante da aresta e é

$$\mathcal{D}_e = \mathcal{U}_a \cap \mathcal{U}_b = \{1, 2, 4\} \cap \{2, 3, 4\} = \{2, 4\}. \quad (3-8)$$

Por fim, o conjunto condicionado para a aresta selecionada é

$$\mathcal{C}_e = \mathcal{C}_{e,a} \cup \mathcal{C}_{e,b} \tag{3-9}$$

onde

$$\mathcal{C}_{e,a} = \mathcal{U}_a \setminus \mathcal{D}_e = \{1, 2, 4\} \setminus \{2, 4\} = \{1\} \tag{3-10}$$

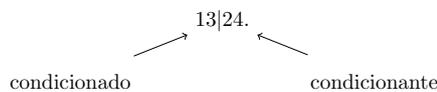
e

$$\mathcal{C}_{e,b} = \mathcal{U}_b \setminus \mathcal{D}_e = \{2, 3, 4\} \setminus \{2, 4\} = \{3\}. \tag{3-11}$$

Logo,

$$\mathcal{C}_e = \mathcal{C}_{e,a} \cup \mathcal{C}_{e,b} = \{1, 3\}. \tag{3-12}$$

A partir deste momento, adotaremos a notação usual utilizada para nomear os vértices e arestas das *R-vines*. Desta forma os conjuntos condicionados e condicionantes serão representados nas árvores como $\mathcal{C}_{e,a}, \mathcal{C}_{e,b} | \mathcal{D}_e$. Por exemplo, utilizando a mesma aresta e do exemplo anterior, temos:



A título de ilustração, renomeamos a Figura 3.1 utilizando a notação usualmente adotada na literatura. A Figura 3.3 ilustra essa nomenclatura.

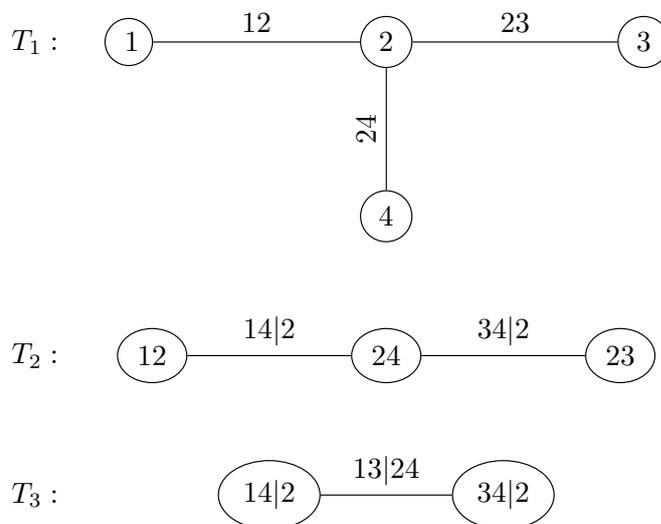


Figura 3.3: *R-vine* quadridimensional utilizando $\mathcal{C}_{e,a}, \mathcal{C}_{e,b} | \mathcal{D}_e$.

A vine regular (*R-vine*) apresentada nada mais é do que um conjunto de árvores cujos vértices e arestas possuem uma nomenclatura particular. Até este momento, não existe nenhuma relação entre a construção pair-cópula e as vines regulares. Entretanto, é intuitivo observar que a nomenclatura dos vértices e arestas se assemelha bastante aos índices das cópulas bivariadas da PCC. De fato, o componente estocástico é introduzido ao associarmos a cada umas das arestas cópulas bivariadas (pair-cópula). Dessa forma, apresentamos agora o link entre um conjunto de variáveis aleatórias $\mathbf{X} = (X_1, \dots, X_d)$ e um R-vine.

Definição 3.1.3 (Especificação cópula R-vine. [45].) $(\mathcal{F}, \mathcal{V}, \mathcal{B})$ é uma especificação de cópula R-vine se:

- (i) Distribuições marginais: $\mathcal{F} = (F_1, \dots, F_d)$ é um vetor de d funções de distribuição marginais univariadas contínuas e invertíveis correspondentes à (X_1, \dots, X_d) .
- (ii) Vine regular: \mathcal{V} é uma R-vine d -dimensional.
- (iii) Cópulas bivariadas: $\mathcal{B} = \{B_e | \ell = 1, \dots, d-1; e \in E_\ell\}$ é um conjunto de cópulas. B_e é uma cópula bivariada e E_ℓ , $\ell = 1, \dots, d-1$ são os conjuntos de arestas da R-vine \mathcal{V} .

A distribuição conjunta \mathbf{F} de um vetor aleatório $\mathbf{X} = (X_1, \dots, X_d)$ é uma realização de uma especificação cópula R-vine se para cada $e \in E_\ell$, $\ell = 1, \dots, d-1$, $e = \{a, b\}$, B_e é a cópula bivariada de $X_{\mathcal{C}_{e,a}}$ e $X_{\mathcal{C}_{e,b}}$ dado $\mathbf{X}_{\mathcal{D}_e} = \{X_k | k \in \mathcal{D}_e\}$. Esta distribuição é denominada distribuição R-vine. [45].

Por exemplo, para a aresta $e = \{\{2, 4\}, \{2, 3\}\}$ (Figura 3.1) temos que $\mathcal{C}_{e,a} = \{3\}$, $\mathcal{C}_{e,b} = \{4\}$ e $\mathcal{D}_e = \{2\}$. Se assumirmos que B_e é uma cópula Gumbel bivariada, então a dependência entre as variáveis X_3 e X_4 condicionadas a $X_2 = x_2$ é representada pela referida cópula Gumbel. Isso implica dizer que a distribuição condicional de $(X_3, X_4 | X_2 = x_2)$ é dada por

$$F(x_3, x_4 | x_2) = B_e(F(x_3 | x_2), F(x_4 | x_2)). \quad (3-13)$$

Uma vez estabelecida a relação entre as cópulas bivariadas e os componentes das árvores, a literatura comumente denota a cópula B_e correspondente a aresta e por $C_{\mathcal{C}_{e,a}, \mathcal{C}_{e,b} | \mathcal{D}_e}$ e $c_{\mathcal{C}_{e,a}, \mathcal{C}_{e,b} | \mathcal{D}_e}$ (densidade cópula). [37, 46].

Por fim, é interessante que expressemos a distribuição R-vine em termos de sua densidade.

Teorema 3.1.1 *Assuma que $(\mathcal{F}, \mathcal{V}, \mathcal{B})$ satisfaça uma especificação cópula R-vine de d elementos. Então existe uma única distribuição R-vine \mathbf{F} com*

densidade dada por

$$f(x_1, \dots, x_d) = \prod_{k=1}^d f_k(x_k) \prod_{l=1}^{d-1} \prod_{e \in E_l} c_{\mathcal{C}_{e,a}, \mathcal{C}_{e,b} | \mathcal{D}_e} (F_{\mathcal{C}_{e,a} | \mathcal{D}_e}(x_{\mathcal{C}_{e,a}} | \mathbf{x}_{\mathcal{D}_e}), F_{\mathcal{C}_{e,b} | \mathcal{D}_e}(x_{\mathcal{C}_{e,b}} | \mathbf{x}_{\mathcal{D}_e})), \quad (3-14)$$

onde $e = \{a, b\}$, $\mathbf{x}_{\mathcal{D}_e} = \{x_i | i \in D_e\}$ e f_k são as densidades marginais univariadas de F_k , $k = 1, \dots, d$.

Para avaliar esta densidade (equação (3-14)), é necessário que calculemos as distribuições condicionais $F_{\mathcal{C}_{e,a} | \mathcal{D}_e}(x_{\mathcal{C}_{e,a}} | \mathbf{x}_{\mathcal{D}_e})$ e $F_{\mathcal{C}_{e,b} | \mathcal{D}_e}(x_{\mathcal{C}_{e,b}} | \mathbf{x}_{\mathcal{D}_e})$ (variáveis transformadas). Para tal, considere $E_\ell \ni e = \{a, b\}$, $a = \{a_1, a_2\}$, $b = \{b_1, b_2\}$ como sendo a aresta que conecta $\mathcal{C}_{e,a}$ com $\mathcal{C}_{e,b}$ dado as variáveis \mathcal{D}_e . Dessa forma, as variáveis transformadas podem ser calculadas como

$$\begin{aligned} F_{\mathcal{C}_{e,a} | \mathcal{D}_e}(x_{\mathcal{C}_{e,a}} | \mathbf{x}_{\mathcal{D}_e}) &= \frac{\partial C_{\mathcal{C}_a | \mathcal{D}_a}(F_{\mathcal{C}_{a,a_1} | \mathcal{D}_a}(x_{\mathcal{C}_{a,a_1}} | \mathbf{x}_{\mathcal{D}_a}), F_{\mathcal{C}_{a,a_2} | \mathcal{D}_a}(x_{\mathcal{C}_{a,a_2}} | \mathbf{x}_{\mathcal{D}_a}))}{\partial F_{\mathcal{C}_{a,a_2} | \mathcal{D}_a}(x_{\mathcal{C}_{a,a_2}} | \mathbf{x}_{\mathcal{D}_a})} \\ &= h(F_{\mathcal{C}_{a,a_1} | \mathcal{D}_a}(x_{\mathcal{C}_{a,a_1}} | \mathbf{x}_{\mathcal{D}_a}), F_{\mathcal{C}_{a,a_2} | \mathcal{D}_a}(x_{\mathcal{C}_{a,a_2}} | \mathbf{x}_{\mathcal{D}_a})) \end{aligned} \quad (3-15)$$

onde $F_{\mathcal{C}_{a,a_1} | \mathcal{D}_a}(x_{\mathcal{C}_{a,a_1}} | \mathbf{x}_{\mathcal{D}_a})$ e $F_{\mathcal{C}_{a,a_2} | \mathcal{D}_a}(x_{\mathcal{C}_{a,a_2}} | \mathbf{x}_{\mathcal{D}_a})$ devem ser obtidas recursivamente. [45]. A equação 3-15 corresponde a equação (2-22) reescrita utilizando a nova notação. $F_{\mathcal{C}_{e,b} | \mathcal{D}_e}(x_{\mathcal{C}_{e,b}} | \mathbf{x}_{\mathcal{D}_e})$ é obtida de maneira similar. [45].

3.2

Matrizes vine regular

O modelo gráfico proposto por Bedford & Cooke [14, 15] facilita consideravelmente a representação da construção pair-cópula. Por outro lado, em termos computacionais, esta representação pode não ser a mais eficiente. Uma representação das *R-vines* em forma matricial foi sugerida por Dißmann [43]. Esta representação é conhecida na literatura como matrizes de R-vine (*Regular Vine Matrices*). Nesta seção resumimos alguns resultados de Dißmann [43].

A ideia é que a estrutura das árvores (conjuntos condicionais e condicionantes - $\{\{\mathcal{C}_{e,a}, \mathcal{C}_{e,b} | \mathcal{D}_e\}\}$ e o conjunto restrição $CR(\mathcal{V})$) sejam armazenados em uma matriz. Isto permite a elaboração de métodos computacionalmente eficientes para inferência, estimação e simulação das *R-vines*.

Iniciaremos esta exposição afirmando que $\mathcal{M} \in \{1, \dots, d\}^{d \times d}$ é uma matriz triangular inferior. Cada elemento dessa matriz pode receber um número inteiro entre 1 e d . Para facilitar o entendimento, utilizaremos como exemplo a seguinte matriz \mathcal{M}^* :

$$\mathcal{M}^* = \begin{pmatrix} 1 & & & & & \\ 5 & 5 & & & & \\ 3 & 4 & 4 & & & \\ 4 & 2 & 3 & 3 & & \\ 2 & 3 & 2 & 2 & 2 & \end{pmatrix}. \quad (3-16)$$

Como dito anteriormente, uma distribuição *R-vine* é caracterizada pelos conjuntos restrição, condicionado e condicionante. Dessa maneira, sua versão matricial também deve ser capaz de armazenar tais informações. Dißmann [43] define as propriedades que uma matriz deve possuir para ser uma matriz de vine regular. Além disso, define as respectivas representações matriciais desses conjuntos e prova que existe uma matriz de vine regular para toda vine regular e vice versa. Por fim, o autor estabelece algoritmos para transformar R-vines em matrizes R-vine.

Definição 3.2.1 (Conjunto restrição - matriz R-vine) *Seja $\mathcal{M} \in \{1, \dots, \}^{d \times d}$ uma matriz triangular inferior cujo elementos são $m_{ij}, j \leq i$. Dessa forma o j -ésimo conjunto restrição de \mathcal{M} é*

$$\mathcal{C}_M(j) = \{(\{m_{j,j}, m_{k,j}\}, D) | k = j + 1, \dots, d, D = \{m_{k+1,j}, \dots, m_{d,j}\}\} \quad (3-17)$$

para $j = 1, \dots, d - 1$. Quando $k = d$, temos $D = \{\}$. O conjunto restrição da matriz \mathcal{M} é a união de todos os $\mathcal{C}_M(j)$, ou seja,

$$\mathcal{C}_M = \mathcal{C}_M(1) \cup \mathcal{C}_M(2) \cup \dots \cup \mathcal{C}_M(d - 1). \quad (3-18)$$

Para os elementos do conjunto restrição $\{(\{m_{j,j}, m_{k,j}\}, D), j = 1, \dots, d - 1, k = j + 1, \dots, d$, o conjunto condicionado é $\{m_{j,j}, m_{k,j}\}$ enquanto que o conjunto condicionante é D .

A Tabela 3.1 apresenta o conjunto restrição para a matriz apresentada em 3-16.

$\mathcal{C}_M(1)$	$\mathcal{C}_M(2)$	$\mathcal{C}_M(3)$	$\mathcal{C}_M(4)$
$(\{1,5\},\{3,4,2\})$	$(\{5,4\},\{2,3\})$	$(\{4,3\},\{2\})$	$(\{3,2\},\{\})$
$(\{1,3\},\{4,2\})$	$(\{5,2\},\{3\})$	$(\{4,2\},\{\})$	
$(\{1,4\},\{2\})$	$(\{5,3\},\{\})$		
$(\{1,2\},\{\})$			

Tabela 3.1: Conjunto restrição da matriz \mathcal{M}^* .

Além do conjunto restrição, antes de definirmos formalmente uma matriz R-vine, é necessário definirmos alguns conjuntos adicionais. Desta forma, seja $\mathcal{M} \in \{1, \dots, \}^{d \times d}$ uma matriz triangular inferior. Para $j = 1, \dots, d - 1$, temos

$$B_M(j) = \{(m_{j,j}, D) | D = \{m_{k,j}, \dots, m_{d,j}\}; k = j + 1, \dots, d\} \quad (3-19)$$

e

$$\tilde{B}_M(j) = \{(m_{k,j}, D) | D = \{m_{j,j}\} \cup \{m_{k+1,j}, \dots, m_{d,j}\}; k = j+1, \dots, d\}. \quad (3-20)$$

As Tabelas 3.2 e 3.3 ilustram os conjuntos $B_M(j)$ e $\tilde{B}_M(j)$.

$B_M(1)$	$B_M(2)$	$B_M(3)$	$B_M(4)$
$(1, \{5, 2, 3, 4\})$	$(5, \{4, 2, 3\})$	$(4, \{3, 2\})$	$(3, \{2\})$
$(1, \{3, 2, 4\})$	$(5, \{2, 3\})$	$(4, \{2\})$	
$(1, \{4, 2\})$	$(5, \{3\})$		
$(1, \{2\})$			

Tabela 3.2: Elementos de $B_M(j)$ utilizando como exemplo \mathcal{M}^* .

$\tilde{B}_M(1)$	$\tilde{B}_M(2)$	$\tilde{B}_M(3)$	$\tilde{B}_M(4)$
$(5, \{1, 2, 3, 4\})$	$(4, \{5, 2, 3\})$	$(3, \{4, 2\})$	$(2, \{3\})$
$(3, \{1, 2, 4\})$	$(2, \{5, 3\})$	$(2, \{4\})$	
$(4, \{1, 2\})$	$(3, \{5\})$		
$(2, \{1\})$			

Tabela 3.3: Elementos de $\tilde{B}_M(j)$ utilizando como exemplo \mathcal{M}^* .

Isto posto, apresentamos agora a matriz R-vine.

Definição 3.2.2 (Matrizes de Rvine - R-vine matrices) *Seja $\mathcal{M} \in \{1, \dots, \}^{d \times d}$ uma matriz triangular inferior. M é uma matriz R-vine se para $i = 1, \dots, d - 2$ e para todo $k = i + 1, \dots, d - 1$ existe um $j \in \{i + 1, \dots, d - 1\}$ de modo que*

$$(m_{k,i}, \{m_{k+1,i}, \dots, m_{d,i}\}) \in B_M(j) \quad \text{ou} \quad (m_{k,i}, \{m_{k+1,i}, \dots, m_{d,i}\}) \in \tilde{B}_M(j). \quad (3-21)$$

A definição 3.2.2 implica algumas propriedades interessantes. A primeira diz que todos os elementos de determinada coluna estarão presentes em todas as colunas à sua esquerda, ou seja,

$$\{m_{i,i}, \dots, m_{d,i}\} \subset \{m_{j,j}, \dots, m_{d,j}\} \quad \text{para} \quad 1 \leq j < i \leq d. \quad (3-22)$$

Além disso, a diagonal principal sempre possui um elemento novo em relação às colunas de maiores índices, ou seja,

$$\{m_{i,i}\} \notin \{m_{i+1,i+1}, \dots, m_{d,i+1}\} \quad \text{para } i = 1, \dots, d-1. \quad (3-23)$$

Estas propriedades podem ser verificadas na matriz \mathcal{M}^* (3-16). Além disso, podemos afirmar que \mathcal{M}^* é uma matriz R-vine pois satisfaz todas as condições presentes na definição 3.2.2. A Tabela 3.4 apresenta os elementos $(m_{k,i}, \{m_{k+1,i}, \dots, m_{d,i}\})$, para $i = 1, \dots, 3$ e $k = i + 1, \dots, 4$, obtidos a partir da matriz \mathcal{M}^* . Com o auxílio das Tabelas 3.2 e 3.3, podemos perceber que $(5, \{2, 3, 4\}) \in B_M(2)$, $(3, \{2, 4\}) \in \tilde{B}_M(3)$, $(4, \{2\}) \in B_M(3)$, $(4, \{2, 3\}) \in B_M(3)$, $(2, \{3\}) \in \tilde{B}_M(4)$ e $(3, \{2\}) \in B_M(4)$.

	$i = 1$	$i = 2$	$i = 3$
$k = 2$	$(5, \{2, 3, 4\})$		
$k = 3$	$(3, \{2, 4\})$	$(4, \{2, 3\})$	
$k = 4$	$(4, \{2\})$	$(2, \{3\})$	$(3, \{2\})$

Tabela 3.4: Elementos $(m_{k,i}, \{m_{k+1,i}, \dots, m_{d,i}\})$ da \mathcal{M}^* .

A representação matricial das R-vines apresentada carrega informações acerca das estrutura da cópula vine. Além do mais, é necessário que armazenemos informações sobre as famílias das cópulas bivariadas e seus respectivos parâmetros. De maneira similar, tais informações podem ser organizadas em um conjunto de matrizes adicionais. Na próxima seção apresentaremos como todas estas informações são de fato organizadas e como estas podem ser lidas.

3.3

Famílias de cópulas e seus parâmetros em forma matricial

Retomemos a matriz \mathcal{M}^* apresentada anteriormente,

$$\mathcal{M}^* = \begin{pmatrix} 1 & & & & & \\ 5 & 5 & & & & \\ 3 & 4 & 4 & & & \\ 4 & 2 & 3 & 3 & & \\ 2 & 3 & 2 & 2 & 2 & \end{pmatrix}.$$

Esta matriz representa a seguinte estrutura:

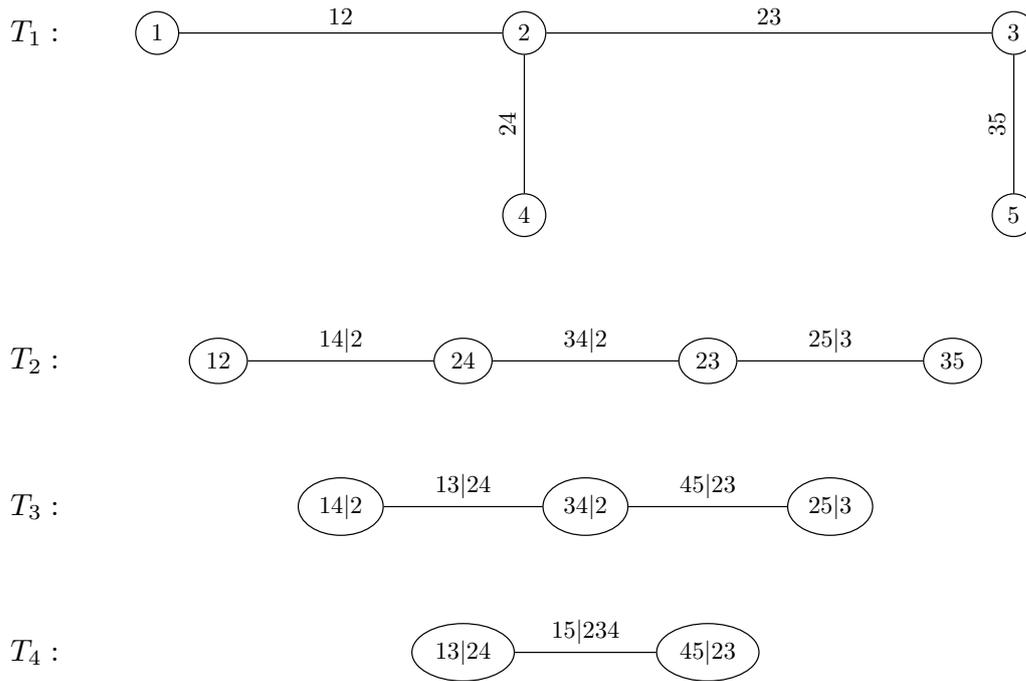


Figura 3.4: Vine regular com 5 variáveis.

A leitura da matriz \mathcal{M}^* é bem simples. Cada linha em conjunto com os elementos da diagonal principal carregam informações sobre as árvores. Esta matriz deve ser interpretada de baixo para cima, de modo que a última linha corresponde a primeira árvore, a penúltima linha corresponde a segunda árvore e assim por diante.

Para determinada coluna, a combinação do elemento da linha com o da diagonal principal formam o conjunto condicionado. Os elementos abaixo dessa linha específica compõem o conjunto condicionante. Por exemplo, considerando a linha 5 de \mathcal{M}^* , podemos concluir que as cópulas que compõem a T_1 são $C_{12}, C_{35}, C_{24}, C_{23}$. Comparando com a Figura 3.4, estas cópulas são precisamente as modeladas na primeira árvore. Vale lembrar que o conjunto condicionante é vazio neste caso.

Considere agora a linha 3. Esta possui informações sobre T_3 . Deste modo, temos as cópulas $C_{1,3|2,4}$ e $C_{4,5|2,3}$, rigorosamente como na Figura 3.4. Tendo em vista a cópula $C_{4,5|2,3}$, a variável X_5 é o elemento da diagonal principal da coluna 2 e a variável X_4 é o elemento da linha 3 coluna 2. Os elementos que pertencem à coluna 2 e que estão abaixo da linha 3 formam o conjunto condicionante e são as variáveis X_2 e X_3 .

Embora a partir da matriz \mathcal{M}^* conhecemos a estrutura das árvores, e conseqüentemente sua construção, ainda carecemos de informações acerca das cópulas bivariadas e seus parâmetros. De modo similar, tais informações são armazenadas em matrizes adicionais.

Com relação as famílias das cópulas bivariadas, estas podem ser organizadas em uma matriz triangular inferior $\mathcal{Z} = (z_{i,j})_{i,j=1,\dots,d}$. A título ilustrativo, apresentamos a matriz \mathcal{Z}^* ,

$$\mathcal{Z}^* = \begin{pmatrix} \Pi & & & & \\ \Pi & F & & & \\ G & N & C & & \\ t & t & C & N & \end{pmatrix}, \quad (3-24)$$

- onde: $\Pi \rightarrow$ cópula de independência.
- $N \rightarrow$ cópula Gaussiana (normal).
- $t \rightarrow$ cópula t-Student.
- $C \rightarrow$ cópula Clayton.
- $G \rightarrow$ cópula Gumbel.

Neste exemplo, a cópula bivariada entre as variáveis X_1 e X_2 seria uma cópula t-Student. A cópula C_{24} seria uma cópula Clayton, enquanto que a $C_{14|2}$ seria uma cópula Gumbel.

Os respectivos parâmetros são organizados em outras matrizes. O número de matrizes necessárias depende do número de parâmetros que as cópulas possuem. Assim sendo, temos $P^w = (p_{i,j}^w)_{i,j=1,\dots,d}$, $1 \leq w \leq \mathcal{W}$, onde \mathcal{W} é o número máximo de parâmetros existente em determinada cópula bivariada. Caso somente cópulas com um único parâmetro sejam utilizadas, as matrizes \mathcal{M} , \mathcal{Z} e \mathcal{P}^1 especificam todo o modelo.

Continuando com o exemplo, as famílias presentes na matriz \mathcal{Z}^* possuem apenas um ou dois parâmetros (cópula t-Student). Isto torna necessário o uso de duas matrizes \mathcal{P}^1 e \mathcal{P}^2 para armazenar todos os parâmetros. Utilizando as matrizes $\mathcal{P}^{1,*}$ e $\mathcal{P}^{2,*}$ como exemplo, temos:

$$\mathcal{P}^{1,*} = \begin{pmatrix} 0 & & & & \\ 0 & -3.1 & & & \\ 4.8 & 0.4 & 1.5 & & \\ 0.8 & 0.5 & 2.0 & 0.7 & \end{pmatrix}, \mathcal{P}^{2,*} = \begin{pmatrix} 0 & & & & \\ 0 & 0 & & & \\ 0 & 0 & 0 & & \\ 4.5 & 6.7 & 0 & 0 & \end{pmatrix} \quad (3-25)$$

Como já mencionado, todas estas matrizes definem a cópula vine. Desta forma, podemos dizer que a cópula entre as variáveis X_1 e X_2 é uma cópula t-Student com $p = 0.8$ e $\nu = 4.5$. A cópula C_{24} é uma cópula Clayton com parâmetro igual a 2. A cópula $C_{25|3}$ é uma Gaussiana com correlação igual a

0.4.

Por fim, vale destacar que é possível expressar a densidade de uma distribuição R-vine em termos de uma matriz R-vine. Considerando as matrizes \mathcal{M} , \mathcal{Z} , \mathcal{P}^w , a seguinte expressão corresponde a densidade de uma distribuição R-vine:

$$f(x_1, \dots, x_d) = \prod_{k=1}^d f(x_k) \prod_{j=1}^{d-1} \prod_{i=j+1}^d c_{m_{j,j}, m_{i,j} | m_{i+1,j}, \dots, m_{d,j}}(\tilde{u}_{j,j}, \tilde{u}_{i,j}; p_{i,j}^1, \dots, p_{i,j}^w, z_{i,j}) \quad (3-26)$$

onde f_k , $k = 1, \dots, d$ representam as distribuições marginais de \mathbf{F} ,

$$\tilde{u}_{j,j} := F_{m_{j,j} | m_{i+1,j}, \dots, m_{d,j}}(x_{m_{j,j}} | x_{m_{i+1,j}}, \dots, m_{d,j}) \text{ e}$$

$$\tilde{u}_{i,j} := F_{m_{i,j} | m_{i+1,j}, \dots, m_{d,j}}(x_{m_{i,j}} | x_{m_{i+1,j}}, \dots, m_{d,j}).$$

3.4 Simulação

Em muitas situações é interessante simular novas observações a partir de uma cópula vine ajustada. Nos Capítulos 4 e 5, a simulação constituirá etapa fundamental para a geração de cenários vazões.

A simulação a partir de uma cópula vine não é uma das tarefas mais difíceis. Esta é realizada através de um procedimento iterativo baseado em uma sequência de distribuições condicionais.

Inicialmente apresentaremos como este procedimento é utilizado para simular observações de uma distribuição multivariada e, em seguida, como este é adaptado para o caso das cópulas. Por fim, ilustraremos o caso das cópulas vine, onde as cópulas condicionais são escritas em termos de cópulas bivariadas.

Esta seção expõe a lógica que rege o algoritmo geral de simulação. O objetivo é ilustrar e esclarecer como este funciona. Consideraremos, sem perda de generalidade, distribuições com poucas variáveis. Obviamente, em casos d -dimensionais tal procedimento precisa ser estabelecido na forma de algoritmos programáveis. Aas et al. [13] estabeleceram algoritmos computacionais para as C-vines e D-vines. Dißmann [43] e Dißmann [45] foram os primeiros a estabelecer tais algoritmos para o caso mais geral de R-vines.

Nesta tese, os algoritmos apresentados em Mai & Scherer [46] foram empregados. Estes são versões mais eficientes do que os supracitados onde cálculos redundantes foram eliminados.

Para esta seção, vamos considerar uma distribuição multivariada $F_{1, \dots, d}(x_1, \dots, x_d)$ com marginais iguais à $F_i(x_i)$, $i = 1, \dots, d$. Neste caso, $F_{2|1}(\cdot | x_1)$, $F_{3|2,1}(\cdot | x_2, x_1)$, \dots , $F_{d|d-1, \dots, 1}(\cdot | x_{d-1}, \dots, x_1)$ são as distribuições condicionais e $F_{2|1}^{-1}(\cdot | x_1)$, $F_{3|2,1}^{-1}(\cdot | x_2, x_1)$, \dots , $F_{d|d-1, \dots, 1}^{-1}(\cdot | x_{d-1}, \dots, x_1)$ são as res-

pectivas funções inversas.

Para amostrarmos observações provenientes da distribuição $F_{1,\dots,d}$, devemos iterar as distribuições condicionais inversas da seguinte maneira:

1º passo: Amostrar d observações independentes de uma distribuição uniforme padrão, i.e., $w_i \stackrel{\text{i.i.d}}{\sim} U(0, 1), i = 1, \dots, d$.

2º passo: Iterar as distribuições inversas.

$$\begin{aligned} x_1 &= w_1, \\ x_2 &= F_{2|1}^{-1}(w_2|x_1), \\ x_3 &= F_{3|2,1}^{-1}(w_3|x_2, x_1), \\ &\vdots \\ x_d &= F_{d|d-1,\dots,1}^{-1}(w_d|x_{d-1}, \dots, x_1). \end{aligned}$$

As observações resultantes (x_1, \dots, x_d) seguem a distribuição multivariada $F_{1,\dots,d}$.

Este procedimento é trivialmente adaptável para o caso das cópulas. Assim, considere uma distribuição cópula multivariada $C_{1,\dots,d}$ de um vetor aleatório $U = (U_1, \dots, U_d) \in [0, 1]^d$ com $U_1, \dots, U_d \sim U(0, 1)$. Para obtermos uma amostra (u_1, \dots, u_d) proveniente da cópula $C_{1,\dots,d}$, devemos proceder de maneira similar ao caso anterior.

1º passo: Amostrar d observações independentes de uma distribuição uniforme padrão, i.e., $w_i \stackrel{\text{i.i.d}}{\sim} U(0, 1), i = 1, \dots, d$.

2º passo: Iterar as cópulas condicionais inversas.

$$\begin{aligned} u_1 &= w_1. \\ u_2 &= C_{2|1}^{-1}(w_2|u_1), \\ u_3 &= C_{3|2,1}^{-1}(w_3|u_2, u_1), \\ &\vdots \\ u_d &= C_{d|d-1,\dots,1}^{-1}(w_d|u_{d-1}, \dots, u_1). \end{aligned}$$

Existem alguns outros métodos para simulação de cópulas (veja [24]). Entretanto, através desse procedimento, a simulação cópulas vine é muito simples, uma vez que as cópulas condicionais podem ser escritas em termos de cópulas bivariadas. Para isto, basta empregarmos recursivamente a equação (2-22) em conjunto com a definição de *h-functions* (equação (2-24)). Uma vez conhecida as cópulas condicionais $C_{d|d-1,\dots,1}$, podemos então obter as respectivas funções inversas em termos das funções inversas das *h-functions* (h^{-1}).

Ilustraremos agora este procedimento para o caso tridimensional. Para esta tarefa seremos auxiliados pelo gráfico de uma R-vine (Figura 3.5) de modo a esclarecer como as distribuições condicionais são obtidas através da aplicação sequencial das *h-functions*.

Neste exemplo, consideraremos a distribuição multivariada $F_{1,2,3}$ de um vetor aleatório $\mathbf{X} = (X_1, X_2, X_3) \in \mathbb{R}^3$, com distribuições marginais dadas por F_1, F_2, F_3 . Nosso objetivo é simular observações (x_1, x_2, x_3) a partir da cópula $C_{1,2,3}$ correspondente a $F_{1,2,3}$.

A Figura 3.5 ilustra essa distribuição. Os vértices na árvore T_1 representam as variáveis transformadas correspondentes a x_1, x_2 e x_3 . Nas arestas, explicitamos quais variáveis são modeladas pelas pair-cópulas em cada momento. Na árvore T_1 existem as cópulas C_{12} e C_{23} calculadas a partir das variáveis transformadas $F_1(x_1)$ e $F_2(x_2)$; $F_2(x_2)$ e $F_3(x_3)$, respectivamente.

Para a cópula seguinte, $C_{13|2}$, precisamos calcular $C_{1|2}$ e $C_{3|2}$. Estas são obtidas através das *h-functions*. Nesta seção, optamos por numerar os índices e explicitar os parâmetros das cópulas bivariadas onde as *h-functions* foram calculadas. Por exemplo, $h_{3|2}(u_3|u_2; \Theta_{23})$ indica a cópula condicional de u_3 dado u_2 obtida a partir da cópula C_{23} parametrizada pelo conjunto de parâmetros Θ_{23} .

Da última árvore podemos obter a cópula $C_{3|2,1}$ ao derivarmos a cópula $C_{1,3|2}(C_{1|2}, C_{3|2})$ com relação à $C_{1|2}$. Em termos de *h-functions*, temos

$$C_{3|2,1} = h_{3|2,1}[h_{3|2}(u_3|u_2; \Theta_{23})|h_{1|2}(u_1|u_2; \Theta_{12}); \Theta_{13|2}]. \quad (3-27)$$

A cópula condicional inversa $C_{3|2,1}^{-1}$ pode ser encontrada a partir da equação (3-27) e das funções inversas das *h-functions* (h^{-1}), ou seja,

$$C_{3|2,1}^{-1} = h_{3|2,1}^{-1}[h_{3|2,1}^{-1}(w_3|h_{1|2}(u_1|u_2; \Theta_{12}); \Theta_{13|2})|u_2; \Theta_{23}] \quad (3-28)$$

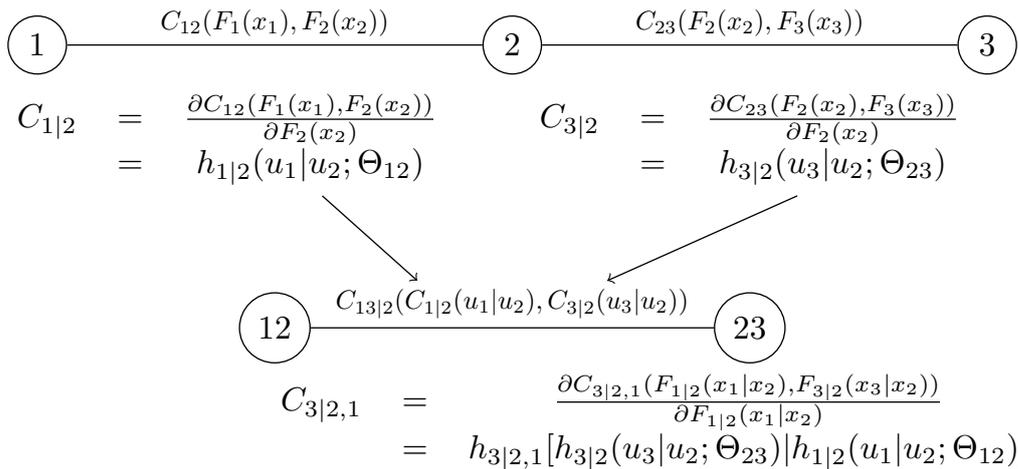


Figura 3.5: R-vine com 3 variáveis e *h-functions*.

Em termos de *h-functions*, o procedimento de simulação é:

1^o passo: Simule $w_i \stackrel{\text{i.i.d.}}{\sim} U(0, 1), i = 1, 2, 3$.

2^o passo: Faça:

$$u_1 = w_1,$$

$$u_2 = h_{2|1}^{-1}(w_2|u_1; \Theta_{12}),$$

$$u_3 = h_{3|2}^{-1}[h_{3|2,1}^{-1}(w_3|h_{1|2}(u_1|u_2; \Theta_{12}); \Theta_{13|2})|u_2; \Theta_{23}].$$

Vale destacar que as observações (u_1, u_2, u_3) são dados cópulas $(U(0, 1))$. Para transformá-los para a escala original basta utilizar as funções quantis (inversas) das correspondentes distribuições marginais, i.e., $x_1 = F_1^{-1}(u_1)$, $x_2 = F_2^{-1}(u_2)$ e $x_3 = F_3^{-1}(u_3)$.

4

R-vine espaciais para simulação de cenários hidrológicos

4.1

Motivação

Além da dependência temporal presente nas séries de vazões dos reservatórios, existe também uma dependência espacial entre as diversas usinas que compõem o Sistema Elétrico Brasileiro (SEB). Deste modo, um modelo de simulação deve ser capaz de modelar ambos os tipos de dependência. Particularmente no SEB onde o número de usinas é elevado, tais modelos também devem ser capazes de incorporar a alta dimensionalidade do problema.

Neste contexto, modelos multivariados para séries temporais podem ser utilizados. [2,5]. Estes são normalmente baseados nos modelos autorregressivos médias móveis (ARMA), cujas desvantagens foram elencadas na introdução.

Hao & Singh [19] citam alguns métodos alternativos tais como o método de desagregação. Inicialmente simula-se séries agregadas (séries anuais, por exemplo) e, posteriormente, decompõem-se em escalas menores (mensais ou semanais, por exemplo). Sharma & O'Neil [8] listam algumas das desvantagens inerentes à ambos os métodos. Dentre elas destaca-se a limitação em representar dependências não lineares e distribuições de probabilidade com formas não usuais. Alguns métodos não paramétricos são utilizados para simulação de cenários multivariados de vazões e, de certa forma, não possuem as desvantagens supracitadas. Contudo, tais métodos são ineficientes em altas dimensões, [22].

Recentemente, as cópulas começaram a ser empregadas para a simulação de cenários hidrológicos. No âmbito da simulação de cenários univariados onde se considera somente a dependência temporal, pode-se citar as referências [9, 11, 16–18, 21].

No âmbito de modelos para geração simultânea de cenários de vazões, onde adicionalmente a dependência temporal, a dependência espacial é incorporada ao modelo, pode-se citar o trabalho de Hao & Sing [19]. Neste, os autores propuseram uma metodologia utilizando entropia para obtenção da cópula (*maximum entropy copula method*) entre diferentes usinas. Tanto a dependência temporal quanto a relação entres as usinas foram modeladas via

cópula. O modelo proposto foi avaliado num conjunto de apenas 3 usinas. A análise dos cenários indicou a adequação do modelo proposto.

Chen et al. [22] desenvolveram um método utilizando cópulas para simulação de vazões de rios. A metodologia considera apenas duas usinas simultaneamente. Ela utiliza a abordagem sugerido por Lee & Salas [17] para a simulação de cenários em uma única localidade (região inicial) e condiciona os valores simulados de outro rio (região de interesse) a estes. Dessa forma, a simulação para qualquer localidade (região de interesse) é realizada através de uma distribuição multivariada composta pelos valores atuais e passados da mesma e os valores presentes da região inicial. Dois conjuntos de rios foram analisados. O primeiro era composto pela vazão mensal de 3 localidades enquanto que o segundo era composto por 6 vazões diárias observadas em diferentes rios. Os resultados indicam que o modelo proposto foi capaz de replicar as características observadas nas séries de vazões originais.

Isto posto, o objetivo deste capítulo é a construção de um modelo de cópulas para a simulação de cenários hidrológicos multivariados. O Capítulo 2 apresentou a cópula vine como uma alternativa bastante eficiente para a construção de modelos em alta dimensão. Embora estes sejam capazes de modelar as mais diversas estruturas de dependência, o número de parâmetros cresce quadraticamente com a dimensão. Assim, uma redução em sua complexidade é desejada. Neste contexto, inserem-se os modelos de cópulas vine espaciais (*Spatial R-vine models*) introduzidos por Erhardt et al. [23]. A principal ideia é utilizar variáveis exógenas (informações geográficas) para reparametrizar a cópula vine de modo a reduzir o número de parâmetros.

O modelo desenvolvido neste capítulo utiliza a distância entre as usinas, as informações sobre precipitações próximas as usinas e a estrutura das bacias para reparametrizar a cópula vine. Ademais, devido à forte sazonalidade presente nas séries de vazões mensais, propomos uma extensão do modelo original para acomodar variações periódicas da dependência entre as usinas.

A contribuição deste capítulo ocorre de diversas formas. O modelo proposto é uma versão aumentada do modelo original ([23]) para o caso onde a dependência espacial não é constante. Assim, este pode ser interpretado como um modelo de cópulas vine espaciais com parâmetros variando periodicamente. O modelo estimado é flexível ao ponto de incorporar inúmeras estruturas de dependência. Sinalizamos um conjunto de variáveis que podem ser aplicadas para aproximar a dependência espacial entre diversas usinas. A reparametrização realizada foi capaz de reduzir em até 96% o número de parâmetros estimados na árvore T_1 . Por fim, salienta-se a aptidão do modelo em simular cenários para um elevado número de usinas. A análise dos cenários indicou que

o modelo é capaz de simular cenários preservando características históricas.

4.2

Cópulas vine espaciais

O modelo de cópula vine espacial (*Spatial R-Vine model*) é uma extensão dos modelos de cópulas vine desenvolvidos com o propósito de modelar a dependência espacial. Esta classe foi inicialmente proposta por Erhardt et al. [23] e, consiste em explorar a relação entre variáveis espaciais (*variáveis exógenas/geográficas*) e os parâmetros das cópulas bivariadas. Esta relação, se corretamente estabelecida, permite uma drástica redução no número de parâmetros.

Antes de expormos esta sofisticada classe de modelos, algumas observações acerca da notação empregada neste capítulo são elucidadas. Uma aresta arbitrária e será denotada como $i(e), j(e); D_e$, onde $C_e = \{i(e), j(e)\}$ é o conjunto condicionado e D_e é o conjunto condicionante. Estes conjuntos foram definidos no Capítulo 3. A cópula associada à esta aresta é $C_{i(e),j(e);D_e}$, cujos parâmetros são $\theta_{i(e),j(e);D_e}$ (primeiro parâmetro da cópula) e $\nu_{i(e),j(e);D_e}$ (segundo parâmetro da cópula, quando existente).¹ O correspondente τ de Kendall é $\tau_{i(e),j(e);D_e}$. Caso este esteja associado à alguma aresta da árvore T_1 , onde o conjunto condicionante é vazio, temos $\tau_{i(e),j(e)}$. Oportunamente, adicionaremos o índice ℓ para explicitar à qual árvore T_ℓ , $\ell = 1, \dots, d - 1$ estaremos nos referindo.

O τ de Kendall pode ser expresso em termos de cópulas. Assim, existe uma função que conecta os parâmetros $\theta_{i(e),j(e);D_e}$ (dado uma família paramétrica bivariada - $b_{i(e),j(e);D_e}$) ao τ de Kendall tal que

$$\tau_{i(e),j(e);D_e} = g_\tau(\theta_{i(e),j(e);D_e}; b_{i(e),j(e);D_e}). \quad (4-1)$$

Esta relação é obtida ao se resolver a equação (2-29) e, é conhecida para as principais cópulas bivariadas. A Seção 2.3 lista algumas dessas funções.

Introduziremos o modelo de vines espaciais através da ilustração de um caso hipotético. O propósito é facilitar seu entendimento através de um exemplo utópico.

Considere dados de $d = 40$ séries de temperatura média mensal e que cada série possua T observações. Assuma ainda que, estes dados foram devidamente tratados e transformados em dados cópulas, resultando em dados $\mathbf{u}^1, \dots, \mathbf{u}^{40}$ com $\mathbf{u}^1 = (u_1^1, \dots, u_T^1)^\top, \dots, \mathbf{u}^{40} = (u_1^{40}, \dots, u_T^{40})^\top$. Estas séries de temperatura foram mensuradas em estações meteorológicas dispersas ao longo de determinada área (região, país, etc.). É razoável supor que a associação entre

¹Consideraremos somente cópulas bivariadas com um ou dois parâmetros.

elas seja inversamente proporcional à distância. Quanto maior a distância, menor é a associação entre as temperaturas. Considere ainda que a distância d_{ij} , $1 \leq i < j \leq 40$, entre todos os pares de estações possa ser calculada. A principal ideia dos modelos de cópulas espaciais é exatamente explorar essa relação entre a distância (ou qualquer outra variável) e a associação entre pares de localidades (representada pelo τ de Kendall) para reduzir o número de parâmetros.

Um simples gráfico, como apresentado na Figura 4.1, ajuda a elucidar a relação do τ de Kendall $\tau_{i,j}$ com a distância $d_{i,j}$, $1 \leq i < j \leq 40$. Neste exemplo hipotético, existe uma relação linear negativa entre o logaritmo da distância e a transformada de Fisher do τ de Kendall ($g_z(\hat{\tau}_{ij})$) estimado a partir dos dados.

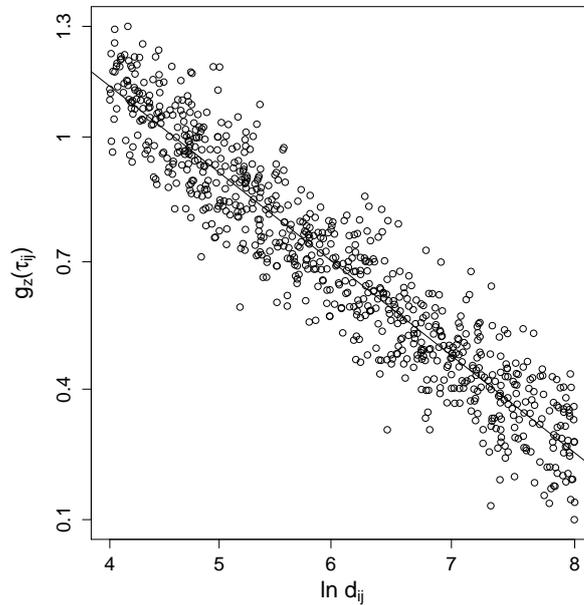


Figura 4.1: Exemplo de associação entre $\ln(d_{ij})$ e $g_z(\hat{\tau}_{ij})$.

Dessa forma, pode-se considerar uma relação linear entre a distância e o τ de Kendall, como

$$g_z(\hat{\tau}_{ij}) = \alpha_0 + \alpha_1 \ln(d_{ij}). \tag{4-2}$$

Em outras palavras, dado uma distância entre duas estações é possível inferirmos o grau de associação entre ambas. Obviamente, esta relação é aproximada.

Denotando g_τ^{-1} como a função inversa da equação (4-1) e utilizando a transformada inversa de Fisher g_z^{-1} introduzida na Seção 2.4, podemos agora expressar o parâmetro da cópula $C_{i(e)j(e);D_e}$ em função de $\ln d_{i(e),j(e);D_e}$ (distância entre estas localidades) como

$$\theta_{i(e)j(e);D_e} := g_\tau^{-1} \left(g_z^{-1} \{ \alpha_0 + \alpha_1 \ln d_{i(e),j(e);D_e} \}; b_{i(e),j(e);D_e} \right). \tag{4-3}$$

Esta relação (equação (4-3)) desempenha papel crucial no modelo espacial, uma vez que estabelece o link entre as variáveis espaciais (no exemplo, distância) e o parâmetro das cópulas bivariadas. Neste exemplo, estamos considerando que a distância descreve razoavelmente tanto $\tau_{i(e),j(e)}$ quanto $\tau_{i(e),j(e),D_e}$. Dessa forma, cada parâmetro de cada aresta passa a ser reescrito em termos de uma equação em comum.

A Figura 4.2 ilustra este fato na árvore T_1 . Cada vértice corresponde a uma estação meteorológica/série temporal.

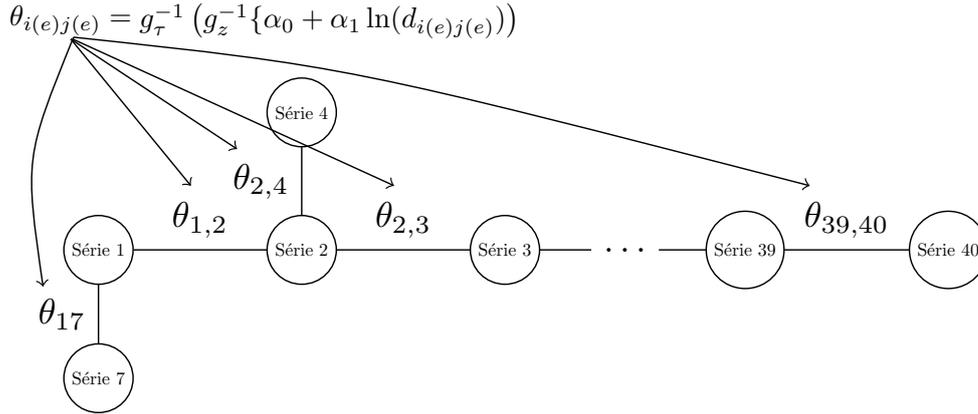


Figura 4.2: Exemplo de reparametrização para a árvore T_1 .

A equação (4-3) será intitulada equação espacial. Ela rege toda a estrutura de dependência entre as variáveis e , ao mesmo tempo, reduz o número de parâmetros.

Em Erhardt et al. [23], os autores estabeleceram uma equação espacial para cada árvore T_ℓ , $\ell = 1, \dots, L_{max}$. Formalmente, a reparametrização é definida por

$$\theta_{i(e)j(e);D_e}^\ell := g_\tau^{-1}\left(g_z^{-1}\{h_\ell(e|\beta_\ell)\}; b_{i(e)j(e);D_e}\right), \quad (4-4)$$

onde a função $h_\ell(e|\beta_\ell)$ controla a inserção das equações espaciais no modelo. Desta maneira, cada árvore T_ℓ da cópula vine possui sua própria reparametrização. [23].

Considerando observações de V variáveis espaciais e que estas são capazes de descrever tanto a dependência condicional quanto a dependência incondicional, a função $h_\ell(e|\beta_\ell)$ assume a seguinte forma:

$$h_\ell(e|\beta_\ell) = \beta_{\ell,0} + \sum_{v=1}^V \beta_{\ell,v}(x_{i(e)j(e)}^v), \quad e \in E_\ell, \quad \ell = 1, \dots, L_{max} \quad (4-5)$$

onde $\beta_\ell = (\beta_{\ell,0}, \dots, \beta_{\ell,V})^\top \in \mathbb{R}^{V+1}$. Todos os parâmetros referentes ao primeiro parâmetro cópula podem ser organizados como $\beta_\theta^{esp} = (\beta_1^\top, \dots, \beta_{L_{max}}^\top) \in \mathbb{R}^{(V+1) \cdot L_{max}}$ onde L_{max} denota o número de árvores consideradas.

A reparametrização apresentada até este momento é válida apenas para o primeiro parâmetro das cópulas bivariadas. No caso específico da cópula t , onde adicionalmente existe um segundo parâmetro, Erhardt et al. [23] identificaram uma tendência quadrática ao longo das árvores e propuseram a seguinte reparametrização:

$$\nu_{i(e)j(e);D_e}^\ell := \exp\{\beta_0^\nu + \beta_1^\nu \cdot \ell + \beta_2^\nu \cdot \ell^2\}. \quad (4-6)$$

Desta forma, os parâmetros que governam os graus de liberdade das cópulas t bivariadas são resumidos à $\beta_\nu^{esp} := (\beta_0^\nu, \beta_1^\nu, \beta_2^\nu)$.

No trabalho original, Erhardt et al. [23] modelaram a dependência espacial entre séries temporais de temperatura na Alemanha entre os anos de 2010 e 2012. A abordagem adotada acompanhou a tradicional para o caso onde os dados possuem dependência serial. Inicialmente aplicou-se modelos de séries temporais para cada série individualmente e, posteriormente, estimou-se a dependência entre as séries através de cópulas. Como variáveis espaciais, empregaram a distância e a diferença de altitude entre as estações meteorológicas. Ademais, os autores conseguiram designar equações espaciais para todas as árvores de sua cópula vine espacial.

4.2.1 Estimação do modelo de cópula vine espacial

A estimação do modelo de cópula vine espacial pode ser realizada através da maximização da função de verossimilhança. Esta é obtida ao se inserir as reparametrizações definidas nas equações (4-4) e (4-6) na função de verossimilhança apresentada no Capítulo 3. Dessa maneira, têm-se

$$\mathcal{L}(\beta_\theta^{esp}, \beta_\nu^{esp} | \mathbf{u}) = \prod_{t=1}^T \prod_{\ell=1}^{L_{max}} \prod_{e \in E_\ell} c_{i(e),j(e);D_e} \left(\tilde{u}_t^{i(e)}, \tilde{u}_t^{j(e)}; \theta_{i(e),j(e);D_e}^\ell, \nu_{i(e),j(e);D_e}^\ell \right), \quad (4-7)$$

onde $\tilde{u}_t^{i(e)} := C_{i(e)|D_e}(u_t^{i(e)} | \mathbf{u}_t^{D_e})$ e $\tilde{u}_t^{j(e)} := C_{j(e)|D_e}(u_t^{j(e)} | \mathbf{u}_t^{D_e})$ com $\mathbf{u}_t^{D_e} := \{u_t^s : s \in D_e\}$. Além disto, $\mathbf{u} = (\mathbf{u}^1, \dots, \mathbf{u}^d)$ com $\mathbf{u}^s = (u_1^s, \dots, u_T^s)^\top$, $s = 1, \dots, d$. [44].

Assim como no caso das cópulas vine, a estimação dos parâmetros pode ser realizada através da maximização por inteira da verossimilhança (*full maximum likelihood estimation*) - onde todos os parâmetros são estimados simultaneamente - ou através da estimação sequencial (*sequential estimation*) - onde cada pair-cópula é estimada individualmente.

Ambas abordagens podem ser empregadas nas cópulas vine espaciais. Entretanto, diferentemente do caso clássico, a estimação sequencial é realizada ao longo das árvores. Em outras palavras, para cada árvore, uma otimização é realizada independentemente do número de variáveis. Comparando os dois

modelos, uma cópula vine com 50 variáveis, por exemplo, teria 49 procedimentos de estimação (um para cada cópula bivariada) somente na T_1 , enquanto que no modelo espacial, apenas uma otimização seria realizada.

4.3

Modelo periódico de cópula vine espacial

A extensão proposta ao modelo de cópula vine espacial é expressa neste momento. Esta é entendida como uma versão do modelo original de [23] adaptado para os casos onde existe uma dependência espacial periódica entre as variáveis.

Na extensão proposta, diferentemente do apresentado na seção 4.2, consideraremos que o parâmetro $\theta_{i(e)j(e);D_e}^\ell$ das cópulas bivariadas variam de acordo com período. No caso específico desta tese, o parâmetro $\theta_{i(e)j(e);D_e}^\ell$ irá ser uma função do mês. Desta forma, o primeiro parâmetro da cópula passa a ser $\theta_{i(e)j(e);D_e}^{\ell,m}$, onde $m = [jan, \dots, dez]$.

Para acomodar esta variação periódica, adicionou-se 11 variáveis *dummy* S_{qm} , $p = 1, \dots, 11$ que assume 1 se $q = m$ ou 0 caso $q \neq m$. Desta maneira, a equação (4-4) no modelo periódico assume a seguinte forma:

$$\theta_{i(e)j(e);D_e}^{\ell,m} := g_\tau^{-1} \left(g_z^{-1} \{h_\ell(e|\boldsymbol{\beta}_\ell)\}; b_{i(e)j(e);D_e}^m \right) \quad (4-8)$$

onde a função $h_\ell(e|\boldsymbol{\beta}_\ell)$ é expressa por

$$h_\ell(e|\boldsymbol{\beta}_\ell) = \beta_{\ell,0} + \sum_{v=1}^V \beta_{\ell,v} x_{i(e)j(e);D_e}^v + \sum_{q=1}^{11} \beta_{\ell,q+V} S_{qm}, \quad e \in \varepsilon_\ell, \ell = 1, \dots, L_{max}. \quad (4-9)$$

Uma vez definida a reparametrização para o primeiro parâmetro da cópula, abordaremos agora o segundo parâmetro (graus de liberdade) da cópula t. Erhardt et al. [23] identificaram uma tendência ao longo das árvores e aproximaram os graus de liberdade através de uma função quadrática.

De maneira diferente, a reparametrização sugerida no presente trabalho ocorre dentro das árvores. Reduziremos todos os parâmetros relativos ao grau de liberdade ($\nu_{i(e)j(e);D_e}^\ell$) de cada árvore à apenas 12 parâmetros, ou seja, um parâmetro ν para cada mês. Por exemplo caso a árvore T_1 possua 30 cópulas t em cada mês, equivalente a $30 \times 12 = 360$ parâmetros ν , de acordo com a parametrização proposta, estes seriam reduzidos à apenas 12. Em outras palavras, cada mês de determinada árvore possui apenas um parâmetro relativo aos graus de liberdade, ou seja,

$$\nu_{i(e)j(e);D_e}^{\ell,m} := \nu^{\ell,m}. \quad (4-10)$$

É de conhecimento comum que a cópula t tende para uma cópula Gaus-

siana conforme os graus de liberdade aumentam. Apenas graus de liberdade relativamente baixos geram dependência de caudas relevantes. Com esta reparametrização, espera-se que os novos parâmetros sejam capazes de estimar baixos graus de liberdade. Obvio que neste caso, estamos assumindo a existência desta característica no conjunto de dados analisados. Caso contrário, a cópula Gaussiana é preferível à cópula t pois possui apenas um parâmetro.

4.4

Modelo periódico de cópulas vine aplicados a séries de vazão

A extensão proposta foi aplicada a um subconjunto de 39 usinas hidrelétricas que compõem o SEB. As séries mensais de vazões naturais foram obtidas junto ao Operador Nacional do Sistema Elétrico (ONS) e estão disponíveis em seu site. [47]. Estas séries tem início em janeiro de 1931 e término em dezembro de 2012 totalizando 984 observações. Iremos denotar y_t^s ($t = 1, \dots, 984$) como a vazão média mensal observada na usina hidrelétrica s , ($s = 1, \dots, 39$) no instante de tempo t .

O termo vazão natural é definido como a vazão que ocorreria em determinada seção do rio caso não existissem ações antrópicas, isto é, efeitos da operação de usinas à montante, evaporações em reservatórios e usos consuntivos tais como irrigação, abastecimento etc. As séries de vazões naturais são reconstituídas através das vazões observadas nos locais e informações relativas às ações antrópicas. [47].

Adicionalmente, um conjunto de 39 séries temporais de precipitação de janeiro de 2010 até dezembro de 2013 fornecidas pelo Instituto Nacional de Meteorologia, [48], e pela Agência Nacional de Águas, [49], foram utilizadas para construir uma variável espacial.

4.4.1

Transformando variáveis originais em uniformes

Antes de transformarmos as variáveis originais em uniformes, é necessário filtrar toda a estrutura de dependência existente nas séries temporais. Para esta tarefa empregamos o modelo PAR (p) apresentado na Seção 2.5. A formulação apresentada supõe a distribuição Gaussiana padrão para os resíduos padronizados. Todavia, após uma análise dos resíduos observou-se que grande parte dos mesmos possuíam assimetria e excesso de curtose. Neste caso, decidimos adotar a distribuição skew- t como distribuição para os resíduos padronizados, ou seja, $\epsilon_t^s \sim skew-t(\xi^s, \omega^s, \zeta^s, \nu^s)$. Sua função densidade de probabilidade é dada por

$$f_{skew-t}(x|\xi, \omega, \zeta, \nu) = \frac{2}{\omega} t_\nu(\tilde{x}) T_{\nu+1} \left\{ \zeta \tilde{x} \left(\frac{\nu+1}{\nu+\tilde{x}^2} \right)^{\frac{1}{2}} \right\} \quad (4-11)$$

onde $\tilde{x} := (x - \xi)/\omega$, t_ν denota a densidade de uma t-Student univariada com ν graus de liberdade e $T_{\nu+1}$ é função de distribuição acumulada de uma t-Student com $\nu+1$ graus de liberdade. Na equação (4-11), $\xi \in \mathfrak{R}$ é o parâmetro de posição, $\omega > 0$ é o parâmetro de escala, $\zeta \in \mathfrak{R}$ é o parâmetro que controla a assimetria e $\nu > 0$ denota os graus de liberdade.

Para a estimação dos modelos marginais, um procedimento em dois estágios foi adotado. Inicialmente, estimou-se os parâmetros do modelo PAR (p) e posteriormente, os parâmetros das distribuições skew-t.

Logo após, calculando

$$u_t^s := F_{skew-t}(\hat{\epsilon}_t^s | \hat{\xi}^s, \hat{\omega}^s, \hat{\alpha}^s, \hat{\nu}^s), \quad t = 1, \dots, 984, s = 1, \dots, 39, \quad (4-12)$$

obtemos os dados cópula a partir dos resíduos do modelo PAR (p). Neste caso, F_{skew-t} é a função de distribuição acumulada skew-t.

Após esta transformação, é necessário verificar se as variáveis transformadas (dados cópula) são realmente uniformes ou não. Para esta tarefa, empregou-se o teste de Anderson-Darling (teste A-D). [50,51]. A hipótese nula diz que os dados transformados seguem uma distribuição uniforme padrão. Na Figura 4.3, cada barra representa o p-valor do teste para cada usina hidrelétrica. A linha preta contínua representa o nível de significância de 5%.

Como pode ser observado, não é possível rejeitar a hipótese nula a um nível de significância de 0,05. Desta maneira, pode-se assumir que os dados transformados seguem uma distribuição uniforme padrão.

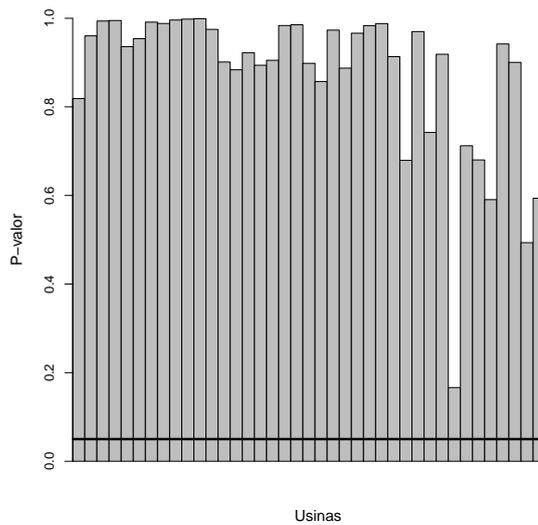


Figura 4.3: Teste de Anderson-Darling para cada uma das 39 usinas.

4.4.2

Definição da estrutura e seleção das pair-cópuas

Para mensurar a variação sazonal entre as usinas hidrelétricas, reorganizamos o conjunto de dados cópula. As observações $\mathbf{u}^s = (u_1^s, \dots, u_{984}^s)$ de cada localidade s são ordenadas de acordo com os meses. Assim, estas são reorganizadas como $\mathbf{u}^s = (\mathbf{u}_{jan}^s, \dots, \mathbf{u}_{dez}^s)$ onde $\mathbf{u}_{jan}^s = (u_{1,jan}^s, \dots, u_{82,jan}^s)^\top, \dots, \mathbf{u}_{dez}^s = (u_{1,dez}^s, \dots, u_{82,dez}^s)^\top$. Por exemplo, \mathbf{u}_{jan}^s contém todas as observações de janeiro da usina s , \mathbf{u}_{fev}^s contém todas as observações de fevereiro e assim sucessivamente. Ademais, $u_{1,jan}^s$ é o primeiro janeiro observado (correspondente à janeiro de 1931), $u_{2,jan}^s$ expressa janeiro do segundo ano histórico (janeiro de 1932) enquanto que $u_{82,jan}^s$ corresponde à observação de janeiro do último ano (janeiro de 2012). Isto posto, podemos calcular o τ de Kendall ($\tau_{ij,m}^{hidro}$) que representa a associação entre as usinas i e j no mês m .

A fim de melhor entender a estrutura de dependência, neste primeiro momento, estimamos 12 cópulas vine (uma para cada mês) por meio do algoritmo proposto por Dißmann et al. [45]. Como peso para as arestas, utilizou-se o τ de Kendall. Este algoritmo está disponível em [52]. As cópulas bivariadas foram selecionadas via BIC. O conjunto de cópulas disponíveis era composto pela cópulas Gaussianas, t-Student, Frank, Gumbel, Clayton e versões rotacionadas destas últimas duas.

Com este modelo inicial, por meio da Figura 4.4 é possível realizar uma análise exploratória do τ de Kendall ao longo das árvores. Vale lembrar que, nesta figura, os coeficientes de Kendall da árvore T_1 são estimados diretamente dos dados enquanto que os das árvores subsequentes são estimados através dos pseudo-dados transformados.

Como pode ser observado, as maiores associações ocorrem na árvore T_1 . Logo a partir da árvore T_2 , o número de associações relevantes diminui sendo em sua grande maioria residual e próximo de zero. Isto sugere a utilização de cópulas de independência em alguns pares de variáveis.

Para determinar em quais pares a cópula de independência deve ser utilizada, estimamos novamente as cópulas vine em conjunto com o teste de independência para cópulas bivariadas (veja [53]) a um nível de significância de 0,05.

A Tabela 4.1 apresenta o número de cópulas bivariadas selecionadas e a log-verossimilhança por árvore até a T_{10} , considerando as 12 cópulas vines estimadas. Como pode ser observado, existe uma predominância da cópula de independência. A partir da T_6 , estas representam quase 90% das cópulas. Por estas razões, truncou-se o modelo na árvore T_6 . O conceito de cópulas truncadas foi introduzido por Brechmann et al. [54] e consiste em utilizar

cópuas de independência para todas as arestas a partir de determinada árvore T_ℓ .

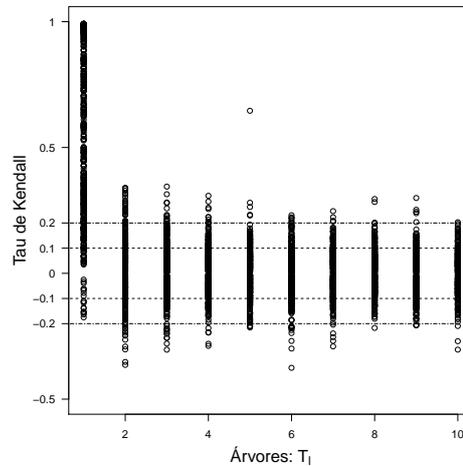


Figura 4.4: τ de Kendall por árvore (T_ℓ).

Esta análise indica que as maiores dependências são captadas nas primeiras árvores. Destaque para a T_1 que corresponde à uma grande parcela da log-verossimilhança. A cópula Gumbel aparece com maior frequência nesta árvore, seguida pela cópula Gaussiana e cópula t. Por sua vez, a cópula t aparece quase que exclusivamente na árvore T_1 . Vale destacar também o número de cópula Arquimedianas cuja presença é um forte indício de estruturas assimétricas de dependência entre os pares de usinas.

Tabela 4.1: Famílias de cópulas selecionadas para T_ℓ , $\ell = 1, \dots, 10$.

	Gaus.	cóp. t	Clayton	Gumbel	Frank	Indep.	% Indep.	LogLik
T_1	83	116	32	159	6	60	13%	40808
T_2	24	4	30	45	6	335	75%	603
T_3	24	0	15	21	4	368	85%	293
T_4	15	0	15	14	3	373	88%	197
T_5	13	2	12	17	6	358	87%	222
T_6	13	1	16	11	4	351	88%	188
T_7	19	1	12	7	4	341	88%	144
T_8	11	2	6	5	2	346	93%	83
T_9	11	0	10	7	4	328	91%	120
T_{10}	11	0	8	8	5	316	90%	107

Com relação a cópula t a Figura 4.5 apresenta os graus de liberdade por mês somente das cópulas t presentes na T_1 . Em geral, os graus de liberdade estimados apresentam valores baixos, o que implica em uma alta dependência de cauda entre os respectivos pares de usinas.

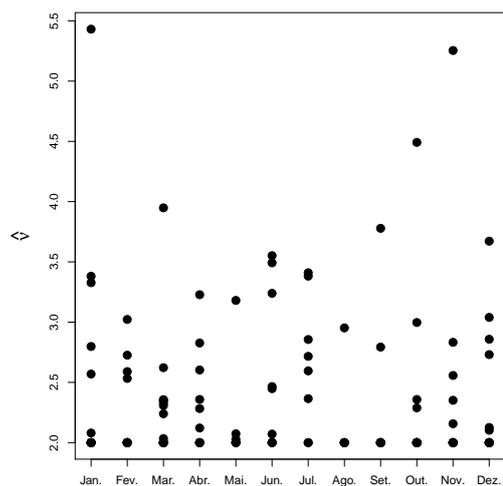


Figura 4.5: Graus de liberdade das cópulas t contidas na T_1 .

Por fim, apresentamos a estrutura da árvore selecionada para a árvore T_1 . Os nomes das respectivas usinas estão listados no Apêndice A. Figuras complementares que ilustram o curso d'água dos correspondentes rios podem ser observadas no Apêndice B.

PUC-Rio - Certificação Digital Nº 1121533/CA



Figura 4.6: Representação da T_1 sobre o mapa brasileiro.

Esta análise preliminar teve como objetivo determinar a estrutura das árvores e as cópulas bivariadas (pair-cópulas) associadas a cada aresta. Desta forma, o modelo espacial proposto empregará a estrutura obtida através do algoritmo de Dißmann et al. [45] e as famílias de cópulas apresentadas na Tabela 4.1 até a T_6 . Em vista do caráter periódico do modelo proposto, ressalta-se entretanto, que a estrutura das árvores é constante ao longo dos meses enquanto que as famílias de cópulas bivariadas variam de acordo com o período.

Uma vez realizada essa análise preliminar, o próximo passo é determinar quais as variáveis exógenas/espaciais que serão empregadas para a estimação do modelo.

4.4.3

Componentes espaciais

Esta seção apresenta as variáveis espaciais e como elas estão relacionadas com os parâmetros das pair-cópulas. Empregamos a distância entre as usinas hidrelétricas, a associação entre as séries de chuvas medidas próximas as usinas e uma variável que representa a estrutura dos rios. Além dessas três, inserimos *dummies* sazonais para representar a variação periódica entre as vazões.

4.4.3.1

Distância

O objetivo é explorar a relação existente entre distância das usinas e o τ de Kendall $\tau_{ij,m}^{hidro}$. A princípio, espera-se que usinas próximas uma da outras possuam um grau de associação maior do que usinas que estão muito afastadas.

Na Figura 4.7 é possível observar o gráfico de dispersão do logaritmo natural da distância contra a transformada de Fisher do τ de Kendall. Como esperado, existe uma associação negativa entre as variáveis, ou seja, de acordo com o aumento da distância entre as usinas, o grau de associação entre as mesmas diminui. Para este gráfico, considerou-se todos os $d(d-1)/2 = 39 * 38/2 = 741$ pares de usinas.

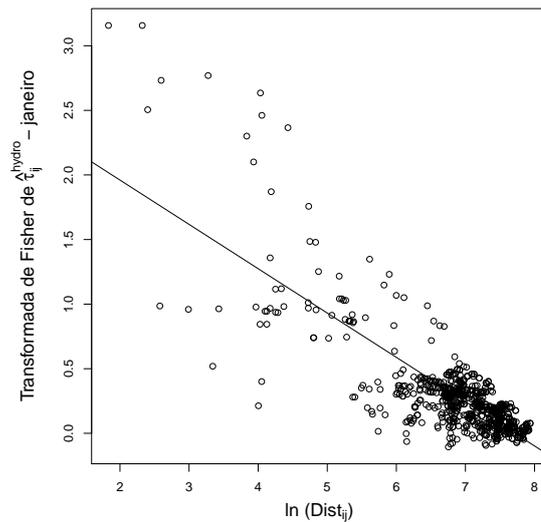


Figura 4.7: Associação entre $\log(\text{Dist})$ e o $g_z(\tau_{ij,m}^{\text{hidro}})$.

4.4.3.2

Precipitação

A ideia é representar a dependência entre as vazões das usinas através da dependência entre os regimes de chuvas, uma vez que as vazões são extremamente dependentes dos regimes de chuva. Em termos práticos, deseja-se utilizar as séries de precipitações para explicar a dependência entre as vazões. Infelizmente, os postos de medições das séries de precipitações (estações meteorológicas) não estão localizados exatamente nos reservatórios das usinas, entretanto, procurou-se utilizar os mais próximos possíveis. Além disso, algumas séries possuem um número baixo de dados incompletos. Estes foram substituídos pelas médias dos respectivos meses.

Desta maneira, para cada usina hidrelétrica uma série de precipitação foi associada. O objetivo é utilizar a dependência entre estas séries como uma *proxy* para as condições meteorológicas das regiões onde as usinas estão localizadas. A Figura 4.8 apresenta as 39 usinas hidrelétricas e as respectivas estações meteorológicas.

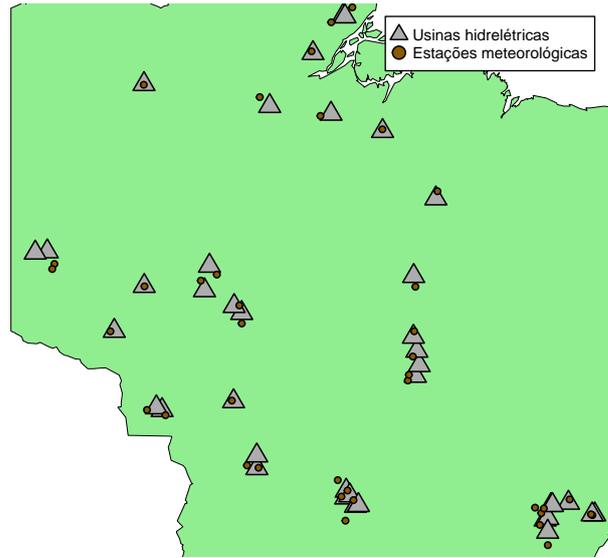


Figura 4.8: Usinas hidrelétricas e estações meteorológicas.

Para cada par de localidade (usina/estação meteorológica), o τ de Kendall (τ_{ij}^{chuva}) foi calculado. Previamente um modelo SARMA (*seasonal autoregressive moving average*) foi estimado para cada série e seus resíduos transformados em dados cópula.

A Figura 4.9 apresenta a relação entre a transformada de Fisher do $\tau_{ij,m}^{hidro}$ versus a transformada de Fisher do τ_{ij}^{chuva} para determinado mês. É possível observar uma associação positiva entre estas variáveis. Novamente, esta análise preliminar foi realizada considerando todos 741 pares de usinas.

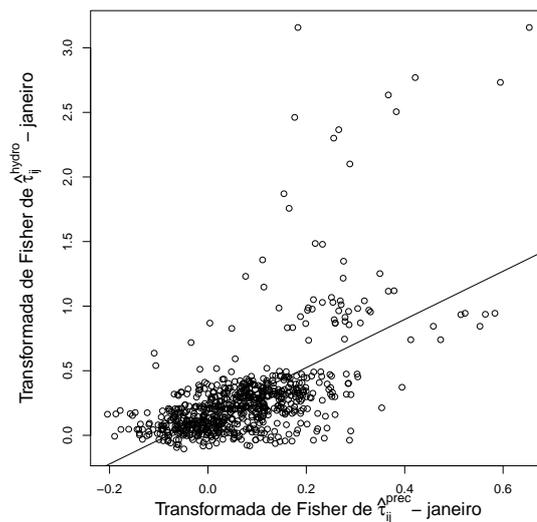


Figura 4.9: Associação entre $g_z(\tau_{ij,m}^{hidro})$ e $g_z(\tau_{ij}^{chuva})$.

4.4.3.3 Topografia dos rios

Algumas usinas hidrelétricas estão dispostas em um sistema de cascata. Isso significa que as vazões utilizadas para produzir energia em determinada planta serão reutilizadas para produzir energia em outras usinas a jusante. Uma simples inspeção nos dados indicou que as maiores dependências ocorrem nestes pares de usinas.

Assim sendo, uma variável dummy é inserida no modelo com o intuito de fornecer à equação espacial informações acerca da disposição das usinas ao longo do rio. Esta variável é definida da seguinte maneira:

$$D_{ij} = \begin{cases} 1, & \text{se as usinas } i \text{ e } j \text{ estão localizadas} \\ & \text{no mesmo rio sem outra usina } k \\ & \text{entre } i \text{ e } j. \\ 0, & \text{caso contrário.} \end{cases} \quad (4-13)$$

4.4.4 Definindo equação espacial para a T_1

As Figuras 4.10 e 4.11, apresentam as variáveis distância e precipitação para os 38 pares de usinas que compõem a árvore T_1 definida na seção 4.4.2. Como pode ser observado, a associação entre as variáveis é mais clara do que as apresentadas nas Figuras 4.7 e 4.9. Isto corrobora a seleção das mesmas para o modelo proposto.

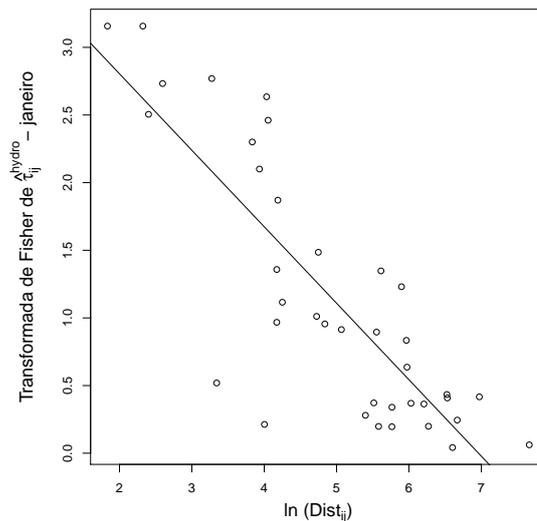


Figura 4.10: $\text{Log}(\text{Dist})$ e $g_z(\tau_{ij,jan}^{hidro})$ para os pares de usinas da árvore T_1 .

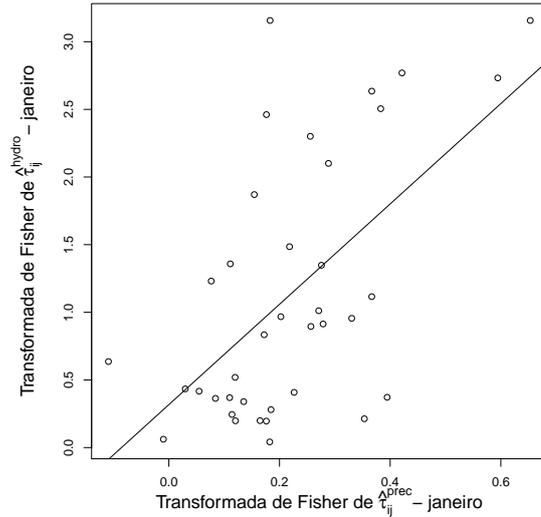


Figura 4.11: $G_z(\tau_{ij, jan}^{hidro})$ e $g_z(\tau_{ij}^{chuva})$ para os pares de usinas da árvore T_1 .

Assim sendo, a função $h_\ell(e|\beta_e)$ que controla a inserção das equações espaciais no modelo assume o seguinte formato:

$$h_1(e|\beta_1) = \beta_{1,0} + \beta_{1,1} \ln(Dist_{i(e)j(e)}) + \beta_{1,2} g_z(\hat{\tau}_{i(e)j(e)}^{prec}) + \beta_{1,3} D_{i(e)j(e)} + \sum_{q=1}^{11} \beta_{1,q+3} S_{qm} \quad (4-14)$$

onde $\beta_1 = (\beta_{1,0}, \dots, \beta_{1,14})^\top \in \mathbb{R}^{15}$.

Uma vez definida a equação que reparametriza a primeira árvore, o próximo passo é definir o conjunto de equações para as árvores subsequentes. Erhardt et al. [23] criaram novas variáveis com base nas variáveis espaciais originais. Assim, seguindo Erhardt et al. [23], para todas as arestas $e \in \varepsilon_\ell$ das árvores T_ℓ , $\ell > 2$, temos

$$\begin{aligned} \overline{Dist}_{i(e), D_e} &:= \frac{1}{\ell - 1} \sum_{k \in D_e} dist_{i(e), k}, & \overline{Dist}_{j(e), D_e} &:= \frac{1}{\ell - 1} \sum_{k \in D_e} dist_{j(e), k} \\ \overline{Prec}_{i(e), D_e} &:= \frac{1}{\ell - 1} \sum_{k \in D_e} \tau_{i(e), k}^{prec}, & \overline{Prec}_{j(e), D_e} &:= \frac{1}{\ell - 1} \sum_{k \in D_e} \tau_{j(e), k}^{prec} \end{aligned} \quad (4-15)$$

que podem ser utilizados como preditores adicionais às variáveis originais apresentadas anteriormente.

Infelizmente, não foi possível identificar qualquer variável explicativa para as árvores subsequentes. Testamos as variáveis definidas na equação (4-15) e as selecionadas para a T_1 . As Figuras 4.12, 4.13 e 4.14 apresentam gráficos de dispersão de algumas variáveis testadas até T_4 .

As conclusões obtidas ao se analisar tais figuras podem ser generalizadas para todas as variáveis e meses considerados nesta tese. Não é possível

identificar qualquer tipo de dependência entre as variáveis analisadas. Além disso, destaca-se a baixa magnitude dos τ de Kendall presentes a partir da T_2 . Estes resultados indicam que uma reparametrização para todas as árvores pode não ser adequada.

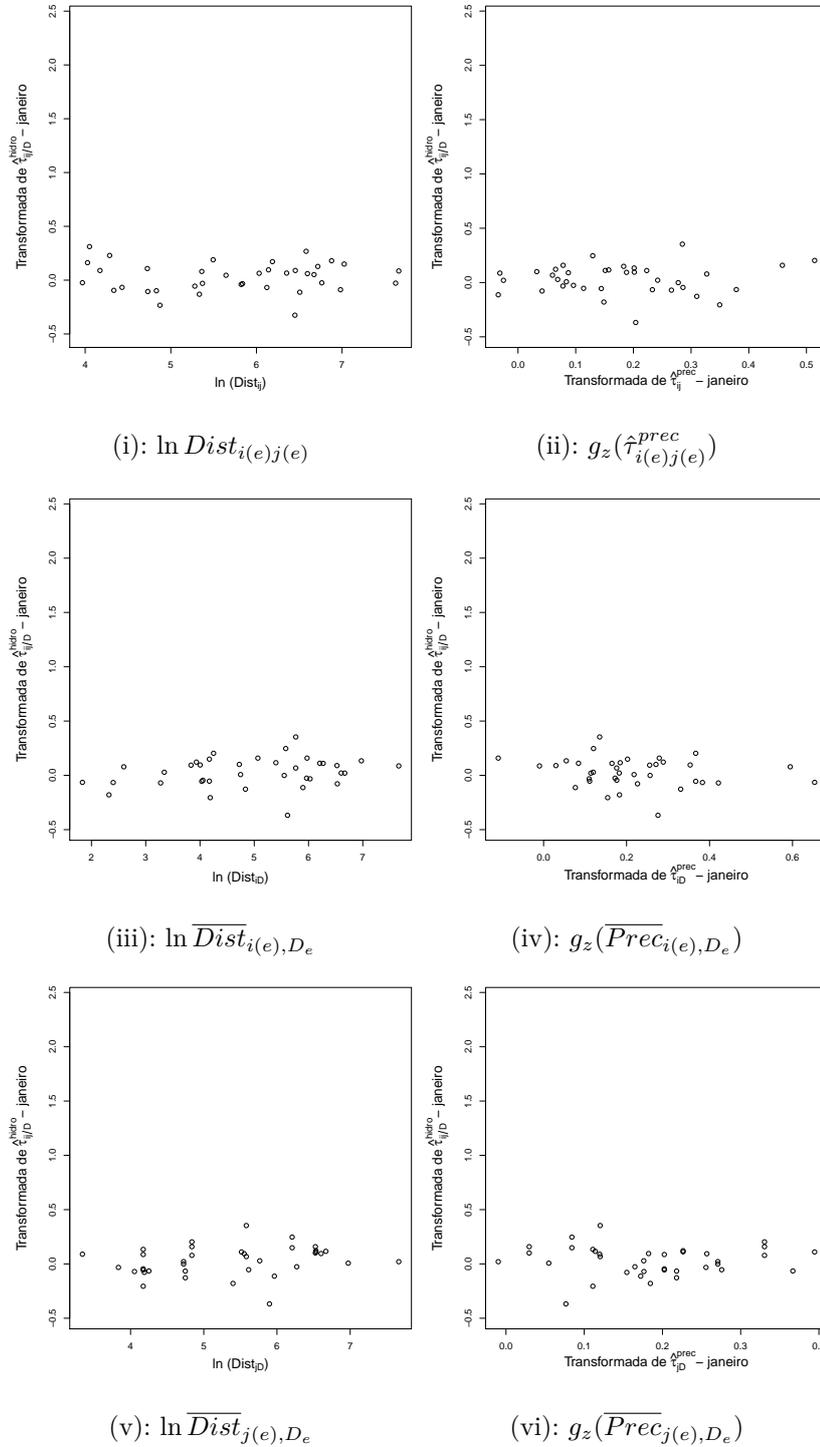


Figura 4.12: Análise das variáveis espaciais para T_2 .

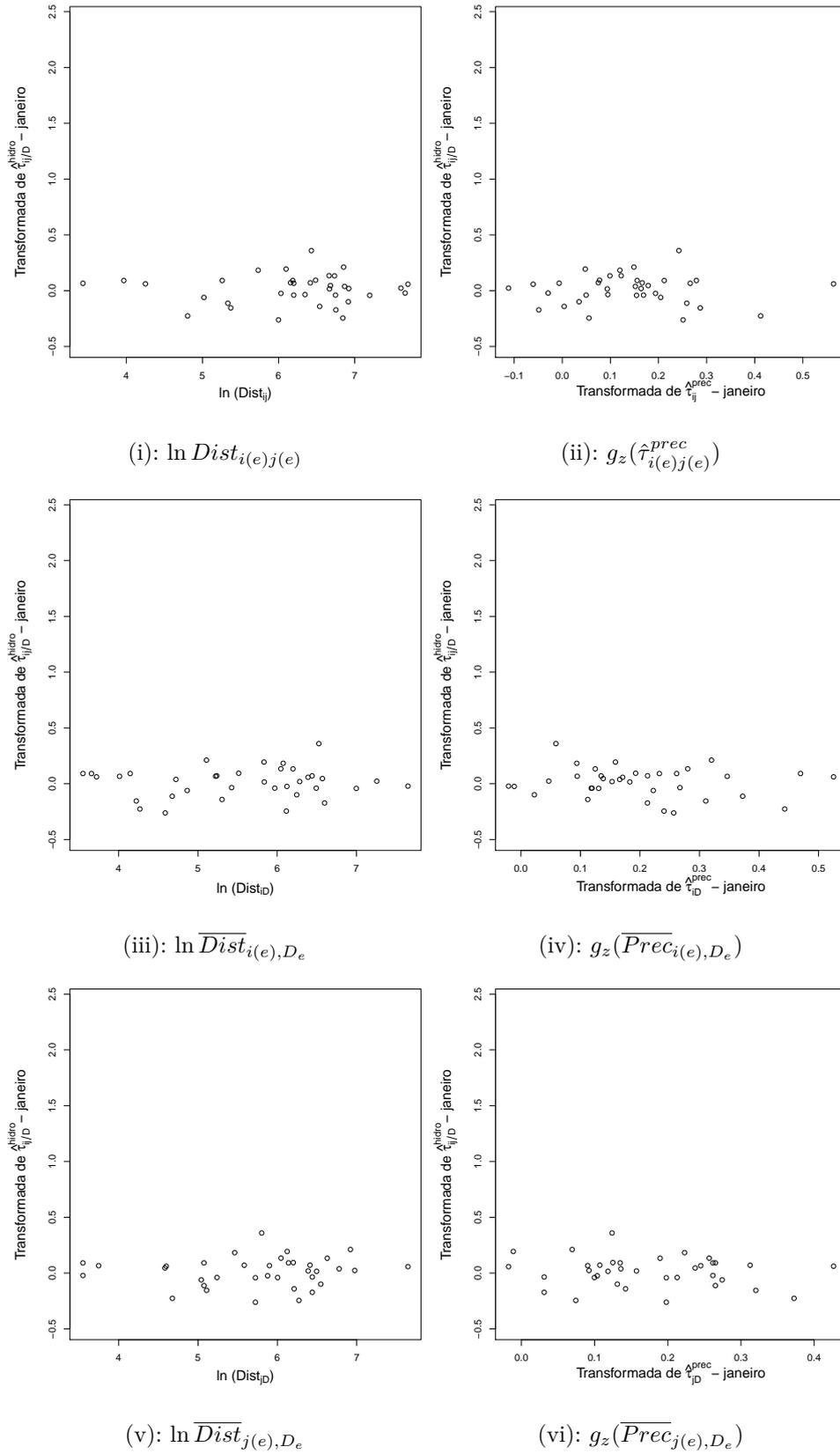
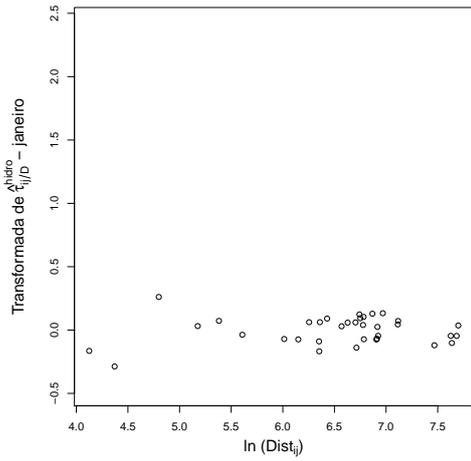
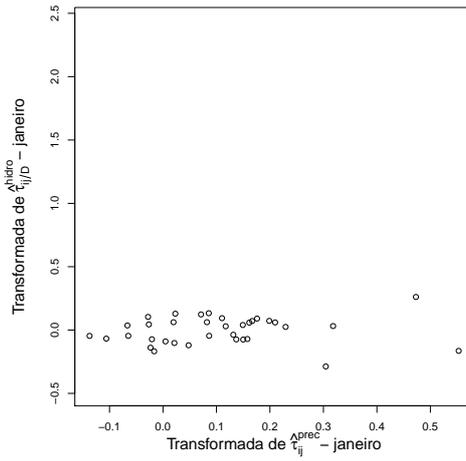


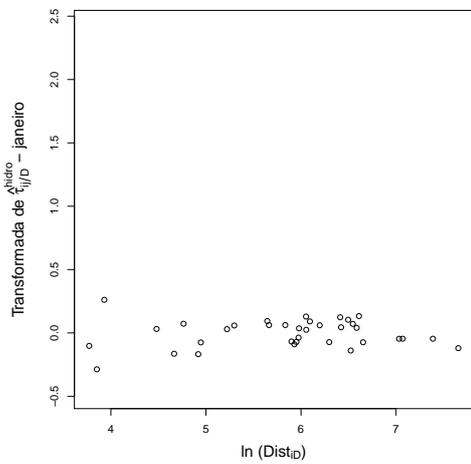
Figura 4.13: Análise das variáveis espaciais para T_3 .



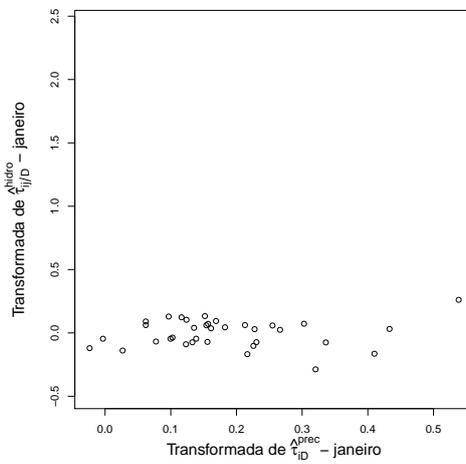
(i): $\ln Dist_{i(e)j(e)}$



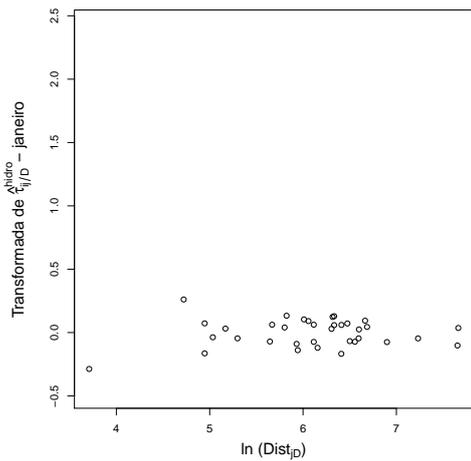
(ii): $g_z(\hat{\tau}_{i(e)j(e)}^{prec})$



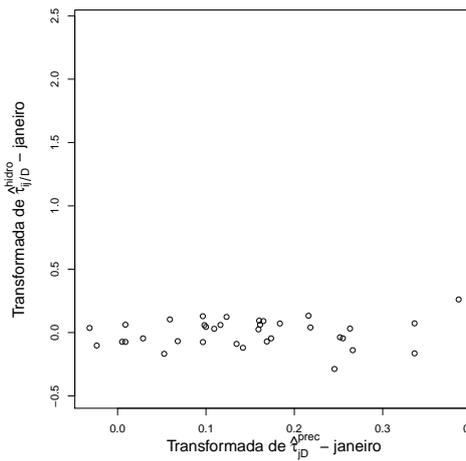
(iii): $\ln \overline{Dist}_{i(e), D_e}$



(iv): $g_z(\overline{Prec}_{i(e), D_e})$



(v): $\ln \overline{Dist}_{j(e), D_e}$



(vi): $g_z(\overline{Prec}_{j(e), D_e})$

Figura 4.14: Análise das variáveis espaciais para T_4 .

4.4.5

Modelo proposto

Apresentamos agora o modelo estimado para as 39 usinas hidrelétricas. Com relação a árvore T_1 , a reparametrização proposta na equação (4-10) para os graus de liberdade (quando existentes) resulta em um vetor $\boldsymbol{\nu}_1 = (\nu_{1,jan}, \dots, \nu_{1,dez})^\top$, ou seja, um parâmetro para cada mês. A equação espacial desenvolvida na seção 4.4.4,

$$h_1(e|\boldsymbol{\beta}_1) = \beta_{1,0} + \beta_{1,1} \ln(Dist_{i(e)j(e)}) + \beta_{1,2} g_z(\hat{\tau}_{i(e)j(e)}^{prec}) + \beta_{1,3} D_{i(e)j(e)} + \sum_{p=1}^{11} \beta_{1,p+3} S_{pm}$$

possui o conjunto de parâmetros $\boldsymbol{\beta}_1 = (\beta_{1,0}, \dots, \beta_{1,14})^\top \in \mathbb{R}^{15}$. Com estas duas reparametrizações, a T_1 possuirá independentemente do número de usinas apenas 27 parâmetros.

Com relação as árvores subsequentes, como já mencionado, não foi possível identificar qualquer tipo de associação entre as variáveis analisadas e o τ de Kendall (condicional). Conseqüentemente, não foi possível estabelecer equações espaciais para as mesmas. Dessa forma, para tais árvores optou-se pela abordagem tradicional de cópulas vine. Os parâmetros das pair-cópulas presentes a partir da árvore T_2 , podem ser organizadas nos conjuntos $\boldsymbol{\beta}_{2:6} = \{\theta_{i(e),j(e);D_e}^{\ell,m} : e \in E_\ell, \ell = 2, \dots, 6; m = 1, \dots, 12\}$ e $\boldsymbol{\nu}_{2:6} = \{\nu_{i(e),j(e);D_e}^{\ell,m} : e \in E_\ell, \ell = 2, \dots, 6; m = 1, \dots, 12\}$.

Os parâmetros são estimados através da maximização da função de verossimilhança utilizando as reparametrizações propostas na equações (4-10) e (4-14). Assim,

$$\mathcal{L}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_{2:6}, \boldsymbol{\nu}_1, \boldsymbol{\nu}_{2:6} | \mathbf{u}^1, \dots, \mathbf{u}^{39}) = \prod_{m=1}^{12} \prod_{t=1}^{82} \prod_{\ell=1}^6 \prod_{e \in E_\ell} c_{i(e),j(e);D_e} \left(\tilde{u}_{t,m}^{i(e)}, \tilde{u}_{t,m}^{j(e)}; \theta_{i(e),j(e);D_e}^{\ell,m}, \nu_{i(e),j(e);D_e}^{\ell,m} \right), \quad (4-16)$$

onde $\tilde{u}_{t,m}^{i(e)} = C_{i(e)|D_e}^m(u_{t,m}^{i(e)} | \mathbf{u}_{t,m}^{D_e})$, $\tilde{u}_{t,m}^{j(e)} = C_{j(e)|D_e}^m(u_{t,m}^{j(e)} | \mathbf{u}_{t,m}^{D_e})$ e $\mathbf{u}_{t,m}^{D_e} := \{u_{t,m}^k : k \in D_e\}$. Como já mencionado, os parâmetros e as cópulas bivariadas variam de acordo com o período enquanto que a estrutura das árvores é constante. Para a maximização da função expressa em 4-16, adotamos a estimação sequencial proposta no contexto de cópulas espaciais por Erhardt et al. [23] para a T_1 e a estimação sequencial tradicional proposta por Aas et al. [13] para as árvores restantes.

Com relação aos valores iniciais para T_1 estimamos uma regressão linear equivalente a equação espacial e fornecemos estes parâmetros como valores iniciais, como sugerido por Erhardt et al. [23]. Tendo em vista os graus de liberdade, os valores iniciais para os mesmos são as médias mensais obtidas na

análise preliminar, ou seja, a média para cada mês da Figura 4.5. Os valores iniciais e finais podem ser observados na Tabela 4.2.

Através da Tabela 4.2, é possível notar que os valores finais são relativamente próximos aos valores iniciais. No caso dos graus de liberdade, todos convergiram para a restrição imposta ($\nu^m \geq 2.5$). Estes parâmetros estão bem próximos das médias mensais estimadas na análise preliminar e, além disso, se observarmos novamente a Figura 4.5, é possível notar uma maior concentração de pontos próximo à 2.

Tabela 4.2: Valores iniciais e finais do modelo estimado.

parâmetro	val. inicial	val. final	parâmetro	val. inicial	val. final
$\beta_{1,0}$	2.2415	1.9357	ν^{jan}	2.6898	2.5
$\beta_{1,1}$	-0.3118	-0.2892	ν^{fev}	2.2872	2.5
$\beta_{1,2}$	-0.1257	0.2168	ν^{mar}	2.2994	2.5
$\beta_{1,3}$	0.7419	0.5666	ν^{abr}	2.3421	2.5
$\beta_{1,4}$	0.0825	0.0467	ν^{mai}	2.2135	2.5
$\beta_{1,5}$	0.0318	0.1535	ν^{jun}	2.5853	2.5
$\beta_{1,6}$	0.0269	0.1351	ν^{jul}	2.6656	2.5
$\beta_{1,7}$	-0.0359	0.0119	ν^{ago}	2.1588	2.5
$\beta_{1,8}$	-0.1719	-0.1669	ν^{set}	2.3215	2.5
$\beta_{1,9}$	-0.3192	-0.2023	ν^{out}	2.3760	2.5
$\beta_{1,10}$	-0.3846	-0.3717	ν^{nov}	2.4294	2.5
$\beta_{1,11}$	-0.3351	-0.3592	ν^{dez}	2.4117	2.5
$\beta_{1,12}$	-0.1653	-0.0763			
$\beta_{1,13}$	-0.0596	0.0351			
$\beta_{1,14}$	-0.0516	-0.0131			

A Tabela 4.3 apresenta os valores da log-verossimilhança obtida por árvore.

Tabela 4.3: Log-verossimilhança por árvore.

	Log-veros.
T_1	29072.95
T_2	760.10
T_3	439.13
T_4	282.05
T_5	182.82
T_6	229.97

Na próxima seção, analisaremos mais a fundo o modelo estimado.

4.5

Resultados

A análise do modelo proposto ocorre em dois momentos distintos. No primeiro, investigamos se o modelo de cópulas estimado é capaz de representar a dependência espacial entre as usinas. Ademais, esta análise visa verificar ainda se a variação periódica existente nos dados foi devidamente incorporada ao modelo. No segundo, examinamos os cenários de vazões simulados. Neste, verificamos se os cenários gerados replicam características observadas nas séries históricas.

4.5.1

Análise da dependência espacial

Esta primeira análise, destina-se a avaliar se o modelo proposto é capaz de replicar as estruturas de dependência existentes entre as usinas. Para tal objetivo, realizamos simulações a partir da cópula estimada e comparamos a dependência espacial simulada com a dependência espacial empírica (calculada com base nos dados históricos).

A partir dos dados cópula históricos $\mathbf{u}^1, \dots, \mathbf{u}^{39}$, para cada mês m é possível estimar o τ de Kendall para todos os pares de usinas i e j . Estes podem ser organizados em uma matriz quadrada simétrica de ordem 39, denominada $\hat{\tau}_m^{emp}$. Em outras palavras, $\hat{\tau}_m^{emp}$ representa a matriz de dependência entre todos os pares de usinas no mês m .

Tendo em vista as simulações, para cada mês m , simulamos $N=5.000$ novas matrizes de dimensão 100×39 . Cada matriz possui 100 observações (dados cópula) para cada uma das 39 usinas.

Para cada mês m e para cada uma das 5000 simulações, estimamos o τ de Kendall $\hat{\tau}_{ij,m}$ para todos os pares de usinas i e j . Em seguida, calculamos o valor médio das simulações e os organizamos em uma matriz $\bar{\tau}_m^{sim}$ cujos elementos são $\bar{\tau}_{ij,m} = \frac{1}{N} \sum_{k=1}^N \hat{\tau}_{ij,m}^k$.

No total, temos 12 matrizes $\hat{\tau}_m^{emp}$ e 12 matrizes $\bar{\tau}_m^{sim}$ representando a associação entre todos os pares de usinas, implicada pelo modelo ou empírica, para determinado mês.

As Figuras 4.15, 4.16, 4.17, 4.18 e 4.19 ilustram as matrizes $\hat{\tau}_m^{emp}$ e $\bar{\tau}_m^{sim}$ através de cores. A transição entre cores ocorre do azul ao vermelho. Tons escuros de azul indicam alta associação positiva entre pares de usinas. Tons escuros de vermelho indicam alta associação negativa. Cores entre azul claro e amarelo claro indicam uma baixa associação ou ausência da mesma.

Como é possível observar, o modelo estimado captura as associações mais significativas (maiores valores do τ de Kendall). Estes pares de usinas

foram modeladas nas primeiras árvores, indicando que a reparametrização proposta é capaz de aproximar razoavelmente a associação entre os pares de usinas mais importantes. É possível observar também que grande parte dos valores simulados estão próximos de zero. Isto decorre diretamente do nível de truncamento adotado (T_6). Os correspondentes pares foram modelados em árvores mais baixas e como mencionado, optamos por utilizar a cópula de independência em tais árvores.

Em problemas de alta dimensão, a maior parte das dependências entre variáveis possui um papel secundário. Geralmente opta-se por capturar apenas as relações mais importantes, deixando as restantes não especificadas. [28].

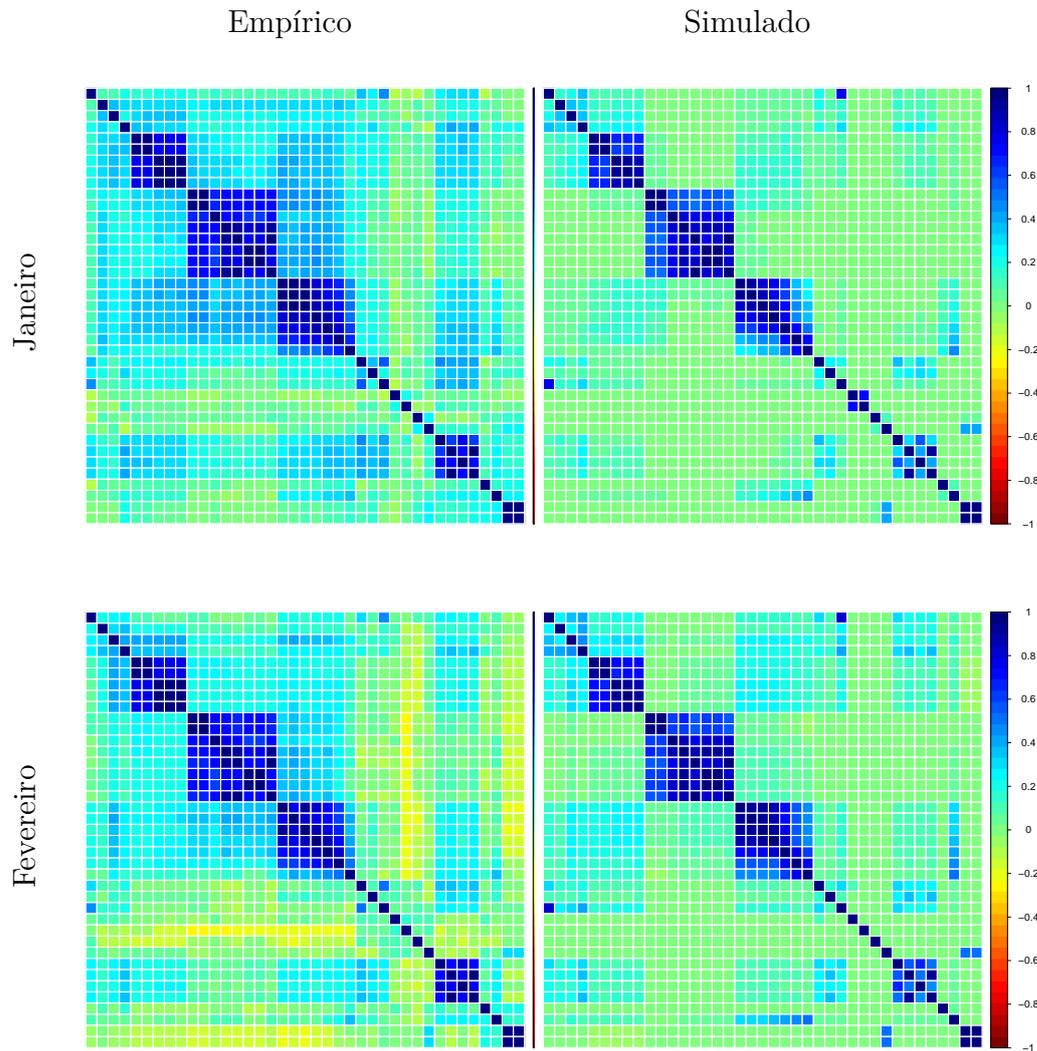
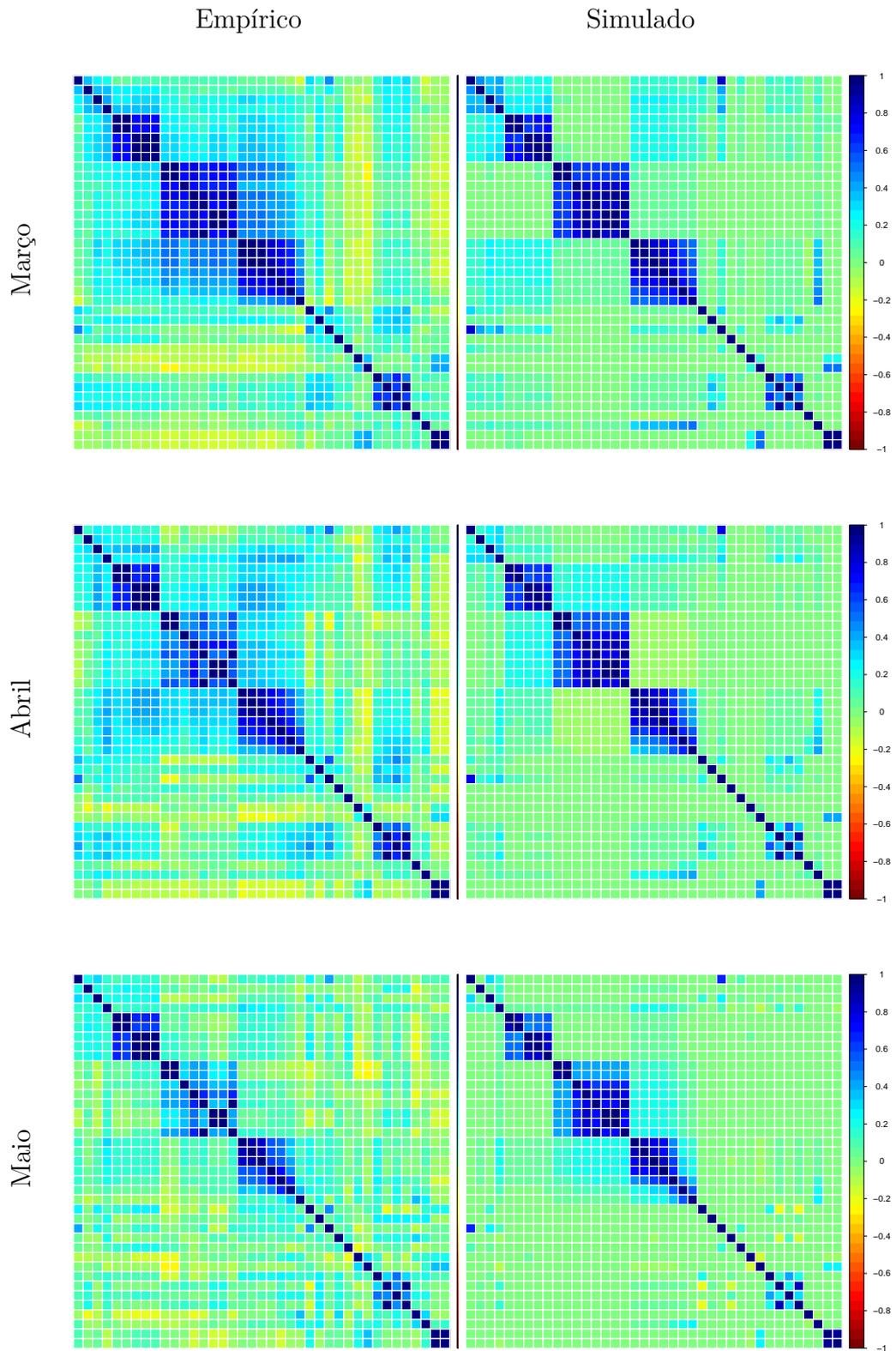


Figura 4.15: Matrizes $\hat{\tau}_m^{emp}$ e $\bar{\tau}_m^{sim}$: janeiro e fevereiro.



PUC-Rio - Certificação Digital N° 1121533/CA

Figura 4.16: Matrizes $\hat{\tau}_m^{emp}$ e $\bar{\tau}_m^{sim}$: março, abril e maio.

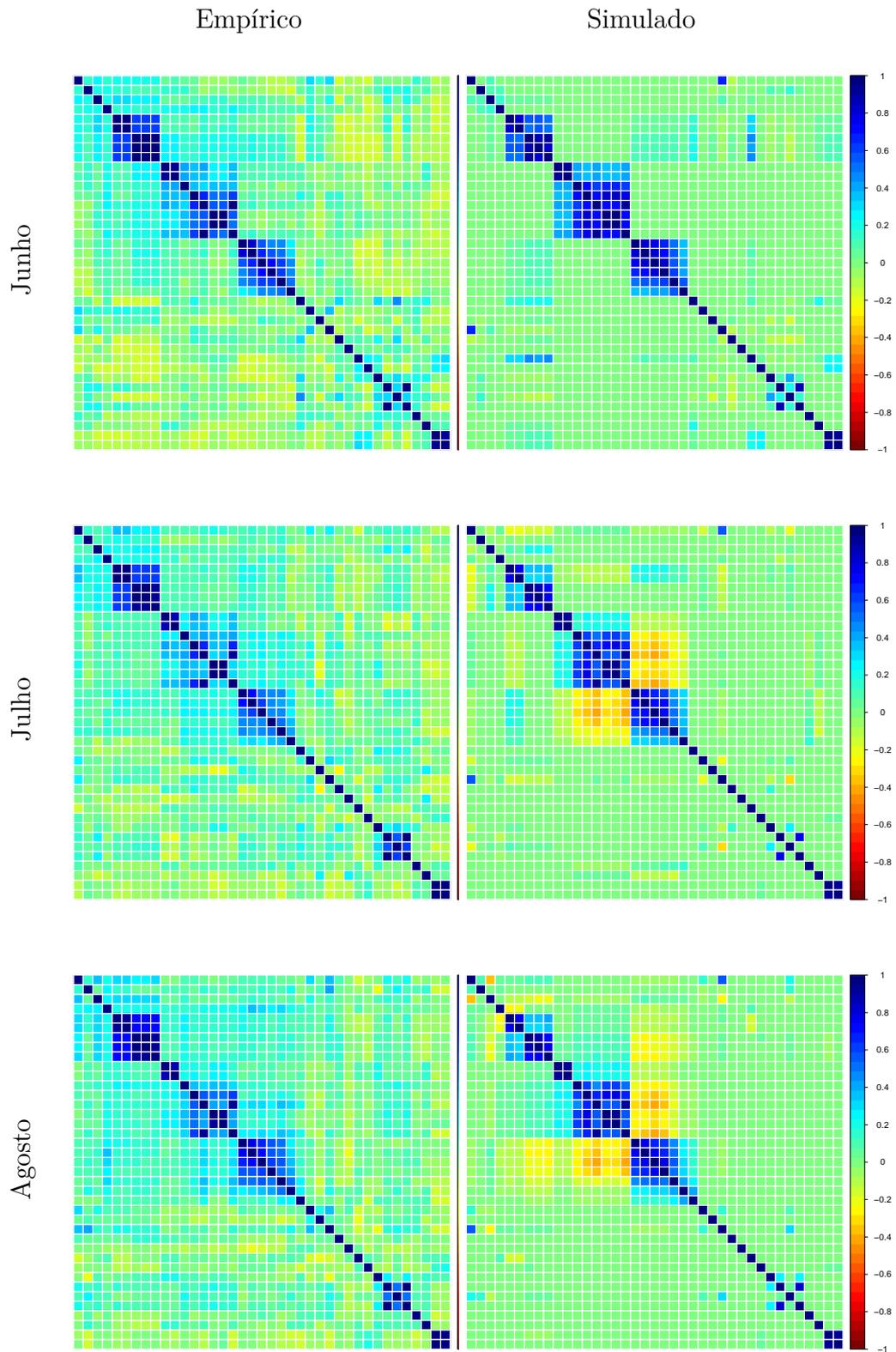


Figura 4.17: Matrizes $\hat{\tau}_m^{emp}$ e $\bar{\tau}_m^{sim}$: junho, julho e agosto.

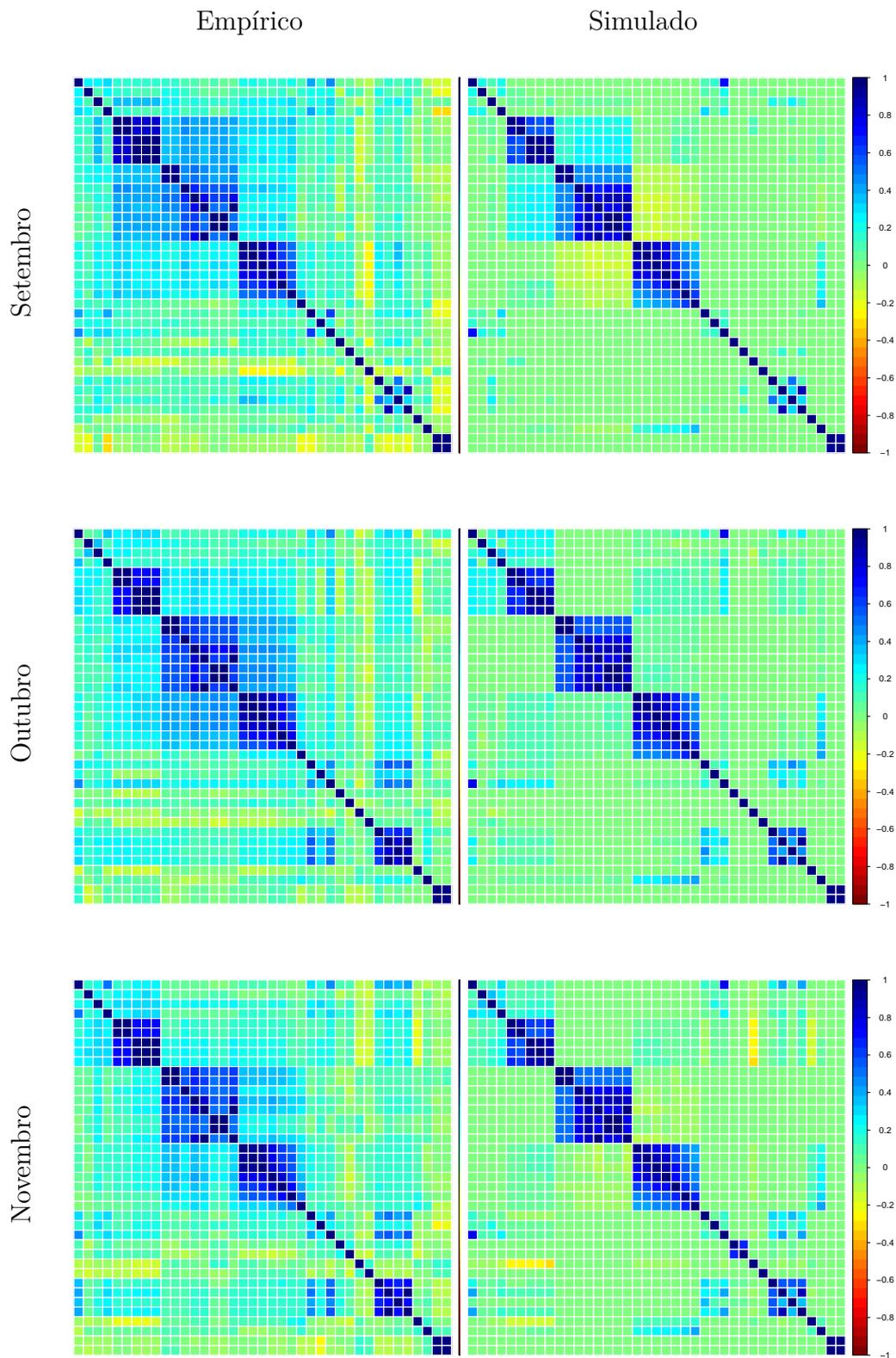


Figura 4.18: Matrizes $\hat{\tau}_m^{emp}$ e $\bar{\tau}_m^{sim}$: setembro, outubro e novembro.

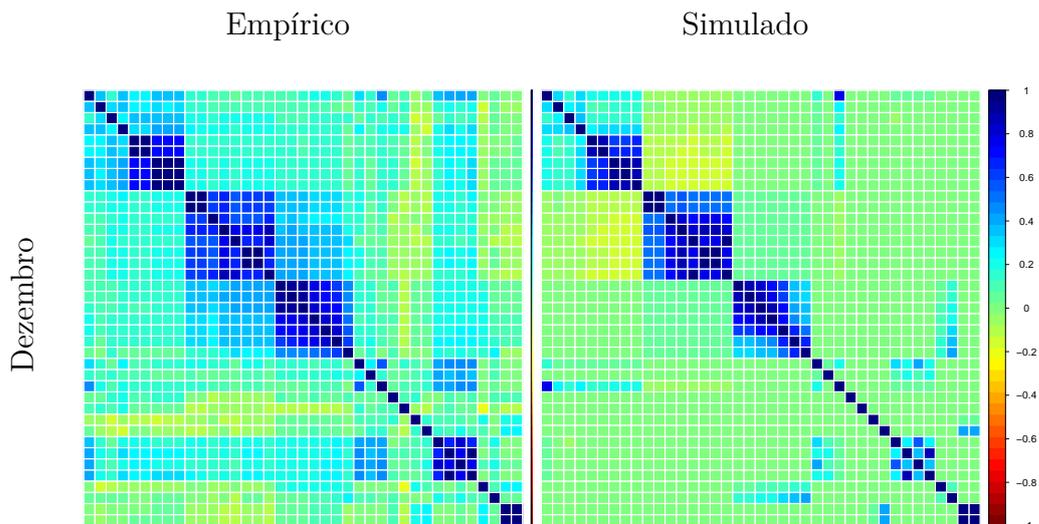


Figura 4.19: Matrizes $\hat{\tau}_m^{emp}$ e $\bar{\tau}_m^{sim}$: dezembro.

No intuito de quantificar a análise realizada, a Figura 4.20 ilustra o absoluto da diferença entre as matrizes $\hat{\tau}_m^{emp}$ e $\bar{\tau}_m^{sim}$ ($|\hat{\tau}_m^{emp} - \bar{\tau}_m^{sim}|$, $m = 1, \dots, 12$). Além disso, a Tabela 4.4 apresenta os percentuais acima de determinados limiares. Isto é, o percentual de diferenças maiores do que 0,05, 0,10, 0,15 e 0,20.

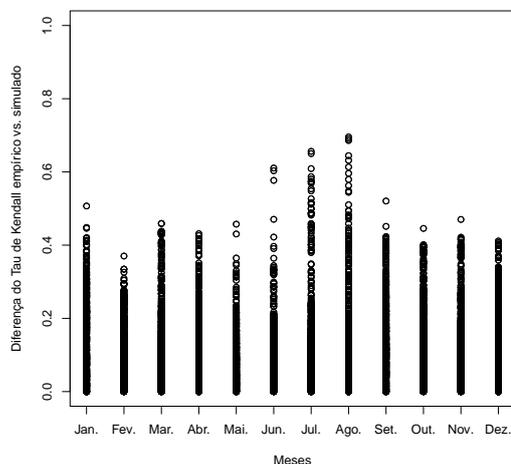


Figura 4.20: Absoluto da diferença entre $\hat{\tau}_m^{emp}$ e $\bar{\tau}_m^{sim}$ para todos os meses.

Tabela 4.4: Absoluto da diferença entre $\hat{\tau}_m^{emp}$ e $\bar{\tau}_m^{sim}$.

	jan.	fev.	mar.	abr.	mai.	jun.	jul.	ago.	set.	out.	nov.	dez.
% > 0,05	0,76	0,65	0,75	0,70	0,60	0,64	0,63	0,65	0,77	0,76	0,71	0,76
% > 0,10	0,56	0,36	0,50	0,49	0,32	0,36	0,37	0,38	0,53	0,50	0,49	0,52
% > 0,15	0,44	0,23	0,26	0,33	0,16	0,17	0,21	0,24	0,31	0,31	0,26	0,32
% > 0,20	0,31	0,12	0,15	0,20	0,06	0,06	0,12	0,15	0,19	0,17	0,13	0,22

Para analisar melhor a reparametrização proposta, realizamos a mesma análise da Tabela 4.4 e da Figura 4.20 somente para os pares de variáveis que compõem a árvore T_1 . Os resultados são apresentados na Figura 4.21 e na Tabela 4.5.

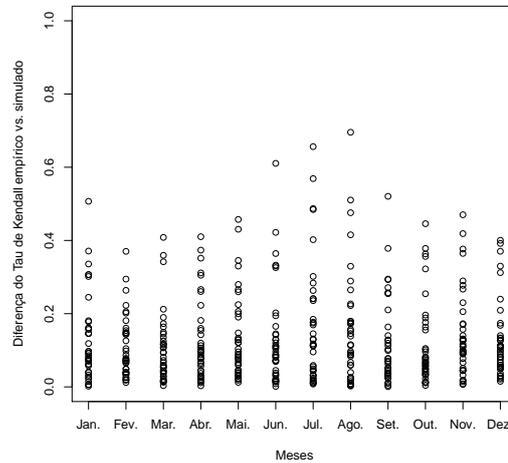


Figura 4.21: Absoluto da diferença ($|\hat{\tau}_m^{emp} - \bar{\tau}_m^{sim}|$): pares presentes na T_1 .

Tabela 4.5: Absoluto da diferença ($|\hat{\tau}_m^{emp} - \bar{\tau}_m^{sim}|$): pares presentes na T_1 .

	jan.	fev.	mar.	abr.	mai.	jun.	jul.	ago.	set.	out.	nov.	dez.
% > 0,05	0,78	0,64	0,67	0,70	0,72	0,70	0,67	0,75	0,62	0,64	0,78	0,81
% > 0,10	0,43	0,40	0,43	0,40	0,48	0,48	0,59	0,54	0,32	0,32	0,59	0,54
% > 0,15	0,32	0,24	0,21	0,29	0,29	0,24	0,40	0,43	0,29	0,29	0,32	0,24
% > 0,20	0,18	0,16	0,10	0,21	0,24	0,18	0,29	0,24	0,16	0,16	0,24	0,18

Por fim, as Figuras 4.22 e 4.23 apresentam comparações entre a variação sazonal empírica e a simulada para alguns pares de usinas presentes na T_1 .

A estrutura do modelo estabelecida pelas variáveis *dummy* periódicas criaram um comportamento em forma de U. Conforme o meio do ano se aproxima, o grau de associação entre usinas diminui. Este padrão também é observado empiricamente em boa parte dos pares de usinas presentes na primeira árvore. Dessa forma, a reparametrização proposta aproxima razoavelmente a variação sazonal presente nos dados. Por outro lado, nos casos onde esse padrão não existe, os valores simulados começam a destoar dos valores reais.

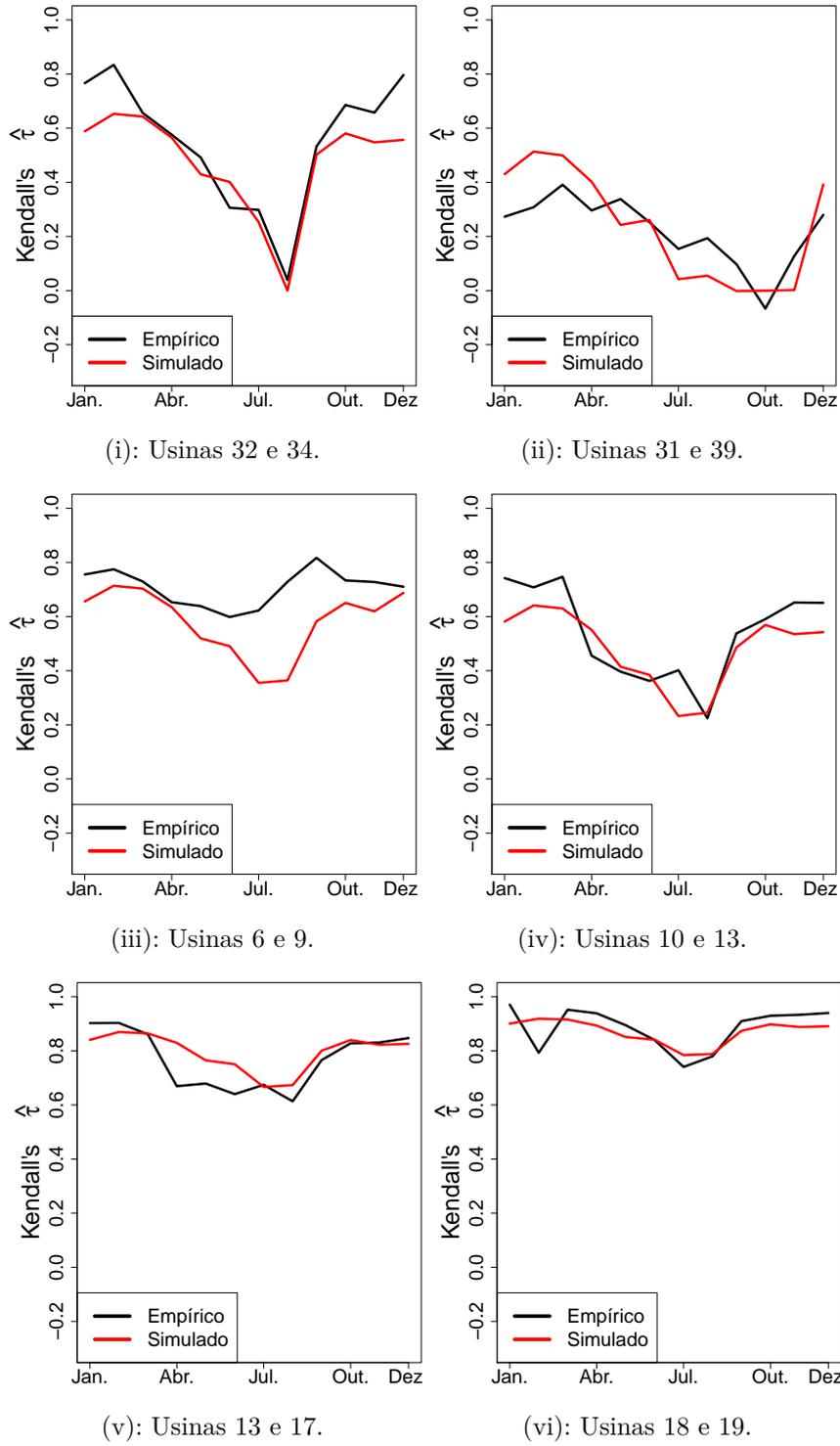


Figura 4.22: Variação sazonal espacial para alguns pares de usinas - I.

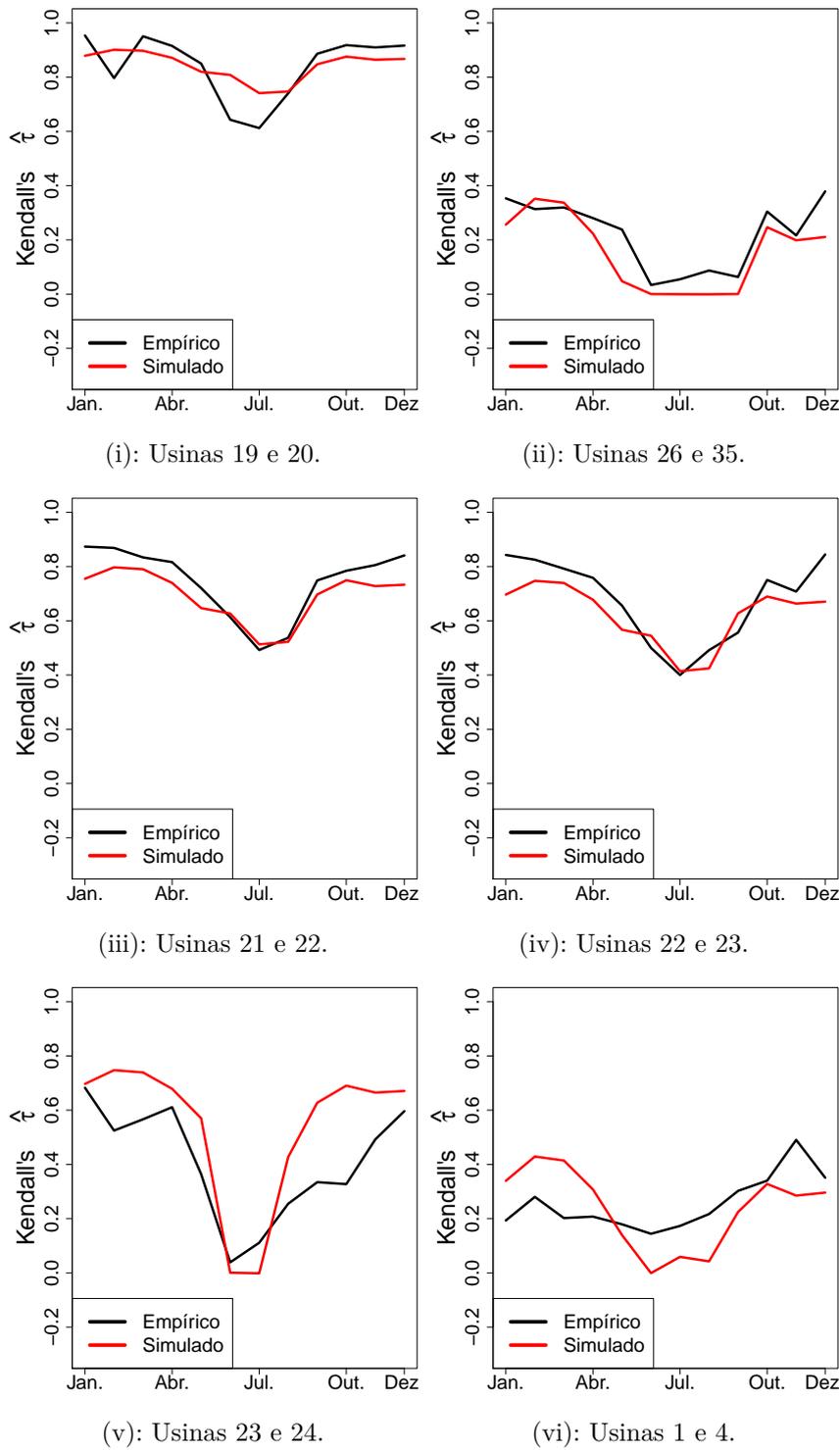


Figura 4.23: Variação sazonal espacial para alguns pares de usinas - II.

4.5.2 Análise dos cenários simulados

Esta seção dedica-se a análise dos cenários de vazões gerados. Estes são obtidos ao injetar-se novos resíduos na equação do modelo de séries temporais (eq. (2-67)). Vale lembrar que estes resíduos são obtidos ao se aplicar a devida

função quantílica às observações sorteadas a partir da cópula vine espacial. Como valores iniciais para geração, empregamos o último ano do histórico disponível.

No total, para cada série de vazão, foram gerados 200 cenários com 120 instantes de tempo (meses), ou seja, 10 anos de cenários. Com esta abordagem, não é possível garantir que todos os cenários sejam estritamente positivos uma vez que os erros sorteados assumem qualquer valor entre $(-\infty, \infty)$.

A Figura 4.24 indica o percentual de cenários válidos do total de 200 cenários simulados para cada uma das 39 usinas. Cada barra representa uma usina hidrelétrica. Como pode ser observado a simulação de cenários com valores negativos é recorrente.

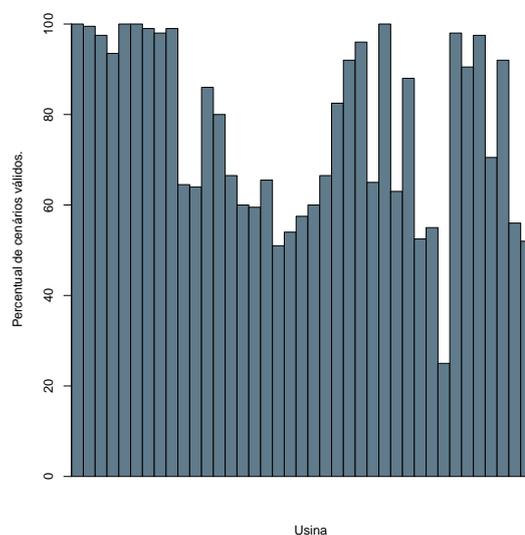
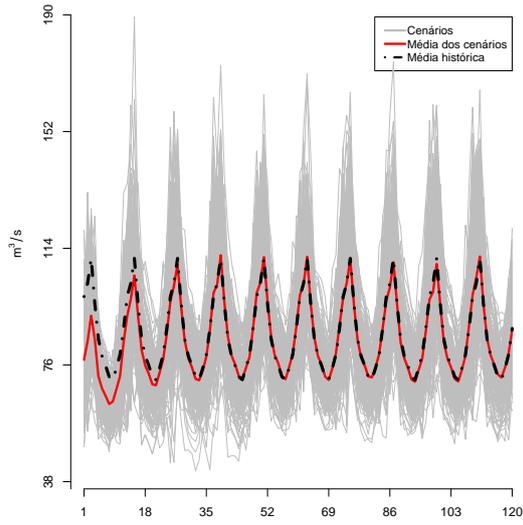


Figura 4.24: Percentual de cenários válidos.

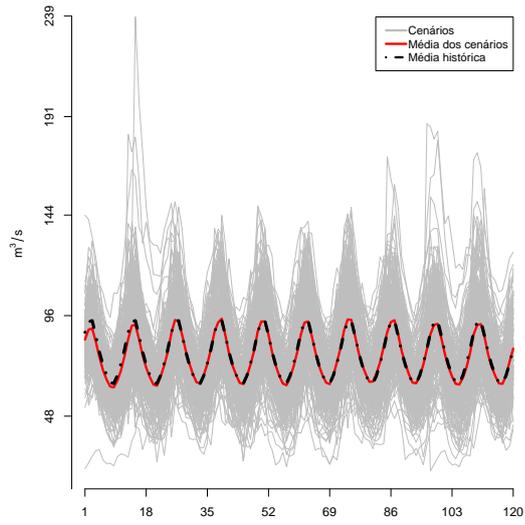
Nas análises seguintes, utilizou-se apenas cenários com valores não negativos (cenários válidos). A partir deste instante, o número de séries simuladas válidas para cada usina é igual a 200.

É desejável que os cenários preservem as principais características das séries históricas. Isto significa que a qualidade do modelo de simulação pode ser auferida ao se verificar se tais características estão presentes nos cenários simulados.

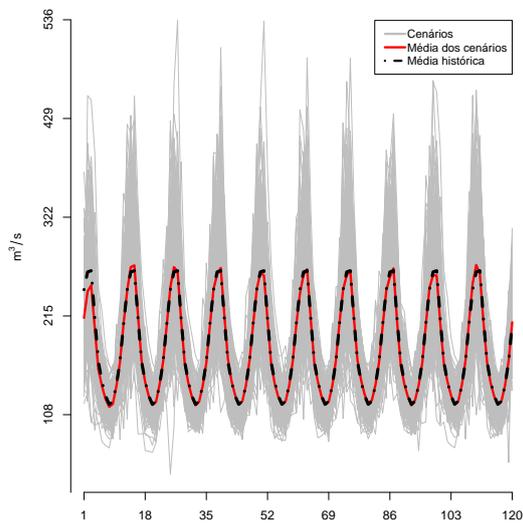
As Figuras 4.25 e 4.26 apresentam uma análise visual dos cenários de vazões simulados para algumas usinas. A linha pontilhada preta representa a média mensal histórica enquanto que a linha vermelha representa a média mensal dos cenários gerados. Estas linhas praticamente se sobrepõem, indicando que o modelo de simulação é capaz de replicar o forte componente periódico existente nas séries de vazões.



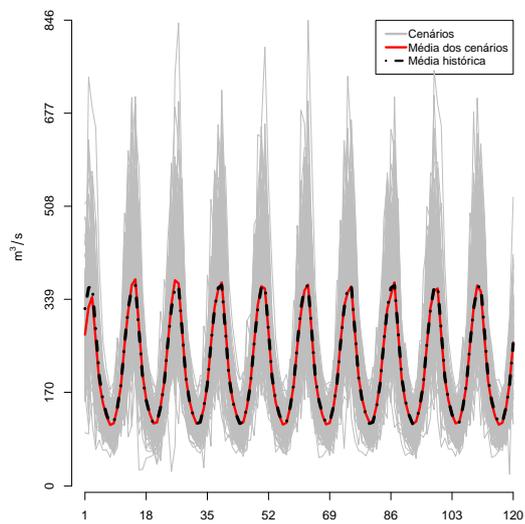
(i): Usina 1.



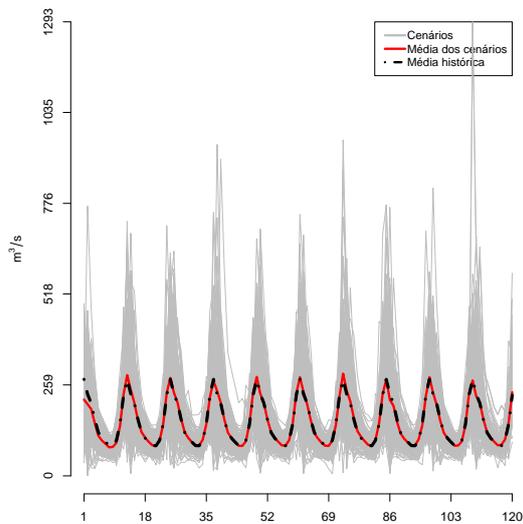
(ii): Usina 2.



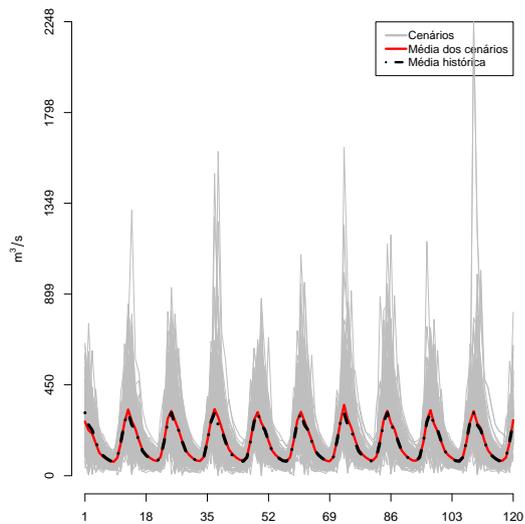
(iii): Usina 5.



(iv): Usina 9.

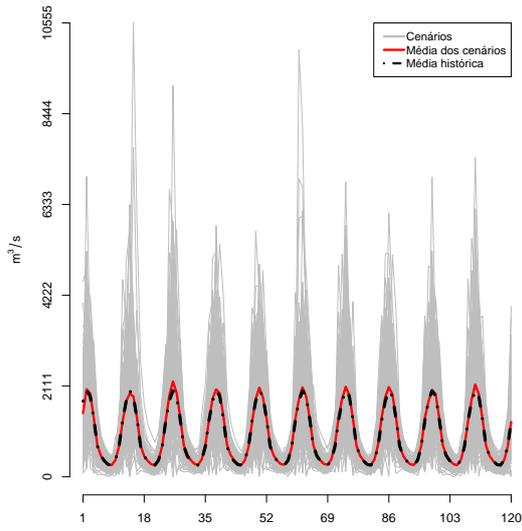


(v): Usina 12.

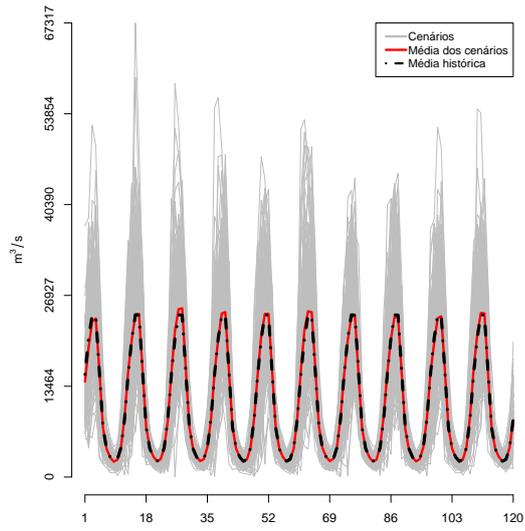


(vi): Usina 16.

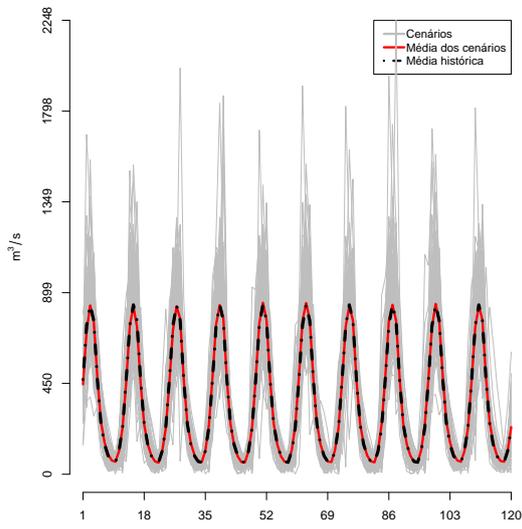
Figura 4.25: Cenários simulados para as usinas 1, 2, 5, 9, 12 e 16.



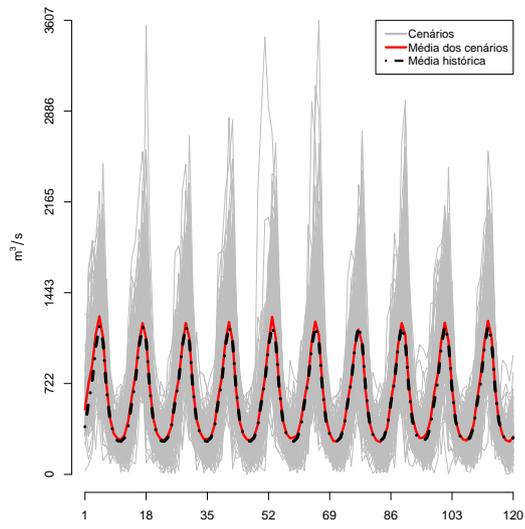
(i): Usina 20.



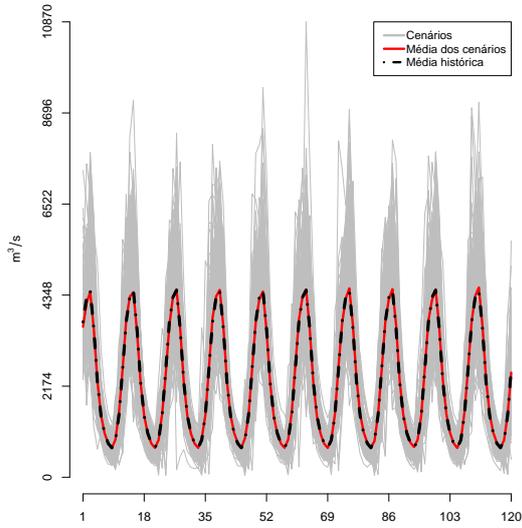
(ii): Usina 24.



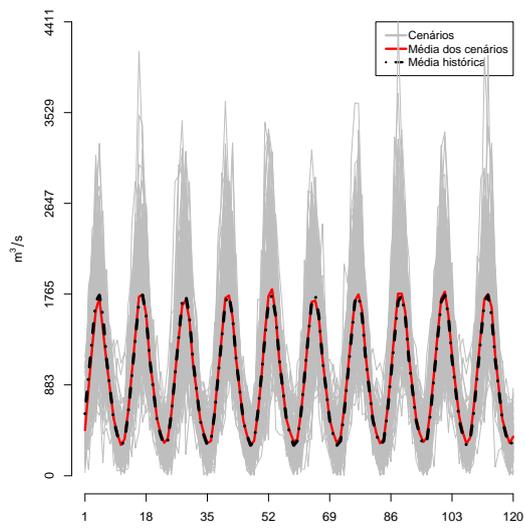
(iii): Usina 28.



(iv): Usina 30.



(v): Usina 34.



(vi): Usina 39.

Figura 4.26: Cenários simulados para as usinas 20, 24, 28, 30, 34 e 39.

De maneira mais formal, realizou-se um conjunto de testes sobre os cenários para averiguar se estes possuíam determinadas propriedades estatísticas observadas historicamente. O primeiro teste empregado foi o teste t para médias. Este tem o objetivo de inferir se as médias mensais dos cenários simulados são estatisticamente iguais às médias mensais históricas. A hipótese nula e alternativa são

$$H_0 : \mu_{hist}^{(m)} = \mu_{cen}^{(m,ano)}$$

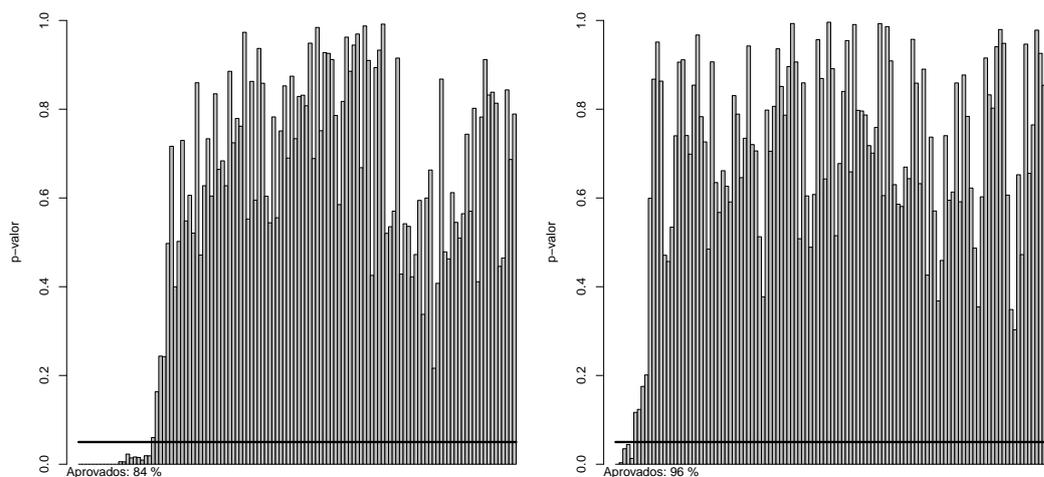
$$H_1 : \mu_{hist}^{(m)} \neq \mu_{cen}^{(m,ano)},$$

onde $m = 1, \dots, 12$ (representam os meses) e $ano = 1, \dots, 10$.

Por exemplo, testa-se se a média de janeiro do primeiro ano simulado ($\mu_{cen}^{(m=1,ano=1)}$) é igual à média histórica de janeiro ($\mu_{hist}^{(m=1)}$). Em seguida, testa-se se a média de fevereiro do primeiro ano simulado ($\mu_{cen}^{(m=2,ano=1)}$) é igual à média histórica de fevereiro ($\mu_{hist}^{(m=2)}$). Estes são realizados sucessivamente para cada um dos 120 períodos simulados. A não rejeição da hipótese nula indica que as médias mensais históricas são estatisticamente iguais às médias mensais dos cenários simulados.

Este teste é realizado individualmente para cada usina, ou seja, têm-se 120 testes por usina. Considerando todas as 39 usinas, no total foram realizados 4680 testes. A um nível de significância de 0.05, o percentual de não rejeição da hipótese nula foi de aproximadamente 96%. As Figuras 4.27 e 4.28 ilustram o teste t para 8 diferentes usinas. A linha pontilhada representa um nível de significância de 0.05.

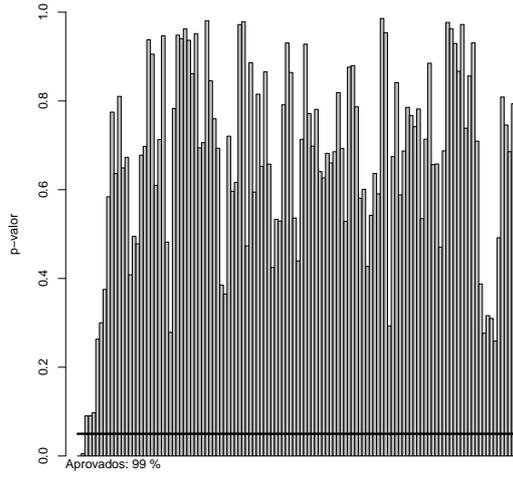
PUC-Rio - Certificação Digital Nº 1121533/CA



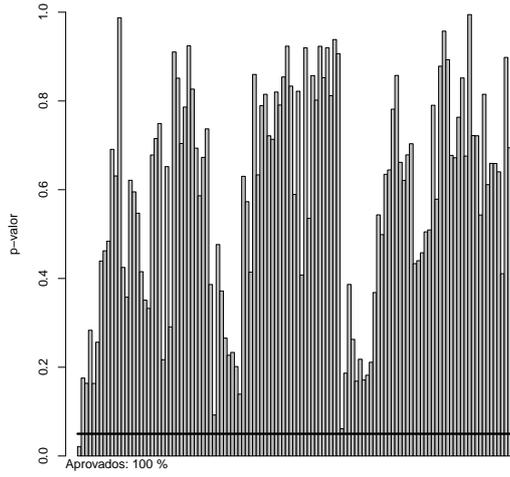
(i): Usina 1.

(ii): Usina 5.

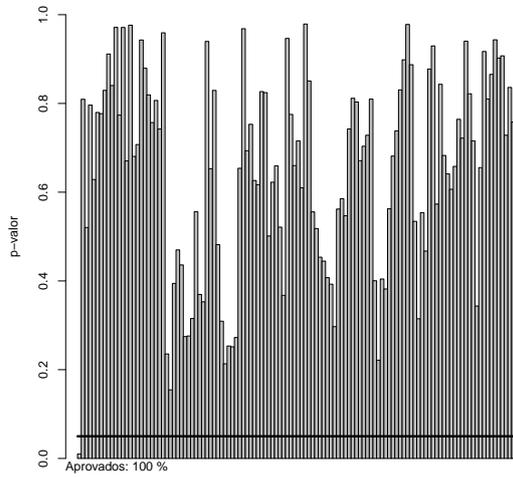
Figura 4.27: Teste t: usinas 1 e 5.



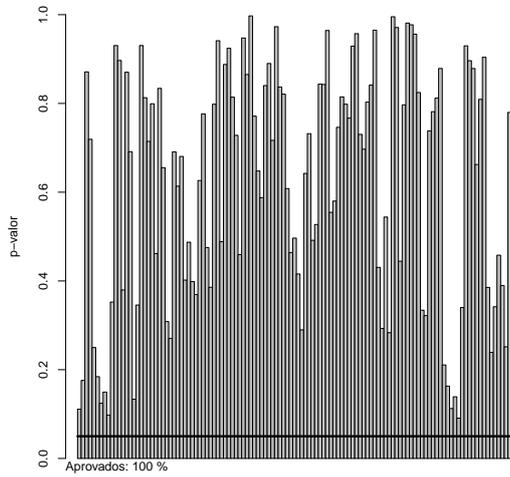
(i): Usina 9.



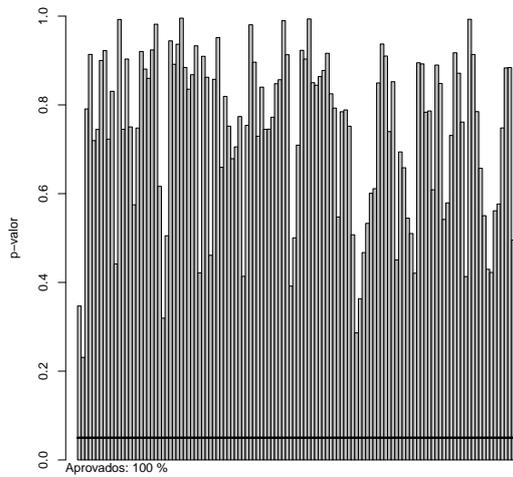
(ii): Usina 16.



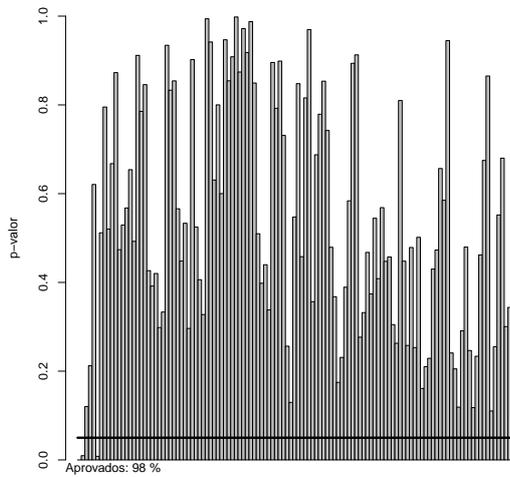
(iii): Usina 20.



(iv): Usina 28.



(v): Usina 34.



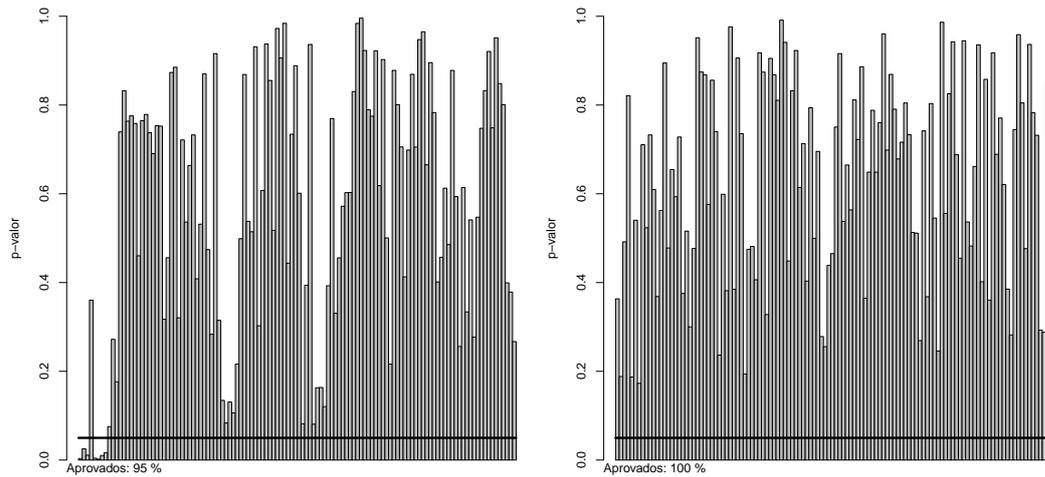
(vi): Usina 39.

Figura 4.28: Teste t: usinas 9, 16, 20, 28, 34 e 39.

O segundo teste realizado foi o teste de Levene. Seu propósito é verificar a igualdade das variâncias. Deseja-se saber se a variância de cada período simulado é estatisticamente igual a respectiva variância no histórico. A hipótese nula e alternativa são

$$\begin{aligned}
 H_0 : \sigma_{hist}^{2(m)} &= \sigma_{cen}^{2(m,ano)}, \\
 H_1 : \sigma_{hist}^{2(m)} &\neq \sigma_{cen}^{2(m,ano)}.
 \end{aligned}
 \tag{4-17}$$

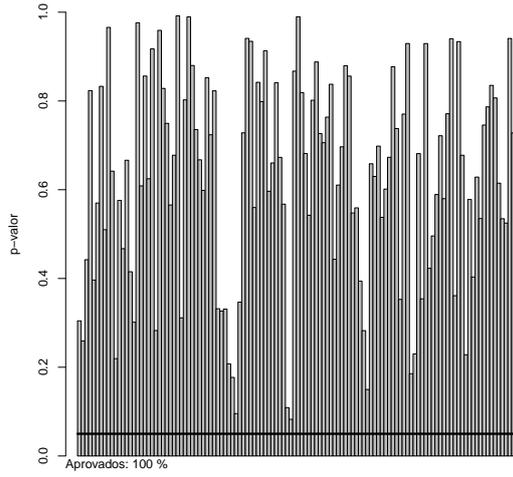
Do total de 4680 testes realizados, aproximadamente 98% indicam que a hipótese nula não pode ser rejeita a um nível de significância de 5%. As Figuras 4.29 e 4.30 apresentam os p-valores para os 120 períodos simulados de algumas usinas selecionadas. A linha pontilhada representa um nível de significância de 0.05.



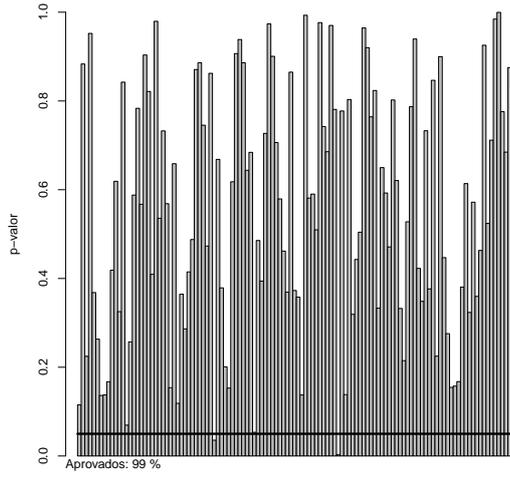
(i): Usina 1.

(ii): Usina 5.

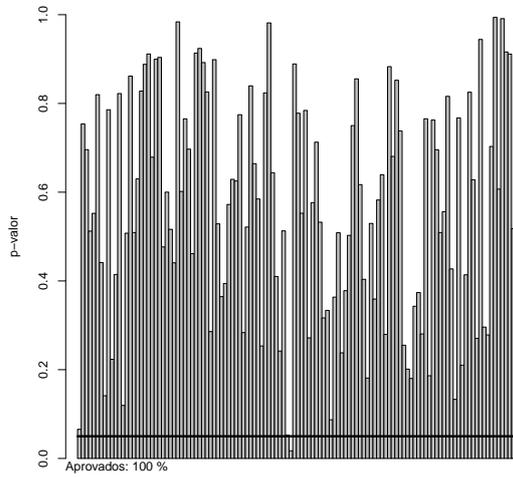
Figura 4.29: Levene: usinas 1 e 5.



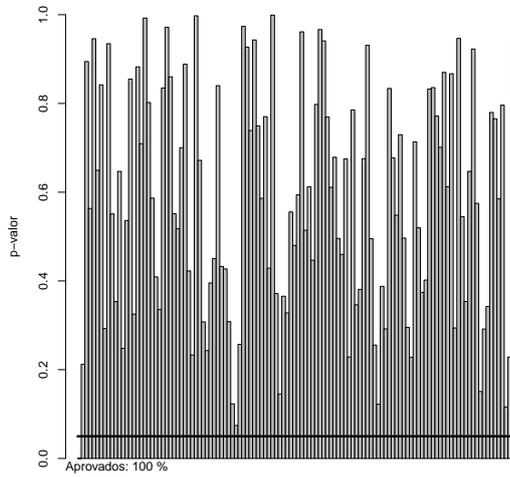
(i): Usina 9.



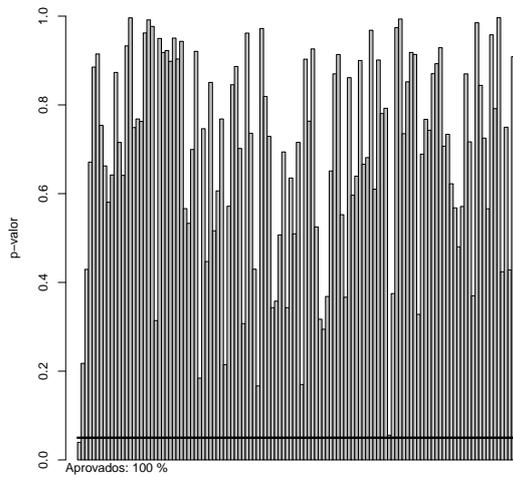
(ii): Usina 16.



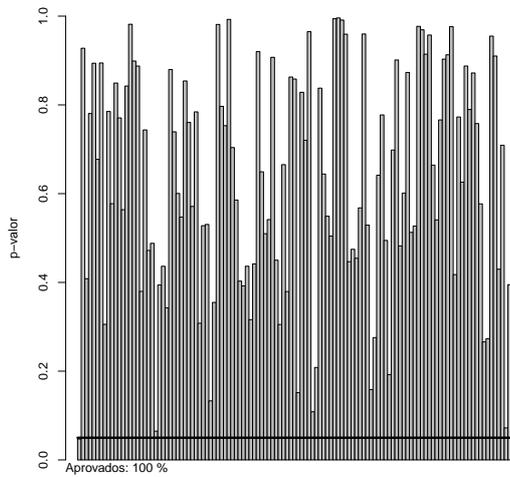
(iii): Usina 20.



(iv): Usina 28.



(v): Usina 34.



(vi): Usina 39.

Figura 4.30: Levene: usinas 9, 16, 20, 28, 34 e 39.

Por fim, o último teste empregado foi o teste de Kolmogorov-Smirnov (K-S). A finalidade desde é verificar se duas amostras são originárias de uma mesma distribuição de probabilidade. Neste sentido, o objetivo é averiguar se, para determinado período, o histórico e os cenários seguem a mesma distribuição. Sendo $F_{hist}^{(m)}(\cdot)$ e $F_{cen}^{(m,ano)}(\cdot)$ as distribuições de probabilidade do histórico e dos cenários respectivamente, as hipóteses nula e alternativa são

$$\begin{aligned}
 H_0 &: F_{hist}^{(m)} = F_{cen}^{(m,ano)}, \\
 H_1 &: F_{hist}^{(m)} \neq F_{cen}^{(m,ano)}.
 \end{aligned}
 \tag{4-18}$$

Nesta análise, em aproximadamente 92% dos casos a hipótese nula não foi rejeitada. As Figuras 4.31 e 4.32 apresentam os resultados dos testes de 8 usinas hidrelétricas selecionadas.

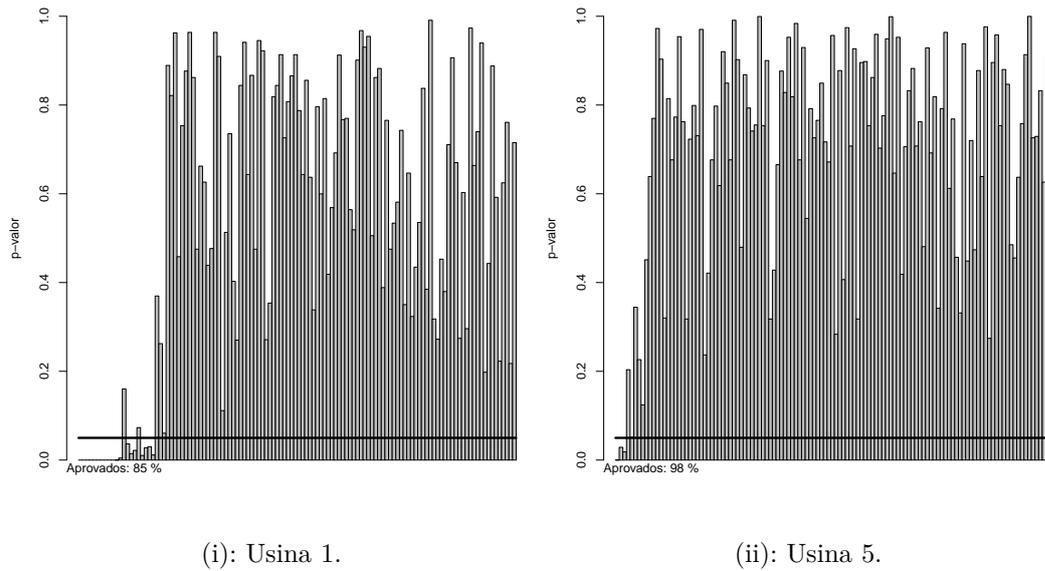
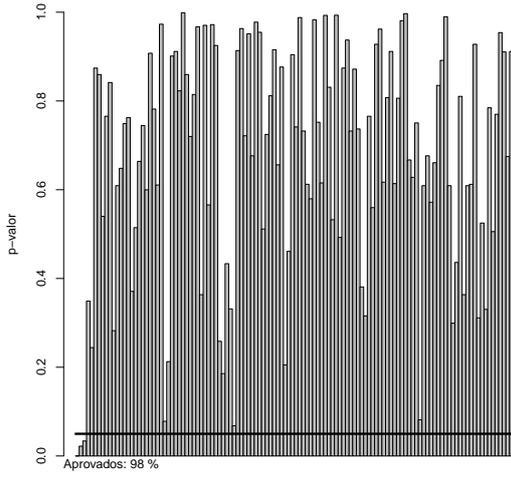
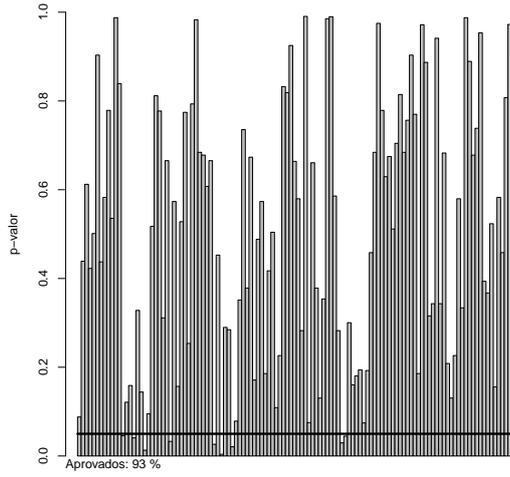


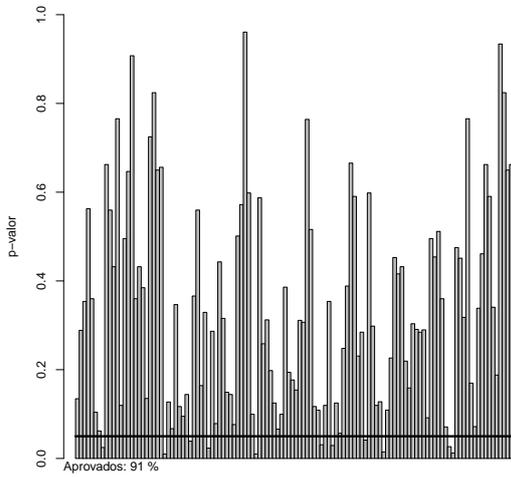
Figura 4.31: K-S: usinas 1 e 5.



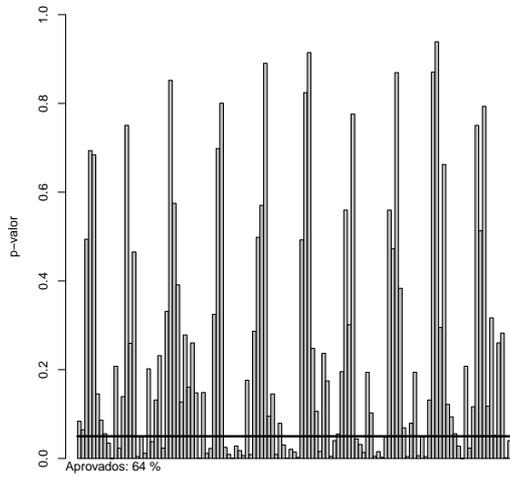
(i): Usina 9.



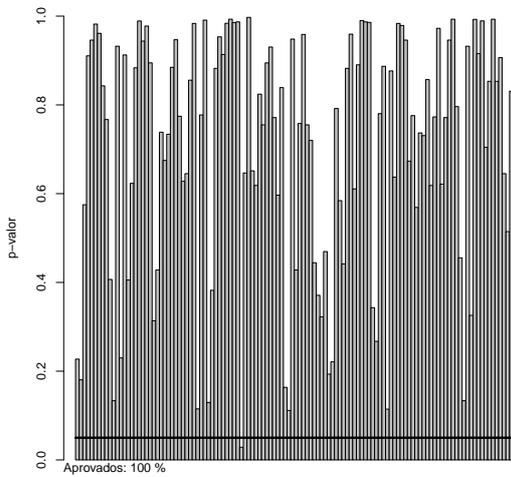
(ii): Usina 16.



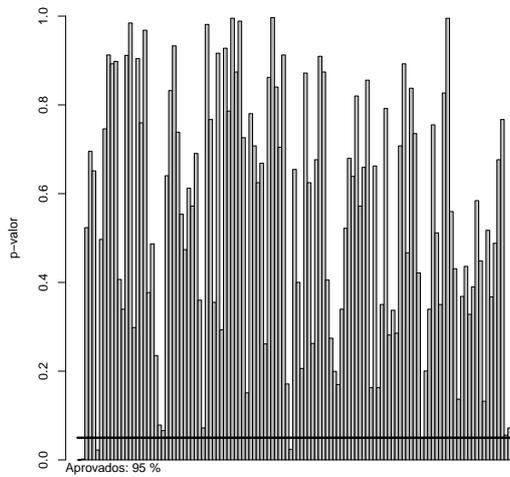
(iii): Usina 20.



(iv): Usina 28.



(v): Usina 34.



(vi): Usina 39.

Figura 4.32: K-S: usinas 9, 16, 20, 28, 34 e 39.

A Tabela 4.6 resume os resultados dos três testes para todas as usinas. Como pode ser observado, o percentual de aprovação é relativamente alto.

Tabela 4.6: Análise dos cenários: Teste t, Levene e Kolmogorov-Smirnov.

	Taxa de aprovação				Taxa de aprovação		
	t	Levene	K-S		t	Levene	K-S
Usina 1	84%	95%	85%	Usina 21	100%	100%	95%
Usina 2	100%	97%	99%	Usina 22	100%	100%	99%
Usina 3	98%	99%	96%	Usina 23	100%	100%	93%
Usina 4	98%	95%	96%	Usina 24	100%	100%	100
Usina 5	96%	100%	98%	Usina 25	93%	99%	90%
Usina 6	96%	100%	99%	Usina 26	100%	95%	83%
Usina 7	99%	100%	98%	Usina 27	62%	88%	40%
Usina 8	99%	100%	98%	Usina 28	100%	100%	64%
Usina 9	99%	100%	98%	Usina 29	100%	97%	95%
Usina 10	94%	97%	95%	Usina 30	97%	100%	85%
Usina 11	94%	95%	95%	Usina 31	99%	99%	97%
Usina 12	98%	98%	98%	Usina 32	100%	99%	93%
Usina 13	96%	95%	98%	Usina 33	100%	100%	100%
Usina 14	92%	97%	91%	Usina 34	100%	100%	100%
Usina 15	100%	99%	95%	Usina 35	100%	100%	99%
Usina 16	100%	99%	93%	Usina 36	87%	98%	77%
Usina 17	92%	97%	90%	Usina 37	100%	99%	99%
Usina 18	100%	100%	89%	Usina 38	98%	100%	97%
Usina 19	100%	100%	88%	Usina 39	98%	100%	95%
Usina 20	100%	100%	91%				

4.6

Algumas alternativas para o modelo desenvolvido.

Algumas configurações diferentes para o modelo anteriormente estimado são testadas nesta seção. As principais modificações estão relacionadas à distribuição dos erros do modelo de séries temporais e as variáveis espaciais. Estas configurações foram implementadas e os principais resultados são discutidos neste instante.

Modelos marginais

Como já mencionado, a abordagem adotada não garante que todos os cenários simulados sejam válidos (positivos). Tal desvantagem pode ser

facilmente contornada ao se aplicar a transformação logarítmica às séries de vazões originais.

Após essa transformação, estimou-se o modelo PAR(p) e verificamos que os resíduos padronizados, em sua grande maioria, seguiam uma distribuição Gaussiana padrão. Desta forma, a função de distribuição acumulada normal padrão foi utilizada para transformar os resíduos do modelo em dados cópula. Na seção anterior, vale lembrar, empregamos a distribuição *skew - t*. Em termos práticos, tal substituição torna o modelo computacionalmente menos custoso uma vez que não é necessário estimar os 4 parâmetros da distribuição *skew - t*. Além de ser mais fácil avaliar as funções acumuladas e quantílicas da distribuição normal.

Variáveis espaciais

Uma regressão linear múltipla entre $g_z(\hat{\tau}_{ij,m}^{hidro})$, $1 \leq i < j \leq 39$, $m = 1, \dots, 12$ e as 14 variáveis exógenas (3 variáveis espaciais e as 11 variáveis dummies) discutidas na Subseção 4.4.3 foi estimada. A Tabela 4.7 apresenta os parâmetros estimados e os respectivos p-valores entre parênteses. Nesta regressão o R^2 ajustado foi de 0,76.

Tabela 4.7: Regressão estimada com as variáveis espaciais.

	val. estimado		val. estimado
const.	2,02 (0,001)	Mai.	-0,21 (0,021)
log(Dist)	-0,27 (0,001)	Jun.	-0,32 (0,001)
log($\hat{\tau}_{ij}^{chuva}$)	0,05 (0,730)	Jul.	-0,39 (0,001)
D	0,80 (0,001)	Ago.	-0,36 (0,001)
Jan.	0,05 (0,521)	Set.	-0,20 (0,002)
Fev.	-0,01 (0,947)	Out.	-0,06 (0,453)
Mar.	0,01 (0,878)	Nov.	-0,07 (0,405)
Abr.	-0,05 (0,567)		

Isto posto, decidimos estimar o modelo espacial levando-se em consideração apenas as variáveis estatisticamente significantes. O nível de significância considerado foi de 0,05. Desta forma, a equação espacial para T_1 passou a ser

$$h_1(e|\beta_1) = \beta_{1,0} + \beta_{1,1} \ln(Dist_{i(e)j(e)}) + \beta_{1,2} D_{i(e)j(e)} + \sum_{q=5}^9 \beta_{1,q-2} S_{qm}. \quad (4-19)$$

Outra modificação em relação ao modelo estimado na Seção 4.4 diz respeito aos graus de liberdade da cópula t presentes na árvore T_1 . Diferente

do caso anterior onde estes foram estimadas periodicamente, optamos neste instante por utilizar a média mensal obtida na análise preliminar. Deste modo, a maximização da verossimilhança ocorre apenas em termos do primeiro parâmetro das cópulas.

Os procedimentos para definição das árvores e seleção das pair-cópulas foram mantidos. Optamos neste caso por truncar o modelo em T_4 . Os parâmetros da equação espacial estimados são apresentados na Tabela 4.8, enquanto que a Tabela 4.9 exhibe a log-verossimilhança por árvore.

Tabela 4.8: Parâmetros da equação espacial estimada.

	const.	log(Dist)	D	Mai.	Jun.	Jul.	Ago.	Set.
$\hat{\alpha}$	1,93	-0,28	0,64	-0,20	-0,24	-0,36	-0,35	-0,12

Tabela 4.9: Modelo alternativo: log-verossimilhança por árvore.

	Log-veros.
T_1	29018,21
T_2	684,52
T_3	408,18
T_4	170,37

Resultados da simulação

As principais análises realizadas na Seção 4.5 foram repetidas. A Figura 4.33 compara as matrizes mensais do τ de Kendall obtidas via simulação ($\hat{\tau}_m^{sim}$) e empiricamente ($\hat{\tau}_m^{emp}$). Uma inspeção visual permite afirmar que o modelo é capaz de aproximar as maiores associações espaciais entre usinas. Comparando as duas configurações, aparentemente não é possível identificar consideráveis diferenças entre as abordagens.

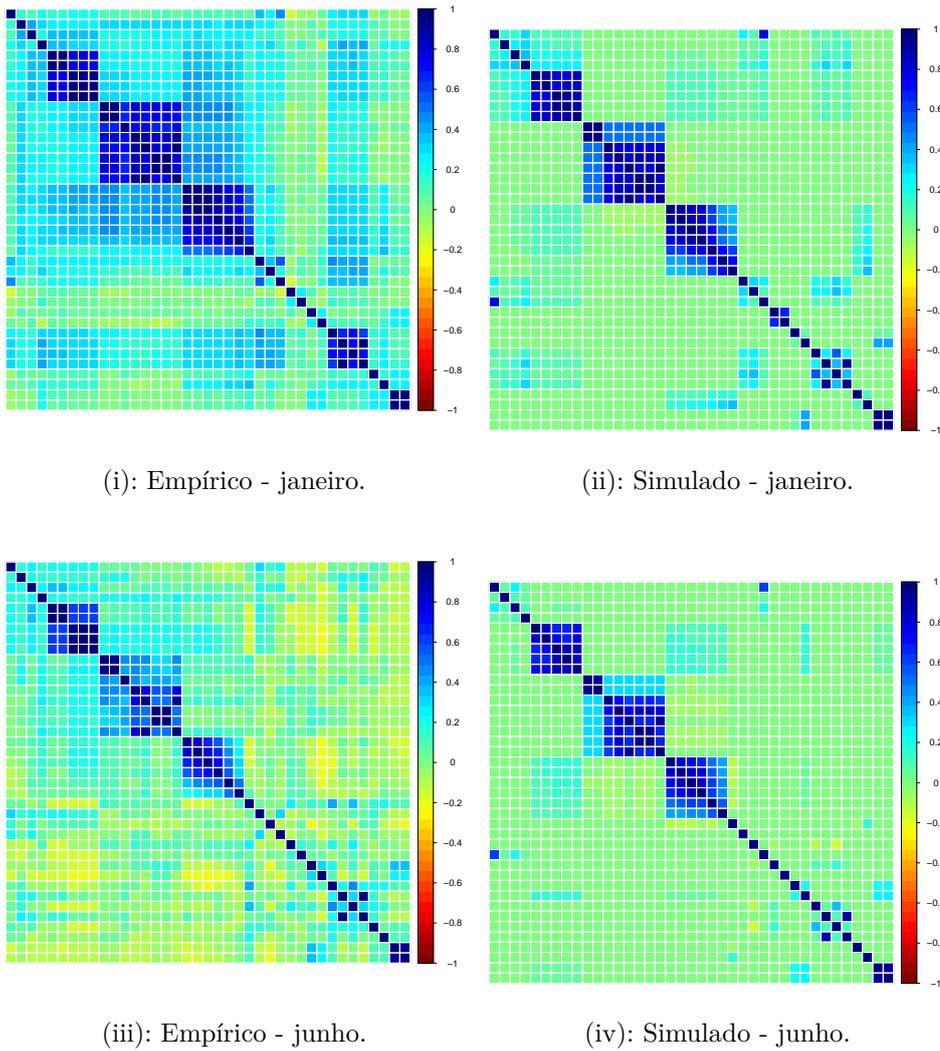


Figura 4.33: Modelo alternativo: matrizes $\hat{\tau}_m^{emp}$ e $\bar{\tau}_m^{sim}$ para janeiro e junho.

As Tabelas 4.11 e 4.10 exibem os percentuais de pares de usinas cujo absoluto da diferença ultrapassam determinados limiares. Novamente, não é possível destacar diferenças consideráveis entre as duas abordagens.

Tabela 4.10: Modelo alternativo: $|\hat{\tau}_m^{emp} - \bar{\tau}_m^{sim}|$ para os pares de usinas da T_1 .

	jan.	fev.	mar.	abr.	mai.	jun.	jul.	ago.	set.	out.	nov.	dez.
% > 0,05	0,78	0,78	0,81	0,70	0,78	0,78	0,78	0,70	0,70	0,75	0,78	0,75
% > 0,10	0,45	0,37	0,43	0,48	0,62	0,51	0,54	0,45	0,54	0,51	0,54	0,37
% > 0,15	0,35	0,27	0,27	0,29	0,35	0,21	0,40	0,29	0,21	0,29	0,32	0,24
% > 0,20	0,21	0,24	0,18	0,18	0,29	0,16	0,35	0,29	0,18	0,16	0,27	0,16

Tabela 4.11: Modelo alternativo: absoluto da diferença entre $\hat{\tau}_m^{emp}$ e $\bar{\tau}_m^{sim}$.

	jan.	fev.	mar.	abr.	mai.	jun.	jul.	ago.	set.	out.	nov.	dez.
% > 0,05	0,75	0,74	0,75	0,76	0,66	0,64	0,66	0,68	0,80	0,79	0,72	0,75
% > 0,10	0,58	0,49	0,52	0,52	0,42	0,40	0,40	0,42	0,56	0,52	0,46	0,52
% > 0,15	0,46	0,32	0,33	0,37	0,22	0,17	0,25	0,20	0,33	0,32	0,27	0,35
% > 0,20	0,36	0,21	0,19	0,22	0,11	0,07	0,17	0,10	0,19	0,19	0,14	0,21

De forma similar na subseção 4.5.1, a Figura 4.34 ilustra a variação sazonal espacial entre alguns pares de usinas.

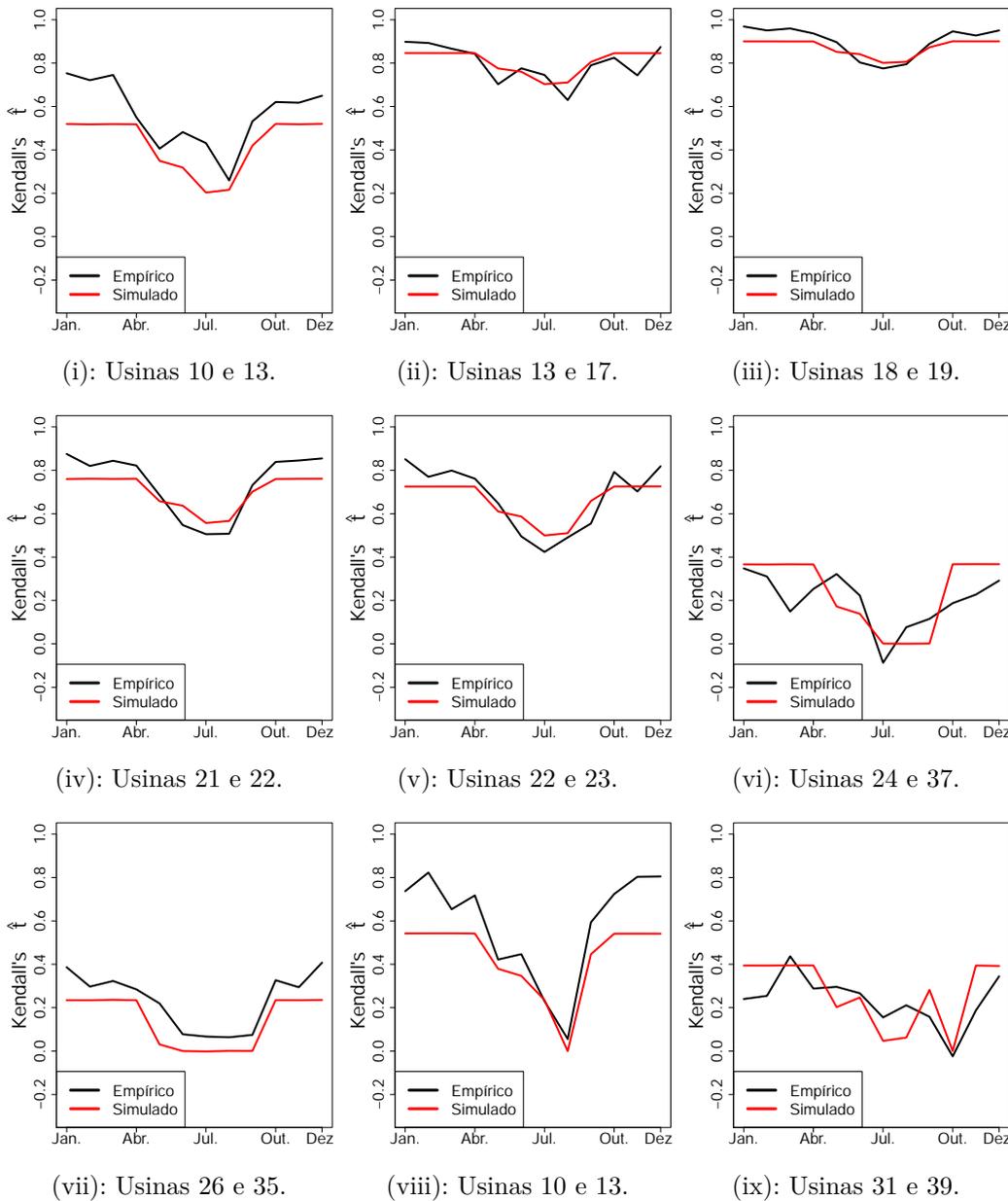


Figura 4.34: Modelo alternativo: variação sazonal espacial.

Por fim, simulamos 200 cenários de tamanho 120 (10 anos) para cada

uma das 39 usinas. Vale lembrar que neste caso, nenhum dos cenários possui valores negativos. Os testes de média, de variância e de Kolmogorov-Smirnov foram empregados novamente. Os respectivos percentuais de aprovação foram de aproximadamente 95%, 95% e 92%.

A Figura 4.35 apresenta os cenários simulados de algumas usinas selecionadas enquanto que a Tabela 4.12 resume o resultado dos testes para cada usina.

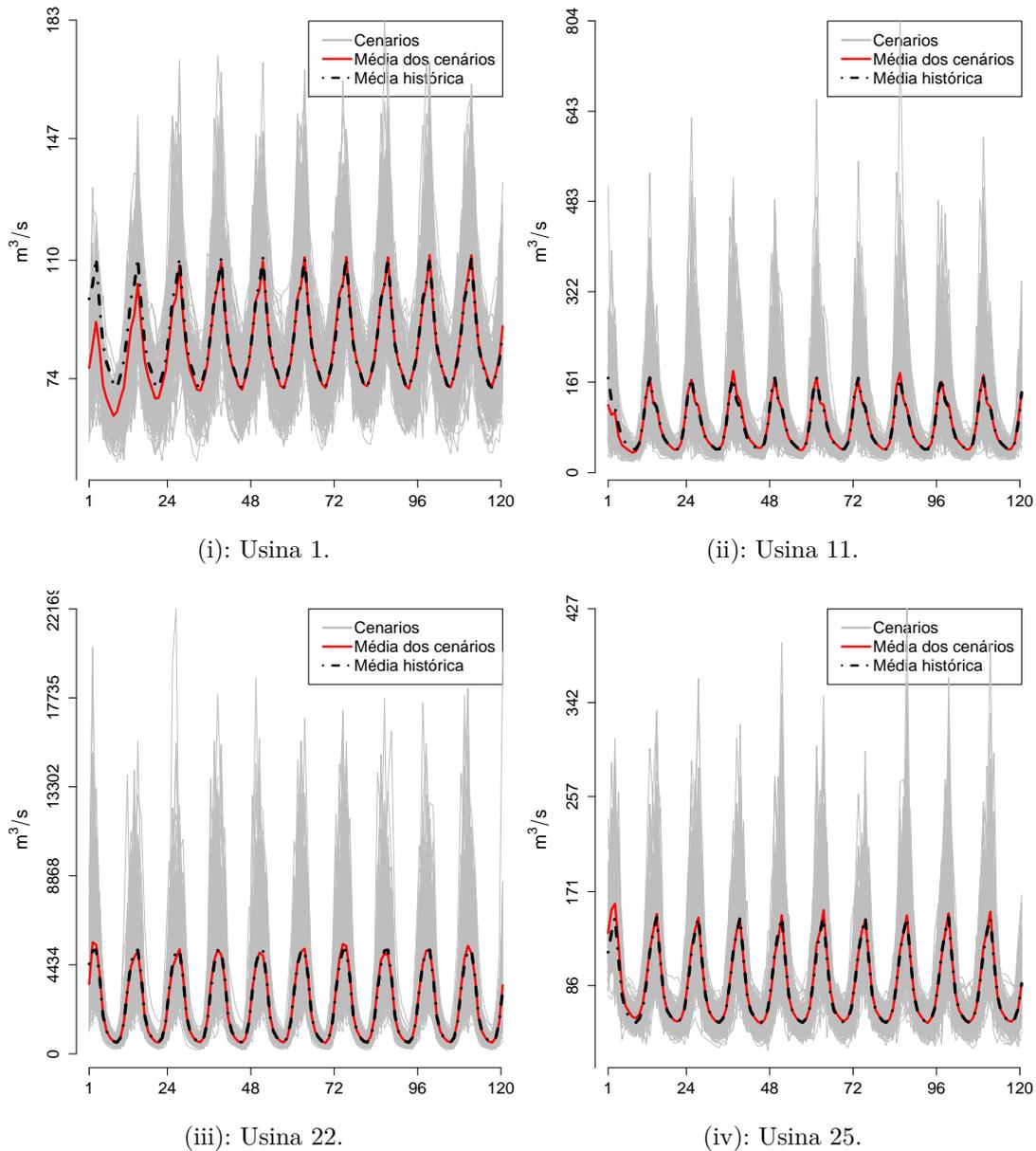


Figura 4.35: Modelo alternativo: cenários para as usinas 1, 11, 22 e 25.

Tabela 4.12: Modelo alternativo: análise dos cenários.

	Taxa de aprovação				Taxa de aprovação		
	t	Levene	K-S		t	Levene	K-S
Usina 1	81%	91%	99%	Usina 21	100%	99%	100%
Usina 2	100%	97%	89%	Usina 22	100%	100%	100%
Usina 3	99%	97%	92%	Usina 23	100%	100%	98%
Usina 4	100%	100%	97%	Usina 24	100%	100%	100%
Usina 5	99%	100%	98%	Usina 25	92%	100%	91%
Usina 6	99%	100%	100%	Usina 26	100%	99%	98%
Usina 7	99%	100%	97%	Usina 27	49%	77%	28%
Usina 8	99%	100%	98%	Usina 28	100%	71%	59%
Usina 9	99%	100%	99%	Usina 29	100%	99%	100%
Usina 10	92%	96%	91%	Usina 30	97%	81%	88%
Usina 11	92%	96%	93%	Usina 31	98%	99%	98%
Usina 12	93%	96%	94%	Usina 32	99%	84%	90%
Usina 13	93%	94%	90%	Usina 33	100%	100%	100%
Usina 14	93%	94%	92%	Usina 34	100%	95%	100%
Usina 15	99%	100%	97%	Usina 35	97%	99%	95%
Usina 16	99%	100%	96%	Usina 36	91%	100%	95%
Usina 17	93%	95%	92%	Usina 37	100%	99%	99%
Usina 18	100%	99%	100%	Usina 38	93%	96%	93%
Usina 19	100%	99%	100%	Usina 39	93%	96%	94%
Usina 20	100%	99%	100%				

4.7

Conclusões do capítulo

A simulação multivariada de cenários hidrológicos foi abordada neste capítulo. Neste contexto, tanto a dependência temporal quanto a dependência espacial devem ser representadas pelo modelo. Adiciona-se a isto, no caso de séries periódicas, uma possível variação sazonal das estruturas probabilísticas. Particularmente no caso do SEB, tais modelos ainda devem ser capazes de lidar com o elevado número de usinas existentes.

Para esta tarefa, propusemos um modelo periódico de cópula vine espacial que incorporou todas as características supracitadas. Seguimos a abordagem padrão para os casos onde os dados apresentam dependência serial. Inicialmente estimou-se o modelo de séries temporais PAR(p) para as vazões e posteriormente, a dependência entre usinas através de cópulas.

Optamos por utilizar uma cópula vine para lidar com a alta dimensionalidade do problema e ao mesmo tempo, mantermos toda a flexibilidade das cópulas bivariadas. A abordagem espacial permitiu ainda que informações geográficas fossem exploradas de modo à reparametrizar e assim reduzir o número de parâmetros do modelo. Como variáveis espaciais empregamos a distância entre usinas, séries de precipitações e informações acerca da topografia dos rios. Por fim, para tratar adequadamente a variação sazonal de dependência espacial, propusemos uma extensão ao modelo original de Erhardt et al. [23]. O modelo resultante pode ser entendido como um modelo de cópulas vine espaciais cujos parâmetros variam periodicamente.

A análise do modelo proposto ocorreu em duas oportunidades distintas. Na primeira, através de simulações de Monte Carlo, comparamos a dependência espacial implicada pelo modelo com a dependência espacial histórica. Os resultados indicaram que a reparametrização sugerida foi capaz de aproximar a dependência entre usinas. Ademais, a extensão proposta para o caso periódico se mostrou razoável.

Na segunda parte, analisou-se os cenários de vazões simulados. Uma inspeção visual indicou que o modelo é capaz de replicar as médias históricas e reproduzir apropriadamente o forte padrão periódico das séries de vazões históricas. Por fim, testes estatísticos indicaram que o modelo proposto é capaz de simular cenários que preservam características observadas nas vazões originais.

Finalizando, testamos algumas configurações diferentes para o modelo. No geral, tais modificações visavam uma redução da complexidade do modelo. Aplicamos a transformação logarítmica nas séries de vazões originais, utilizamos a distribuição Gaussiana para os resíduos do modelo PAR(p) e reduzimos o número de variáveis espaciais. Quando comparado ao modelo inicial, os resultados não indicaram quaisquer diferenças relevantes entre as abordagens.

5

Cópuas periódicas para simulação de cenários hidrológicos univariados

Propõe-se neste capítulo uma metodologia baseada nos modelos de cópuas vine para simulação estocástica de séries univariadas de vazões.

O modelo resultante, pode ser entendido como um modelo autorregressivo periódico (não-)linear de ordem p e, é uma alternativa ao modelo PAR(p). Os trabalhos realizados no âmbito da simulação hidrológica univariada, ([9,11,16–18,21]), consideraram apenas a dependência temporal de lag 1. Assim, propomos uma metodologia, baseada em cópuas vine, capaz de modelar séries de vazões que possuem auto-dependência maiores do que um. Além do mais, a abordagem periódica adotada permite que a persistência temporal varie de acordo com o período. Isto faz com que o forte componente sazonal comumente presente em tais séries seja modelado adequadamente.

5.1

Motivação

Em se tratando de modelos univariados para simulação estocástica de séries hidrológicas, os modelos lineares pertencentes a família de Box & Jenkins são extensivamente empregados. [3–5, 39].

Geralmente tais modelos assumem a hipótese de normalidade, muitas das vezes não é observada empiricamente. [9,11]. Ademais, para a simulação de cenários via Monte Carlo, a normalidade dos resíduos produz valores simulados entre $(-\infty, \infty)$, o que não garante que todos os cenários sejam estritamente positivos. Uma alternativa seria a utilização de transformações na série original. Entretanto, de acordo com Hao & Sing [10] e Li et al. [11], estas podem incorporar algum tipo de viés nos cenários. Por fim, salienta-se que tais modelos não são capazes de capturar associações não lineares presentes em séries de vazões.

Recentemente, com o advento das cópuas, uma nova via para simulação estocástica de cenários de vazões surgiu. O número de pesquisas que utilizam cópuas para simulação de cenários vem crescendo ao longo dos anos, entretanto, seu potencial ainda não foi completamente investigado.

É de conhecimento comum que cópulas podem modelar a dependência temporal de uma série univariada. [24, 55–57].

Aspectos teóricos a respeito do uso de cópulas para dependência serial foram abordados em [56–61]. Em finanças, Mendes & Aíube [25] empregaram cópulas para modelar o primeiro e o segundo momento de séries univariadas de retornos. Mendes & Accioly [26] utilizaram pair-cópulas para a previsão de séries de volatilidade realizada do mercado brasileiro. No caso de séries temporais multivariadas, Smith [62] propôs o uso de cópulas para capturar a dependência serial e a interdependência entre séries temporais. Brechmann & Czado [27] propuseram o modelo COPAR (*Copula autoregressive model*), o que é uma alternativa ao modelo VAR (Vetor autorregressivo). Smith et al. [63] expressaram a distribuição conjunta de uma série temporal em termos de pair-cópulas. Além disso, os autores propuseram um método Bayesiano para a estimação de cópulas para dados longitudinais.

Joe [24] representa alguns modelos de séries temporais estacionários através de cópulas. Estas formulações estendem inúmeros modelos Gaussianos autorregressivos, tais como $AR(p)$, $MA(q)$ e $ARMA(p, q)$. Em particular, uma série temporal Markoviana de ordem p representada via cópula¹ pode ser entendida como um processo autorregressivo não-linear. A forma geral de uma série temporal Markoviana de ordem 1 é $y_t = g(\varepsilon_t; y_{t-1})$, onde $\{\varepsilon_t\}$ é uma sequência de choques aleatórios independentes e identicamente distribuídos e $\{y_t\}$ é uma sequência de uma variável aleatória observada. Neste caso, $\{\varepsilon_i\}$ é independente de y_j para $j < i$. Para $p \geq 2$, a forma geral é $y_t = g(\varepsilon_t; y_{t-p}, \dots, y_{t-1})$. Isso significa que y_t depende do passado através das últimas p observações.

Se Y_t é uma variável aleatória contínua cuja a função de distribuição acumulada é F_Y , pode-se assumir que $u_t = F_Y(y_t)$ com $u_t \sim U(0, 1)$. Quando $p = 1$, têm-se a cópula entre U_{t-1} e U_t . Caso p seja igual a 2, têm-se a cópula $C_{t-2, t-1, t}(u_{t-2}, u_{t-1}, u_t)$, onde as cópulas marginais $C_{t-2, t-1}$ devem ser idênticas a $C_{t-1, t}$. Para defasagens maiores do que 2 ($p > 2$), a cópula é $C_{t-p, \dots, t}(u_{t-p}, \dots, u_t)$, onde as cópulas marginais correspondentes ao mesmo *lag* temporal devem ser iguais. Tais restrições são trivialmente satisfeitas por uma D-vine. [24].

Em uma D-vine, cada vértice na T_1 possui grau menor ou igual a 2. Na Figura 5.1, é possível observar uma D-vine com 4 variáveis representando um processo Markoviano de ordem 3.

¹Neste caso a cópula não é uma cópula Gaussiana. Caso as distribuições marginais sejam normais e a cópula seja Gaussiana, o modelo resultante é o modelo linear $AR(p)$. [56, 63].

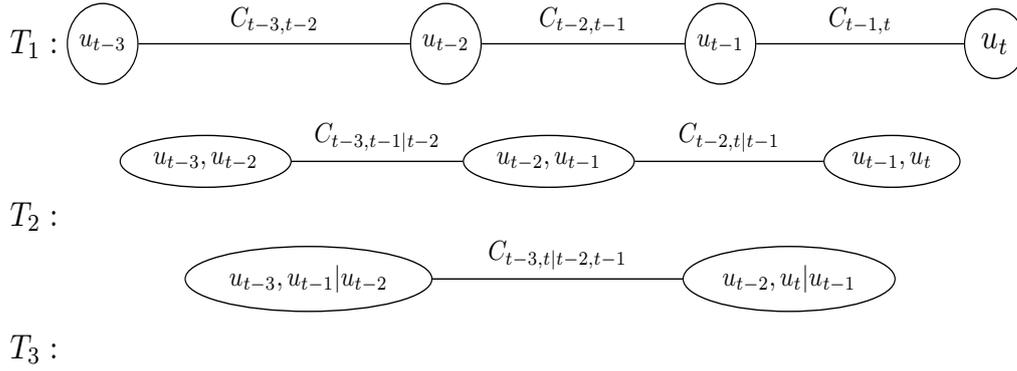


Figura 5.1: Exemplo de uma D-vine com 4 variáveis.

Um exemplo de como uma série temporal Markoviana de ordem $p = 2$ é decomposta em termos de cópulas bivariadas é apresentado. Para isto, considere a série temporal univariada $\mathbf{Y} = \{Y_1, \dots, Y_T\}$. A função de densidade conjunta de \mathbf{Y} é $f(y_T, \dots, y_1)$ e, pode ser decomposta como

$$f(y_T, \dots, y_1) = f_1(y_1) \cdot \prod_{t=2}^T f(y_t | y_{t-1}, \dots, y_1). \quad (5-1)$$

Assumindo um processo Markoviano de ordem 2, esta densidade passa a ser reescrita como

$$f(y_T, \dots, y_1) = f_1(y_1) \cdot \prod_{t=2}^T f(y_t | y_{t-1}, y_{t-2}). \quad (5-2)$$

De maneira similar a decomposição apresentada na Subseção 2.1.2, temos

$$f(y_t | y_{t-1}, y_{t-2}) = c_{t,t-2|t-1}(u_{t|t-1}, u_{t-2|t-1}) \cdot c_{t,t-1}(u_t, u_{t-1}) \cdot f_t(y_t), \quad (5-3)$$

onde $u_{t|t-1} = F_{t|t-1}(y_t | y_{t-1})$ e $u_{t-2|t-1} = F_{t-2|t-1}(y_{t-2} | y_{t-1})$. Inserindo a equação (5-3) na equação (5-2), obtêm-se

$$f(y_T, \dots, y_1) = f_1(y_1) \cdot \prod_{t=2}^T c_{t,t-2|t-1}(u_{t|t-1}, u_{t-2|t-1}) \cdot c_{t,t-1}(u_t, u_{t-1}) \cdot f_t(y_t). \quad (5-4)$$

Por exemplo, se as distribuições marginais são Gaussianas e todas as pair-cópulas são Gaussianas bivariadas, então a equação (5-4) representa um processo estacionário Gaussiano autorregressivo - AR(2). [63].

O uso de cópulas para a simulação de cenários de vazões univariadas foi inicialmente estudado por Lee & Salas [17]. Eles citam alguns modelos comumente empregados em hidrologia e suas limitações. Segundo os autores, a característica de que as cópulas acomodam qualquer estrutura de dependência é a razão pela qual estas devem ser estudadas a fundo para simulação hidrológica. Em seu trabalho, foram analisadas apenas cópulas bivariadas (Gumbel, Gaussiana, Clayton e Frank). Um estudo de Monte Carlo foi feito

para verificar as características de cada cópula em termos de simulação. Por fim, foi realizado um estudo de caso considerando a série anual de vazão do rio Nilo.

Hao & Sing [18] propuseram um método utilizando cópulas e entropia para a simulação de cenários de vazões mensais. A estimação das distribuições marginais foi realizada através do método de entropia enquanto que as distribuições conjuntas foram representadas por cópulas. Em termos de estrutura de dependência temporal, os autores se concentraram no caso bivariado, ou seja, modelaram apenas a dependência entre os meses t e $t - 1$. Para tal, utilizaram cópulas bivariadas onde para cada mês, uma cópula foi estimada. Uma versão do modelo inserindo uma variável exógena (variável agregada - soma dos últimos 12 instantes de tempo) foi proposta. Neste caso, cópulas tridimensionais foram empregadas.

Zachariah & Reddy [9] estimaram um modelo utilizando os princípios de cópula e entropia para a simulação de cenários de vazões mensais da barragem de Hirakud, Índia. As distribuições marginais foram obtidas via entropia enquanto que a dependência entre dois meses subsequentes foi representada pela cópula Gumbel-Hougaard bivariada. De maneira similar, Kong et al. [21] estimaram um modelo baseado em cópulas e entropia para a simulação de cenários mensais do rio Xiangxi na China.

Li et al. [11] estimaram a distribuição conjunta entre dois meses subsequentes condicionadas a outras variáveis tais como variáveis agregadas, variáveis climáticas, etc. A estrutura de dependência entre os meses foi capturada por cópulas bivariadas. Para a estimação do modelo, os parâmetros da função de densidade conjunta foram aproximados por redes neurais. Os autores concluem que o modelo proposto é capaz de reproduzir nos cenários as estatísticas observadas no histórico e a auto-dependência de lag 1. Entretanto, os autores destacam a limitação do modelo em reproduzir adequadamente auto-dependência de ordens superiores.

Jeong & Lee [20] utilizaram a mesma abordagem proposta por Lee & Salas [17] em conjunto com o modelo PDAR (periodic discrete autoregressive). Em seu artigo, os autores utilizaram cópulas para modelar a estrutura de dependência temporal de séries de vazões mensais. Para lidar com a forte sazonalidade, eles adotaram um procedimento periódico onde para cada mês, uma cópula foi estimada. Novamente, os autores se limitaram a analisar apenas cópulas bivariadas.

Como pode ser observado, ao nosso conhecimento, os trabalhos de simulação de séries hidrológicas univariadas ficaram restritos ao caso das cópulas bivariadas, ou seja, se preocuparam exclusivamente com a auto-

dependência de lag 1. Em alguns casos, houve uma extensão para o caso tridimensional, entretanto, a terceira variável acrescentada sempre era uma variável exógena. Com isto em mente, o objetivo principal desse capítulo é propor uma metodologia baseada em cópulas vine para a simulação de séries hidrológicas considerando estruturas de dependência superiores a 1 e o forte componente periódico.

5.2

Modelo periódico de cópula vine para séries de vazões

A metodologia proposta para simulação de cenários hidrológicos periódicos univariados é apresentada nesta seção. Considere y_t , $t = 1, \dots, T$ como uma série temporal de vazão mensal com s períodos (número de intervalos dentro de um ano). Aqui, n representa o número de anos, ou seja, $T/s = n$. Isto posto, pode-se reescrever o índice temporal t como $t = t(r, m) = (r - 1)s + m$, onde $r = 1, \dots, n$ e $m = 1, \dots, s$. Neste caso específico de séries mensais, r denota os anos enquanto que m representa os meses.

Além do mais, $u_{t(r,m)} = F^m(y_{t(r,m)})$ e $u_{t(r,m)-i} = F^{m-i}(y_{t(r,m)-i})$. Desta forma, a distribuição multivariada do período m com dimensão d é dada por

$$F^m(y_{t(r,m)-d+1}, \dots, y_{t(r,m)-1}, y_{t(r,m)}) = C^m(u_{t(r,m)-d+1}, \dots, u_{t(r,m)-1}, u_{t(r,m)}) \quad (5-5)$$

A ideia principal da metodologia proposta é utilizar para cada período m uma cópula d -dimensional (cópula vine), permitindo que tanto a dimensão quanto as cópulas variem de acordo com o período. O modelo resultante possui estrita conexão com o modelo PAR(p) e, pode ser entendido como uma versão (não-)linear do mesmo.

5.2.1

Simulação de cenários

Como discutido na Seção 3.4, a simulação de novas observações a partir de uma função de distribuição d -dimensional \mathbf{F} pode ser realizada por meio do procedimento de transformação inversa (*inverse transformation procedure*). De maneira geral, inicialmente amostra-se $w_i \sim U[0, 1]$, $i = 1, \dots, d$ e posteriormente itera-se $u_1 := w_1$, $u_2 := F_{2|1}^{-1}(w_2|u_1)$, $u_3 := F_{3|1,2}^{-1}(w_3|u_1, u_2)$ até $u_d := F_{d|1,\dots,d-1}^{-1}(w_d|u_{d-1}, \dots, u_1)$. Vale a pena lembrar que qualquer função de distribuição condicional pode ser expressa em termos de suas cópulas. Assim sendo, $u_1 := w_1$, $u_2 := C_{2|1}^{-1}(w_2|u_1)$, $u_3 := C_{3|1,2}^{-1}(w_3|u_1, u_2), \dots, u_d := C_{d|1,\dots,d-1}^{-1}(w_d|u_1, \dots, u_{d-1})$.

Observe, entretanto, que no caso da simulação de cenários de vazões (séries temporais), a única variável na qual estamos realmente interessados é

a simulação de u_t condicionada as $d - 1$ observações passadas. Assumindo que t pertence a um determinado período m , têm-se

$$u_t := C_{t|t-1,t-2,\dots,t-d+1}^{-1}(w_t|u_{t-1}, u_{t-2}, \dots, u_{t-d+1}). \quad (5-6)$$

A cópula condicional inversa ($C_{t|t-1,t-2,\dots,t-d+1}^{-1}$ - equação (5-6)), pode ser escrita em termos de *h-functions* e suas respectivas funções inversas (*inverse of h-functions*, h^{-1}). Vale lembrar que a variável simulada está na escala uniforme; $u_t \sim U(0, 1)$. Assim, esta deve ser transformada para a escala original (m^3/s) empregando a distribuição marginal adequada. Mais detalhes sobre simulação de cópulas podem ser encontrados nas referências [13, 28, 46].

5.3 Estudo de caso

A metodologia proposta foi testada na série de vazão mensal (m^3/s) da usina hidrelétrica de Manso². Os dados foram obtidos junto ao Operador Nacional do Sistema (ONS) com início em janeiro de 1931 e término em dezembro de 2012 totalizando 82 anos completos.

A Figura 5.2 apresenta a série de vazão do rio Manso (esquerda). Do lado direito, pode-se observar a mesma série organizada de acordo com os períodos. Como pode ser observado, é nítido o forte componente sazonal nesta série. Durante os meses próximos ao verão (período chuvoso), tanto a média quanto a variância são superiores aos meses de inverno (período seco).

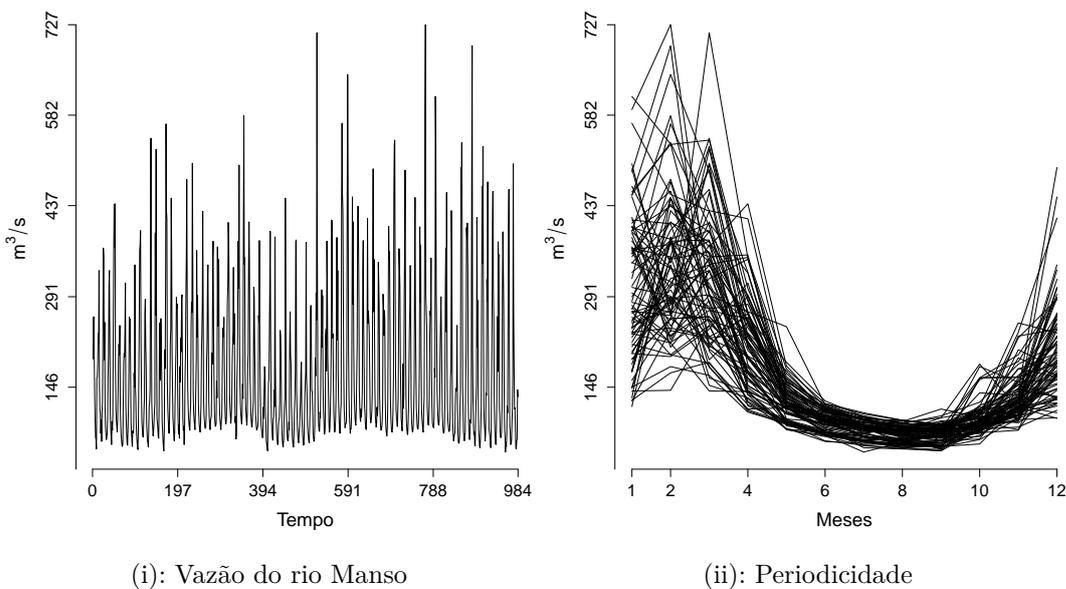


Figura 5.2: Série de vazão do rio Manso.

²Esta série temporal também foi empregada no estudo de caso do Capítulo 4.

Antes da estimação do modelo, vale a pena ressaltar que o primeiro passo é transformar as variáveis originais (série temporal, nesse caso específico) em dados cópula (dados que sejam uniformemente distribuídos).

Para esta transformação, após uma análise inicial, utilizou-se a distribuição gama como distribuição marginal. Para cada mês uma distribuição gama foi estimada. Sua função densidade de probabilidade é dada por

$$f(y|\alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} y^{\alpha-1} e^{-y/\beta} \quad (5-7)$$

onde $y \geq 0$, $\alpha > 0$ é o parâmetro de forma e $\beta > 0$ é o parâmetro de escala.

Os dados cópula são obtidos através da transformada integral de probabilidade,

$$\hat{u}_{t(r,m)} = F^m(y_{t(r,m)}|\hat{\alpha}^m, \hat{\beta}^m) \quad (5-8)$$

onde F^m é a função de distribuição acumulada gama no período m . Vale destacar a versatilidade para se modelar as distribuições marginais. Embora tenhamos optado por uma distribuição gama para cada mês, poderíamos ter empregado diferentes famílias para cada mês. Em outras palavras, não existe a obrigatoriedade de que as distribuições marginais mensais sejam as mesmas.

Após a transformação, é crucial verificar se os dados cópula seguem uma uniforme padrão. Esta análise pode ser realizada observando-se os histogramas das variáveis ou através de um teste estatístico (teste de Anderson-Darling, por exemplo.). Ambos foram realizados e os resultados indicam que a distribuição gama é uma escolha razoável para as distribuições marginais mensais. A Figura 5.3 apresenta os p-valores para cada um dos 12 testes realizados. A linha preta representa o nível de significância de 5%.

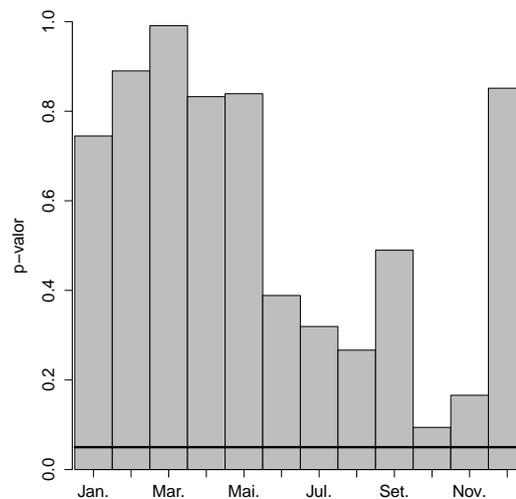


Figura 5.3: Teste de Anderson-Darling para cada um dos 12 meses.

Realizada a estimação das distribuições marginais, o próximo passo é

estimar as cópulas vine mensais. Neste momento surge um questionamento sobre como determinar a dimensão mais adequada para cada mês. Nesta tese permitiu-se que a cópula vine periódica assumisse dimensões entre 2 e 4, ou seja, $d = 2$, $d = 3$ ou $d = 4$. Isto significa que modelos (não-)lineares autorregressivos com ordem iguais a 1, 2 ou 3 serão empregados.

Para definir as dimensões, sugerimos um procedimento iterativo em conjunto com o teste de independência assintótico para cópulas bivariadas. Para mais detalhes sobre o teste, veja [53]. Para cada período, inicia-se com uma cópula bivariada. Na sequência, estima-se uma cópula tri-dimensional e verifica-se se a cópula condicional $C_{t-2,t|t-1}$ é uma cópula produto utilizando o teste de 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 a 4. Por outro lado, se $C_{t-3,t|t-1,t-2}$ é uma cópula de independência, reduzimos a dimensão para 3.

A ideia por trás dessa abordagem é verificar a última árvore da cópula vine. Isto equivale a testar se a cópula condicional $C_{t-d+1,t|t-1,\dots,t-d+2}$ é uma cópula de independência ou não. Em outras palavras, estamos analisando 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 serial.

Com relação as cópulas bivariadas, empregou-se um conjunto formado pela cópulas de independência (I), Gaussiana (N), t-Student (S-t), Gumbel (G), Clayton (C) e Frank (F).

A seleção das cópulas bivariadas foi realizada via BIC (Bayesian Information Criterion). A Tabela 5.1 apresenta as cópulas bivariadas selecionadas. Como pode ser observado, a cópula Clayton, que possui dependência de cauda inferior, ocorre com maior frequência. Em termos práticos, isto indica que vazões baixas são seguidas por vazões baixas. Existe um baixo número de cópulas Gaussiana e t-Student. O elevado número de cópulas Arquimedianas indica a presença de estruturas assimétricas de dependência.

A Tabela 5.2 apresenta as ordens selecionadas para os meses. As cópulas bivariadas são predominantes. Entretanto, destaca-se que as maiores dimensões ocorrem entre maio e agosto, o que coincide com o período do inverno (período de seca) no Brasil.

Tabela 5.1: Número de cópulas selecionadas.

I	N	S-t	C	G	F
2	3	2	7	3	2

Tabela 5.2: Dimensões selecionadas para cada mês.

	Jan.	Fev.	Mar.	Abr.	Mai.	Jun.	Jul.	Ago.	Set.	Out.	Nov.	Dez.
Dim.	2	2	3	2	4	4	2	3	2	2	2	2

5.3.1 Análise do modelo estimado

Antes das simulações dos cenários, é interessante que se averigüe algumas características do modelo estimado em termos de ajuste.

A Figura 5.4 apresenta o comparativo entre a série real e a série ajustada pelo modelo (previsão um passo à frente). Visualmente, pode-se perceber que a série ajustada (linha vermelha) acompanha o comportamento da série histórica.

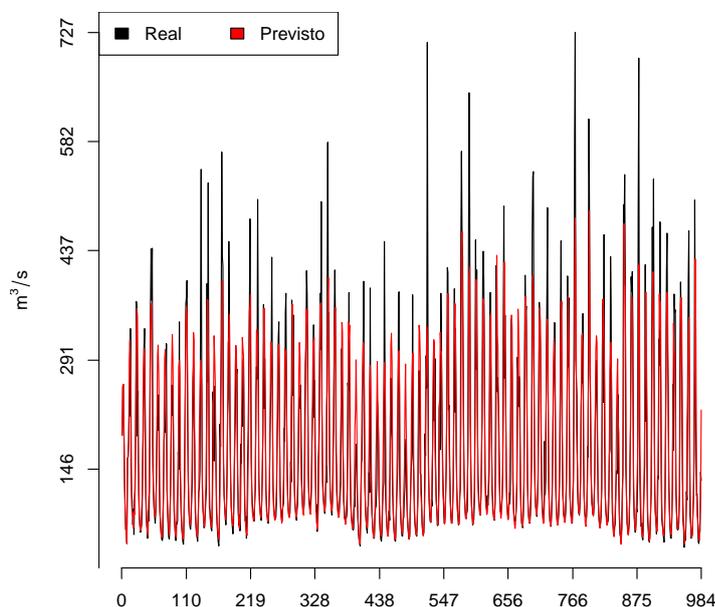


Figura 5.4: Vazão do rio Manso e vazão ajustada pelo modelo.

Vale lembrar que a série ajustada não é obtida analiticamente como no caso do modelo autorregressivo Gaussiano. Tal série deve ser obtida via simulação. No total, para cada previsão um passo à frente, foram simuladas

5000 observações condicionadas aos valores históricos da série original. A previsão é obtida ao se calcular a média dessas simulações. O MAPE (*Mean absolute percentage error*) obtido foi igual a 18.24%. A título de comparação, estimamos um modelo PAR(p) cujo MAPE foi de 18,76%.

Por fim, apresentamos a função de auto-dependência dos resíduos com base no τ de Kendall. Como pode ser observado na Figura 5.5, o modelo proposto foi capaz de capturar toda a estrutura de dependência existente na série analisada. As linhas pontilhadas foram obtidas a partir do teste de independência baseado no τ de Kendall. Sob a hipótese nula de que o τ de Kendall é igual a zero, a estatística de teste segue uma distribuição normal com média zero e variância $2(2n + 5)/9n(n - 1)$ onde n é o número de pares. Considerando um nível de significância de 5%, os intervalos podem ser aproximados por $\pm 1.96(\sqrt{2(2n + 5)/9n(n - 1)})$. Para mais detalhes, veja [53].

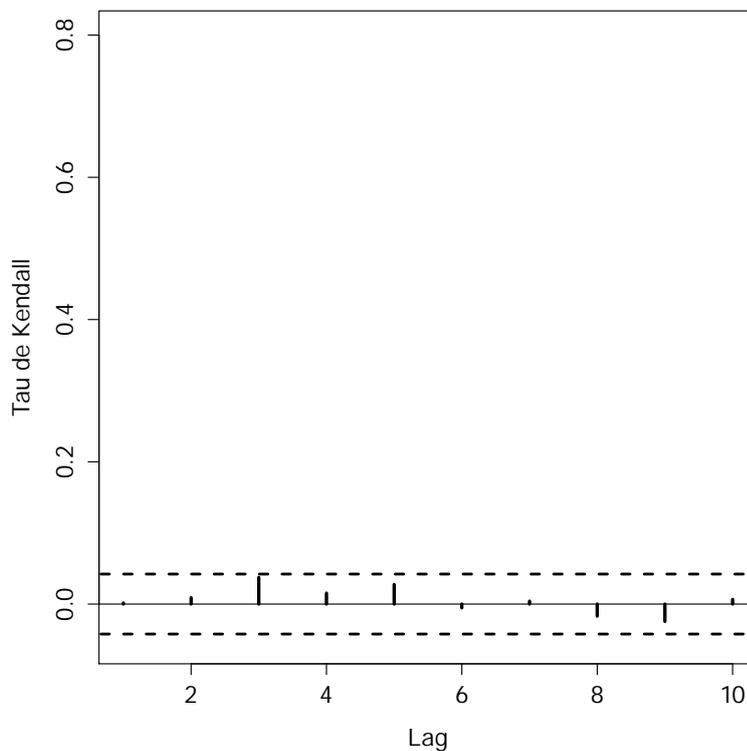


Figura 5.5: Auto-dependência dos resíduos.

Adicionalmente, a Figura 5.6 apresenta as auto-dependências dos resíduos para cada um dos 12 meses (modelos) estimados. Como pode ser observado, os resíduos mensais não apresentam dependência temporal.

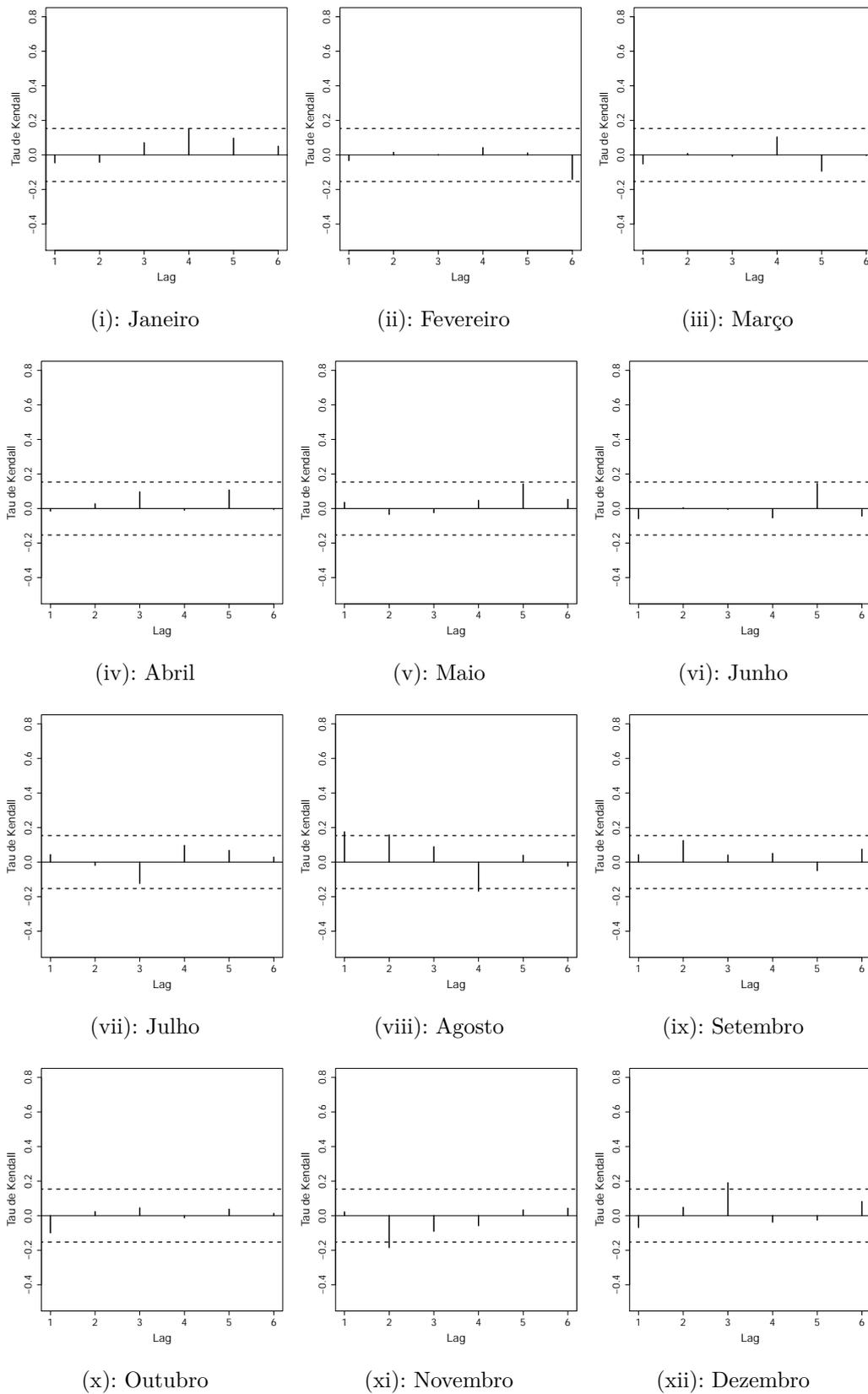


Figura 5.6: Auto-dependência dos resíduos de cada mês.

5.3.2

Simulação dos cenários

Os cenários simulados obtidos através do modelo desenvolvido são apresentados a seguir. No total, simulou-se 200 cenários contendo 60 meses, totalizando 5 anos de séries sintéticas. Para inicializar o procedimento, utilizou-se o último ano disponível da série histórica. Todos os códigos foram implementados em R com auxílio do pacote VineCopula [52].

A Figura 5.7 apresenta todos os 200 cenários simulados (linhas cinza) assim como suas médias periódicas (linha azul) e a média mensal histórica (linha vermelha). Através de uma análise visual, a metodologia proposta consegue representar a forte periodicidade que existe na série do rio Manso. Além disso, podemos observar claramente que a média dos cenários praticamente se sobrepõe a média histórica.

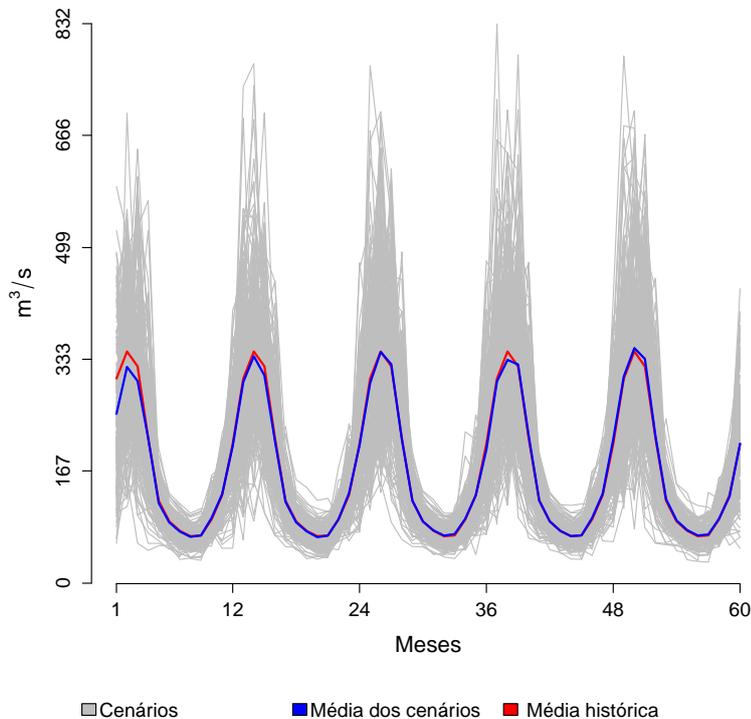


Figura 5.7: Cenários simulados.

Para analisar os cenários simulados, alguns testes estatísticos foram empregados. As variáveis de interesse foram 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, a variância e a distribuição do primeiro mês simulado era igual a média, a variância e a distribuição de janeiro. Esta análise é repetida para cada um dos 60 meses simulados.

Para tais análises, o teste t, o teste de Levene e o de Kolmogorov–Smirnov foram utilizados. Com relação ao último, este é um teste não-paramétrico que compara se duas amostras são oriundas da mesma distribuição ou não.

A Figura 5.8 resume os três testes estatísticos. 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 indicam que a metodologia proposta é capaz de simular cenários que replicam as características originais observadas na série temporal. As taxas de aprovação foram 99%, 99% e 94% para o teste-t, o teste de Levene e o teste K-S respectivamente.

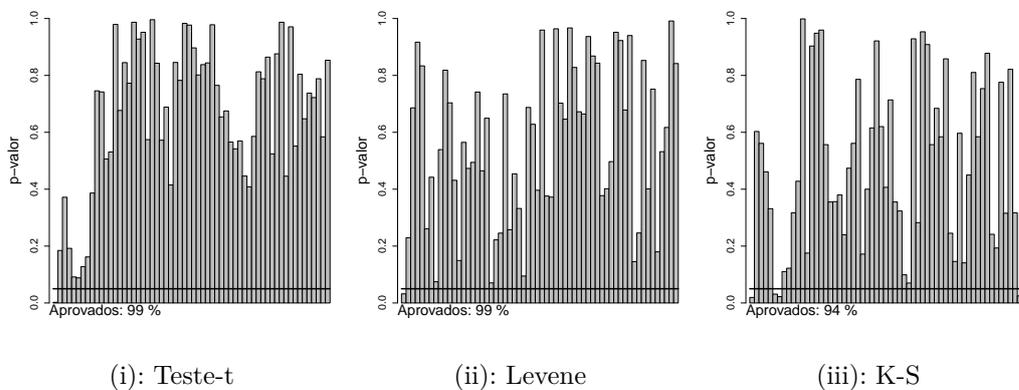


Figura 5.8: Análise dos cenários simulados.

Por fim, realizamos uma análise da estrutura de dependência dos cenários simulados. O objetivo é averiguar se estes preservam a dependência serial observada na série histórica. Simulamos $SCE = 200$ cenários com 100 anos cada. Para cada cenário $sce = 1, \dots, SCE$, calculamos os *lags* de cada modelo mensal e os armazenamos. Por fim, computamos os *lags* médios dessas simulações. As Figuras 5.9, 5.10 e a Tabela 5.3 trazem os resultados dessa análise.

A Figura 5.9 apresenta o *lags* 1, 2, 3, 4 e 5 (real e simulado) para cada um dos 12 modelos mensais. Como pode ser observado, para praticamente todos os meses e *lags*, a média simulada é próxima dos valores históricos. Isto é corroborado pelos respectivos valores apresentados na Tabela 5.3.

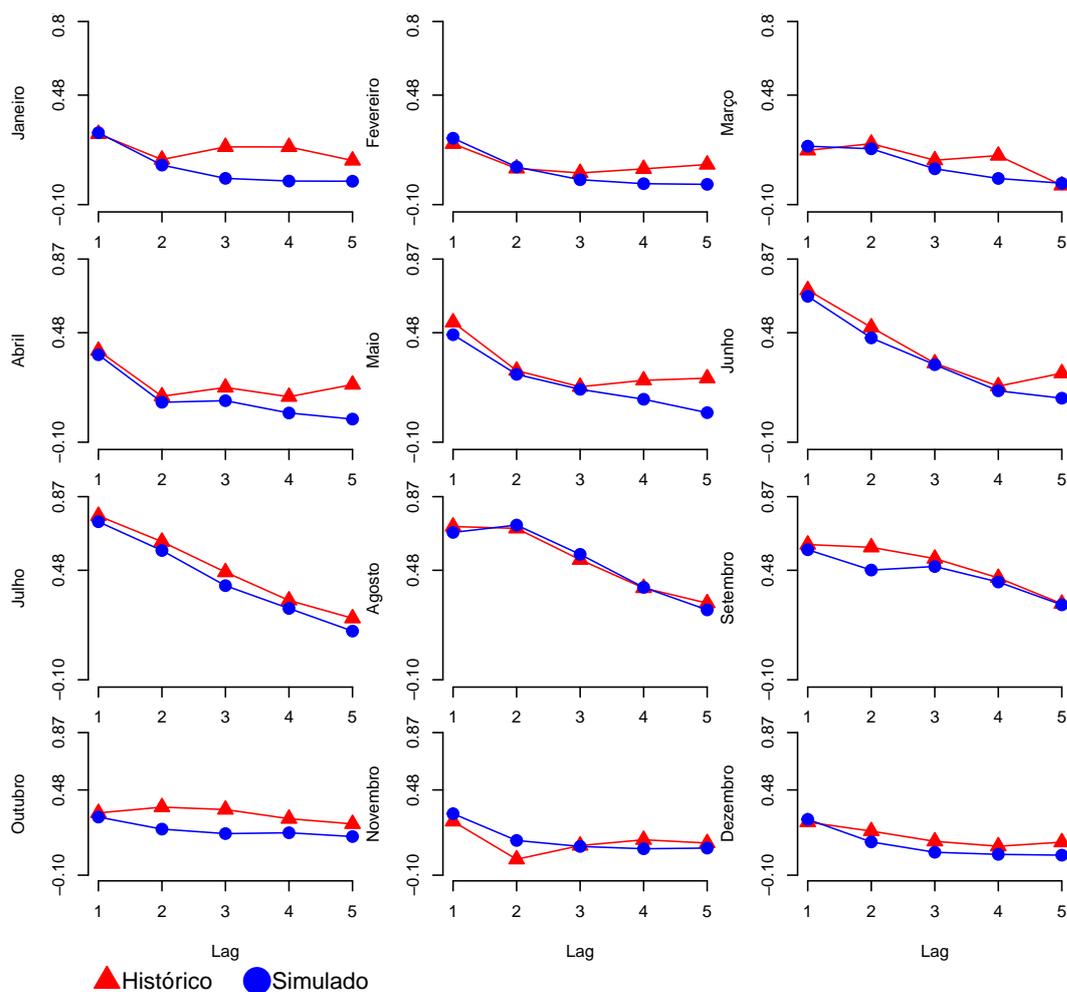


Figura 5.9: Comparação da auto-dependência de lags 1, 2, 3, 4 e 5.

Tabela 5.3: Auto-dependência mensal: valores históricos e simulados.

		Lag 1	Lag 2	Lag 3	Lag 4	Lag 5
Jan.	Histórico	0,2746	0,1385	0,2056	0,2055	0,1335
	Simulado	0,2805	0,1092	0,0390	0,0248	0,0238
Fev.	Histórico	0,2219	0,0916	0,0677	0,0892	0,1121
	Simulado	0,2516	0,0992	0,0319	0,0107	0,0072
Mar.	Histórico	0,1871	0,2227	0,1349	0,1596	0,0018
	Simulado	0,2095	0,1955	0,0896	0,0385	0,0131
Abr.	Histórico	0,3847	0,1432	0,1898	0,1407	0,2052
	Simulado	0,3631	0,1114	0,1197	0,0546	0,0223
Mai.	Histórico	0,5350	0,2791	0,1935	0,2277	0,2388
	Simulado	0,4694	0,2593	0,1799	0,1273	0,0563
Jun.	Histórico	0,7055	0,5085	0,3180	0,1966	0,2654
	Simulado	0,6726	0,4522	0,3105	0,1721	0,1329
Jul.	Histórico	0,7683	0,6315	0,4702	0,3204	0,2257
	Simulado	0,7369	0,5846	0,3984	0,2778	0,1577
Ago.	Histórico	0,7116	0,7018	0,5358	0,3861	0,3059
	Simulado	0,6805	0,7194	0,5643	0,3893	0,2700
Set.	Histórico	0,6162	0,6015	0,5418	0,4398	0,3025
	Simulado	0,5888	0,4811	0,5000	0,4177	0,2968
Out.	Histórico	0,3234	0,3629	0,3463	0,2842	0,2487
	Simulado	0,2953	0,2132	0,1826	0,1884	0,1627
Nov.	Histórico	0,2659	0,0079	0,1014	0,1406	0,1185
	Simulado	0,3182	0,1360	0,0954	0,0796	0,0842
Dez.	Histórico	0,2607	0,1998	0,1301	0,0965	0,1243
	Simulado	0,2797	0,1259	0,0557	0,0414	0,0360

A Figura 5.10 apresenta diversos *box-plots* onde é possível obtermos uma ideia a respeito da variabilidade dos valores simulados. No geral, a maioria dos valores históricos estão dentro das caixas o que indica que o modelo é capaz de simular cenários que respeitam as estruturas de dependência temporal observadas na série histórica.

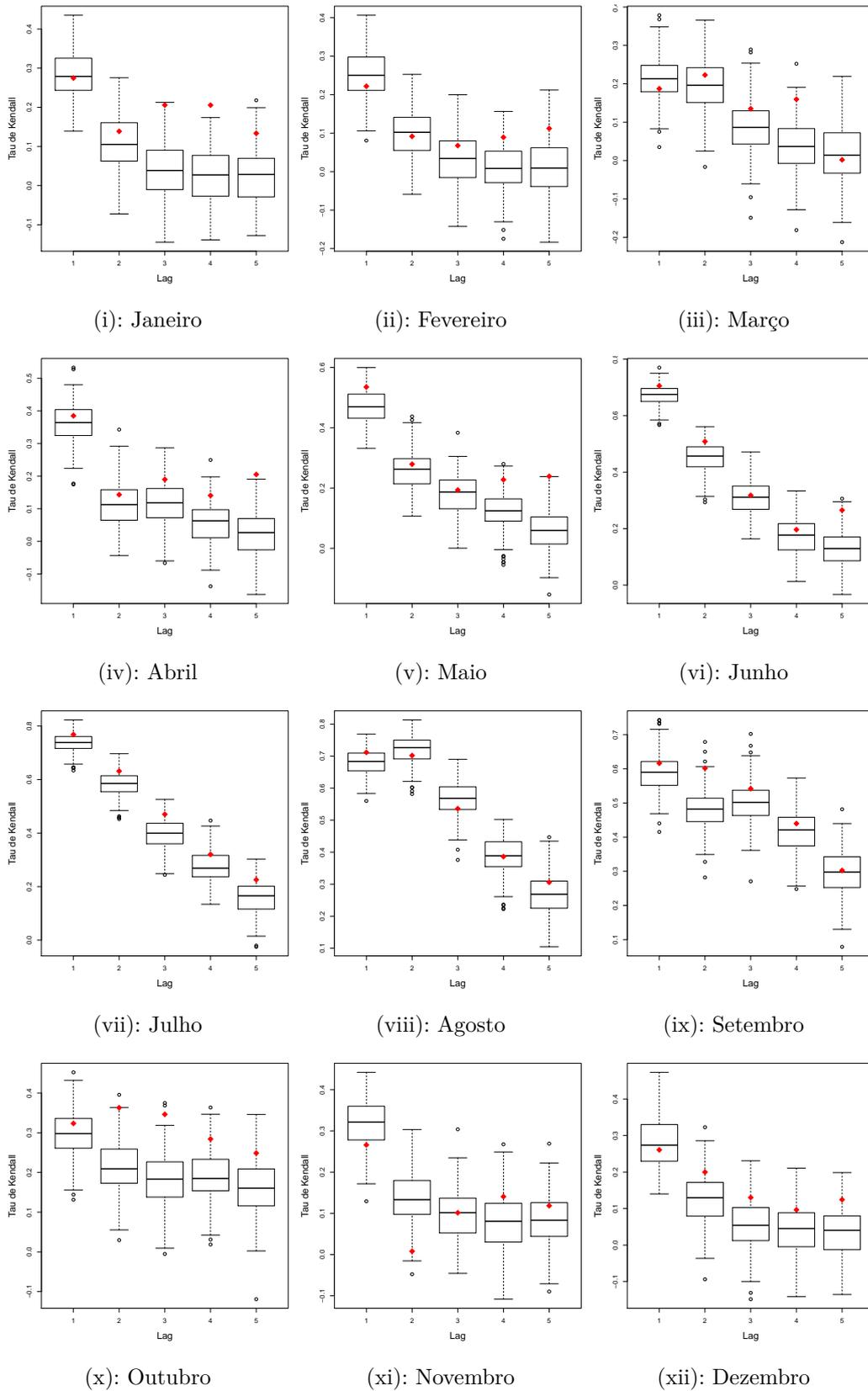


Figura 5.10: Box-plot para auto-dependência de lags 1, 2, 3, 4 e 5.

Conclusões do capítulo

Neste capítulo foi proposto uma nova metodologia para simulação de cenários univariados de vazões. Ao nosso conhecimento, os trabalhos com cópulas para vazões existentes se concentraram unicamente em capturar apenas a auto-dependência de lag um. Neste sentido, a principal contribuição deste capítulo é propor uma metodologia que incorpora aos cenários estruturas de dependência temporal com ordens superiores a um.

A metodologia proposta utiliza cópulas vine para lidar com a dependência serial entre os meses. O modelo resultante pode ser entendido como uma versão (não-)linear do tradicional modelo PAR(p). Além de ser mais flexível em termos da estrutura temporal, esta metodologia permite que características tradicionalmente observadas em dados hidrológicos como assimetria e curtose possam ser adequadamente modeladas. Ressalta-se ainda o fato de que os cenários gerados serão sempre positivos. Para isto basta que as distribuições marginais possuam suporte positivo.

Os resultados da análise dentro da amostra e da simulação de Monte Carlo, indicaram que esta metodologia modela adequadamente a dependência serial. Além do mais, a análise dos cenários simulados demonstra que o modelo é capaz de simular cenários preservando determinadas características observadas na série histórica.

Referências bibliográficas

- [1] BOX, G. E. P.; JENKINS, G. M.. **Time series analysis: forecasting and control**. Holden-Day, San Francisco (USA), Revised edition, 1976.
- [2] SOUTO, M.; MOREIRA, A.; VEIGA, A.; STREET, A.; GARCIA, J. ; EPPRECHT, C.. **A high-dimensional VARX model to simulate monthly renewable energy supply**. In: IEEE POWER SYSTEMS COMPUTATION CONFERENCE, p. 1–7, Wroclaw, Poland, 2014.
- [3] PEREIRA, G. A. A.; SOUZA, R. C.. **Long memory models to generate synthetic hydrological series**. *Mathematical Problems in Engineering*, 2014:1–8, 2014.
- [4] SOUZA, R. C.; MARCATO, A. L. M.; DIAS, B. H. ; OLIVEIRA, F. L. C.. **Optimal operation of hydrothermal systems with hydrological scenario generation through bootstrap and periodic autoregressive models**. *European Journal of Operational Research*, 222(3):606–615, 2012.
- [5] SALAS, J. D.; DELLEUR, J. W.; YEVJEVICH, J. W. ; LANE, W. L.. **Applied modeling of hydrologic time series**. Water Resources Publications, Michigan (USA), first edition, 1980.
- [6] FERNANDEZ, B.; SALAS, J. D.. **Periodic gamma autoregressive processes for operational hydrology**. *Water Resources Research*, 22(10):1385–1396, 1986.
- [7] VECCHIA, A. V.. **Periodic autoregressive-moving average (parma) modeling with applications to water resources**. *JAWRA Journal of the American Water Resources Association*, 21(5):721–730, 1985.
- [8] SHARMA, A.; O'NEILL, R.. **A nonparametric approach for representing interannual dependence in monthly streamflow sequences**. *Water resources research*, 38(7):5–1–5–10, 2002.
- [9] ZACHARIAH, M.; REDDY, M. J.. **Development of an entropy-copula-based stochastic simulation model for generation of monthly**

- inflows into the hira kud dam. *ISH Journal of Hydraulic Engineering*, 19(3):267–275, 2013.
- [10] HAO, Z.; SINGH, V. P.. **Single-site monthly streamflow simulation using entropy theory**. *Water Resources Research*, 47(9), 2011.
- [11] LI, C.; SINGH, V. P. ; MISHRA, A. K.. **Monthly river flow simulation with a joint conditional density estimation network**. *Water Resources Research*, 49(6):3229–3242, 2013.
- [12] SKLAR, M.. **Fonctions de répartition à n dimensions et leurs marges**. *Publications de l'Institut de Statistique de L'Universit? de Paris*, 8:229–231, 1959.
- [13] AAS, K.; CZADO, C.; FRIGESSI, A. ; BAKKEN, H.. **Pair-copula constructions of multiple dependences**. *Insurance: Mathematics and Economics*, 44(2):182–198, 2009.
- [14] BEDFORD, T.; COOKE, R. M.. **Probability density decomposition for conditionally dependent random variables modeled by vines**. *Annals of Mathematics and Artificial Intelligence*, 32(1):245–268, 2001.
- [15] BEDFORD, T.; COOKE, R. M.. **Vines: A new graphical model for dependent random variables**. *Annals of Statistics*, 30(4):1031–1068, 2002.
- [16] LEE, T.; SALAS, J. D.. **Using copulas for stochastic streamflow generation**. In: *WORLD ENVIRONMENTAL AND WATER RESOURCES CONGRESS 2008*, p. 1–10, Ahupua'A (USA), 2008.
- [17] LEE, T.; SALAS, J. D.. **Copula-based stochastic simulation of hydrological data applied to Nile river flows**. *Hydrology Research*, 42(4):318–330, 2011.
- [18] HAO, Z.; SINGH, V. P.. **Entropy-copula method for single-site monthly streamflow simulation**. *Water Resources Research*, 48(6), 2012.
- [19] HAO, Z.; SINGH, V. P.. **Modeling multisite streamflow dependence with maximum entropy copula**. *Water Resources Research*, 49(10):7139–7143, 2013.
- [20] JEONG, C.; LEE, T.. **Copula-based modeling and stochastic simulation of seasonal intermittent streamflows for arid regions**. *Journal of Hydro-environment Research*, 9(4):604–613, 2015.

- [21] KONG, X. M.; HUANG, G. H.; FAN, Y. R. ; LI, Y. P.. **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, 2015.
- [22] CHEN, L.; SINGH, V. P.; GUO, S.; ZHOU, J. ; ZHANG, J.. **Copula-based method for multisite monthly and daily streamflow simulation.** Journal of Hydrology, 528:369–384, 2015.
- [23] ERHARDT, T. M.; CZADO, C. ; SCHEPSMEIER, U.. **R-vine models for spatial time series with an application to daily mean temperature.** Biometrics, 71(2):323–332, 2015.
- [24] JOE, H.. **Dependence Modeling with Copulas.** CRC Press, Taylor Francis Group, Boca Raton (USA), 2014.
- [25] MENDES, B. V. M.; AÍUBE, C.. **Copula based models for serial dependence.** International Journal of Managerial Finance, 7(1):68–82, 2011.
- [26] MENDES, B. V. M.; ACCIOLY, V. B.. **Robust pair-copula based forecasts of realized volatility.** Applied Stochastic Models in Business and Industry, 30(2):183–199, 2014.
- [27] BRECHMANN, E. C.; CZADO, C.. **Copar—multivariate time series modeling using the copula autoregressive model.** Applied Stochastic Models in Business and Industry, 31(4):495–514, 2015.
- [28] KUROWICKA, D.; COOKE, R. M.. **Uncertainty analysis with high dimensional dependence modeling.** John Wiley & Sons, Great Britain.
- [29] MCNEIL, A. J.; FREY, R. ; EMBRECHTS, P.. **Quantitative risk management: concepts, techniques, and tools.** Princeton University Press, USA, 2010.
- [30] MENDES, B. V. M.. **Introdução à análise de eventos extremos.** E-Papers Serviços Editoriais, Rio de Janeiro, Brasil., 2004.
- [31] CHERUBINI, U.; LUCIANO, E. ; VECCHIATO, W.. **Copula methods in finance.** Wiley Finance Series. John Wiley & Sons, 2004.
- [32] JOE, H.. **Families of m -variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters.** In: Rüschendorf, L.; Schweizer, B. ; Taylor, M. D., editors, DISTRIBUTIONS WITH

- FIXED MARGINALS AND RELATED TOPICS, volumen 28, Lecture Notes-Monograph Series, p. 120–141. Institute of Mathematical Statistics, 1996.
- [33] MORALES-NÁPOLES, O.; COOKE, R. ; KUROWICKA, D.. **About the number of vines and regular vines on n nodes.** 2010.
- [34] ACAR, E. F.; GENEST, C. ; NEŠLEHOVÁ, J.. **Beyond simplified pair-copula constructions.** Journal of Multivariate Analysis, 110:74–90, 2012.
- [35] HAFF, I. H.; AAS, K. ; FRIGESSI, A.. **On the simplified pair-copula construction—simply useful or too simplistic?** Journal of Multivariate Analysis, 101(5):1296–1310, 2010.
- [36] STOEBER, J.; JOE, H. ; CZADO, C.. **Simplified pair copula constructions—limitations and extensions.** Journal of Multivariate Analysis, 119:101–118, 2013.
- [37] CZADO, C.. **Statistical modelling with copulas.** Lecture Notes, 2015.
- [38] FISHER, R. A.. **Frequency distribution of the values of the correlation coefficients samples from an indefinitely large population.** Biometrika, 10(4):507–521, 1915.
- [39] HIPEL, K. W.; MCLEOD, A. I.. **Time series modelling of water resources and environmental systems.** Elsevier, Netherlands, 1994.
- [40] PENNA, D. D. J.. **Definição da árvore de cenários de afluências para o planejamento da operação energética de médio prazo.** Tese de doutorado, Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2009.
- [41] MACEIRA, M. E. P.; DAMÁZIO, J. M.. **Use of the par (p) model in the stochastic dual dynamic programming optimization scheme used in the operation planning of the brazilian hydropower system.** Probability in the Engineering and Informational Sciences, 20(01):143–156, 2006.
- [42] MCLEOD, A. I.. **Diagnostic checking periodic autoregression models with application.** The Journal of Time Series Analysis, 15(2):221–233, 1994.
- [43] DISSMANN, J. F.. **Statistical inference for regular vines and application.** Diplomarbeit, Fakultät für Mathematik, Technische Universität München, Deutschland, 2010.

- [44] ERHARDT, T. M.. **Predicting temperature time series using spatial vine copulae**. Masterarbeit, Fakultät für Mathematik, Technische Universität München, Deutschland, 2013.
- [45] DISSMANN, J.; BRECHMANN, E. C.; CZADO, C. ; KUROWICKA, D.. **Selecting and estimating regular vine copulae and application to financial returns**. Computational Statistics and Data Analysis, 59(1):52–69, 2013.
- [46] MAI, J. F.; SCHERER, M.. **Simulating copulas: stochastic models, sampling algorithms, and applications**. Imperial College Press, World Scientific, London (UK).
- [47] ONS - OPERADOR NACIONAL DO SISTEMA ELÉTRICO. **Atualização de séries históricas de vazões - período 1931 a 2014**. http://www.ons.org.br/operacao/vazoes_naturais.aspx. Acesso em: Agosto de 2014.
- [48] INMET - INSTITUTO NACIONAL DE METEOROLOGIA. **BDMEP - banco de dados meteorológicos para ensino e pesquisa**. <http://www.inmet.gov.br/porta1/>. Acesso em: Agosto de 2014.
- [49] ANA - AGÊNCIA NACIONAL DE ÁGUAS. **Sistema nacional de informações sobre recursos hídricos**. <http://www.snirh.gov.br/>. Acesso em: Agosto de 2014.
- [50] ANDERSON, T. W.; DARLING, D. A.. **Asymptotic theory of certain goodness of fit criteria based on stochastic processes**. The annals of mathematical statistics, 23(2):193–212, 1952.
- [51] MARSAGLIA, G.; MARSAGLIA, J.. **Evaluating the anderson-darling distribution**. Journal of Statistical Software, 9(2):1–5, 2004.
- [52] SCHEPSMEIER, U.; STOEBER, J.; BRECHMANN, E. C.; GRAELER, B.; NAGLER, T. ; ERHARDT, T.. **VineCopula: Statistical Inference of Vine Copulas**, 2015. R package version 1.5.
- [53] GENEST, C.; FAVRE, A. C.. **Everything you always wanted to know about copula modeling but were afraid to ask**. Journal of hydrologic engineering, 12(4):347–368, 2007.
- [54] BRECHMANN, E. C.; CZADO, C. ; AAS, K.. **Truncated regular vines in high dimensions with application to financial data**. Canadian Journal of Statistics, 40(1):68–85, 2012.

- [55] PATTON, A. J.. **A review of copula models for economic time series.** *Journal of Multivariate Analysis*, 110:4–18, 2012.
- [56] JOE, H.. **Multivariate models and multivariate dependence concepts.** Chapman & Hall/CRC Press, USA, first edition, 1997.
- [57] X. CHEN; FAN, Y.. **Estimation of copula-based semiparametric time series models.** *Journal of Econometrics*, 130(2):307–335, 2006.
- [58] DARSOW, W. F.; NGUYEN, B. ; OLSEN, E. T.. **Copulas and markov processes.** *Illinois Journal of Mathematics*, 36(4):600–642, 1992.
- [59] BEARE, B. K.. **Copulas and temporal dependence.** *Econometrica*, 78(1):395–410, 2010.
- [60] IBRAGIMOV, R.. **Copula-based characterizations for higher order markov processes.** *Econometric Theory*, 25(03):819–846, 2009.
- [61] CHOROŚ, B.; IBRAGIMOV, R. ; PERMIAKOVA, E.. **Copula estimation.** In: *COPULA THEORY AND ITS APPLICATIONS*, p. 77–91. Springer, 2010.
- [62] SMITH, M. S.. **Copula modelling of dependence in multivariate time series.** *International Journal of Forecasting*, 31(3):815–833, 2015.
- [63] SMITH, M.; MIN, A.; ALMEIDA, C. ; CZADO, C.. **Modeling longitudinal data using a pair-copula decomposition of serial dependence.** *Journal of the American Statistical Association*, 105:1467–1479, 2010.
- [64] ANA - AGÊNCIA NACIONAL DE ÁGUAS. **HidroWeb - Sistema de informações hidrológicas.** <http://hidroweb.ana.gov.br/HidroWeb.asp?TocItem=4100>. Acesso em: Novembro de 2016.
- [65] ANEEL - AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. **SIGEL - sistema de informações georreferenciadas do setor elétrico.** <http://sigel.aneel.gov.br/sigel.html>. Acesso em: Agosto de 2014.
- [66] ONS - OPERADOR NACIONAL DO SISTEMA ELÉTRICO. **Diagrama esquemático das usinas hidroelétricas do SIN. horizonte: 2014-2018.** http://www.ons.org.br/conheca_sistema/mapas_sin.aspx. Acesso em: Agosto de 2014.

A

Lista das usinas hidrelétricas utilizadas.

A classificação segue a nomenclatura dos *shapefiles* das bacias hidrográficas fornecidos pela ANA [64].

- Bacia 6 - Rio Paraná:

01. Jauru,
02. Ponte de Pedra,
03. Itiquira,
04. Manso,
05. Salto,
06. Salto Verdinho,
07. Caçu,
08. Barra dos Coqueiros,
09. Foz do Rio Claro.

- Bacia 5 - Atlântico, Trecho Leste:

10. Guilman Amorim,
11. Sá Carvalho,
12. Candonga,
13. Baguari,
14. Mascarenhas,
15. Salto Grande,
16. Porto Estrela,
17. Aimorés.

- Bacia 2 - Rio Tocantins:

18. Serra da Mesa,
19. Cana Brava,
20. São Salvador,
21. Peixe Angical,
22. Lajeado,
23. Estreito Tocantins,
24. Tucuruí.

- Bacia 1 - Rio Amazonas:

25. Rondon 2,
26. Dardanelos,
27. Guaporé,
28. Samuel,
29. Santo Antônio,
30. Balbina,
31. Santo Antônio do Jari,
32. Salto Apiacás,
33. SINOP,
34. Teles Pires,
35. Colíder,
36. Curuá-Una,
37. Belo Monte.

- Bacia 3 - Atlântico, T. Norte/Nordeste:

38. Cachoeira Caldeirão,
39. Ferreira Gomes.

B

Figuras suplementares: árvore T_1 estimada no Capítulo 4.

As figuras B.1, B.2, B.3, B.4, B.5 e B.6 foram elaboradas com base nos *shapefiles* das bacias hidrográficas brasileiras fornecidos pela ANA [64] e no diagrama esquemático das usinas hidroelétricas do SIN disponibilizado pela ONS [66]. As latitudes e longitudes das respectivas usinas foram obtidas junto ao Sistema de Informações Georreferenciadas do Setor Elétrico da ANEEL [65].

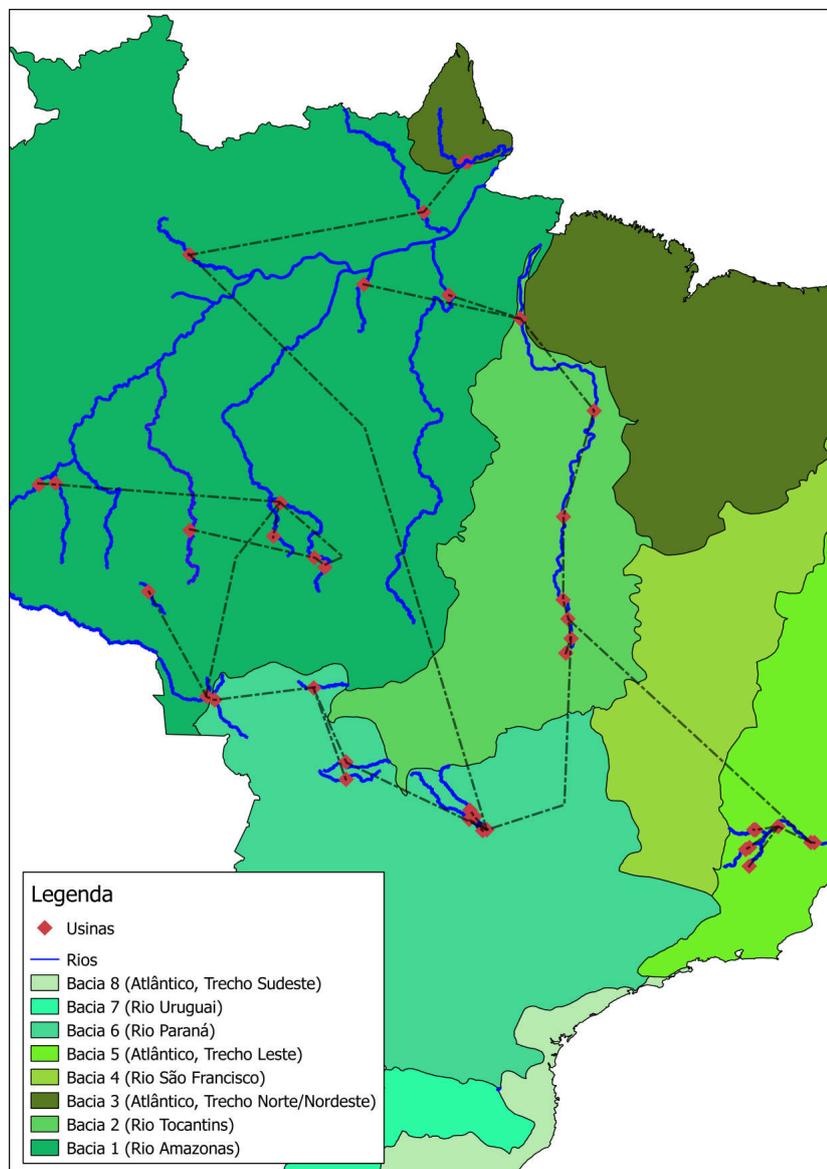


Figura B.1: A árvore T_1 sobre o mapa brasileiro.

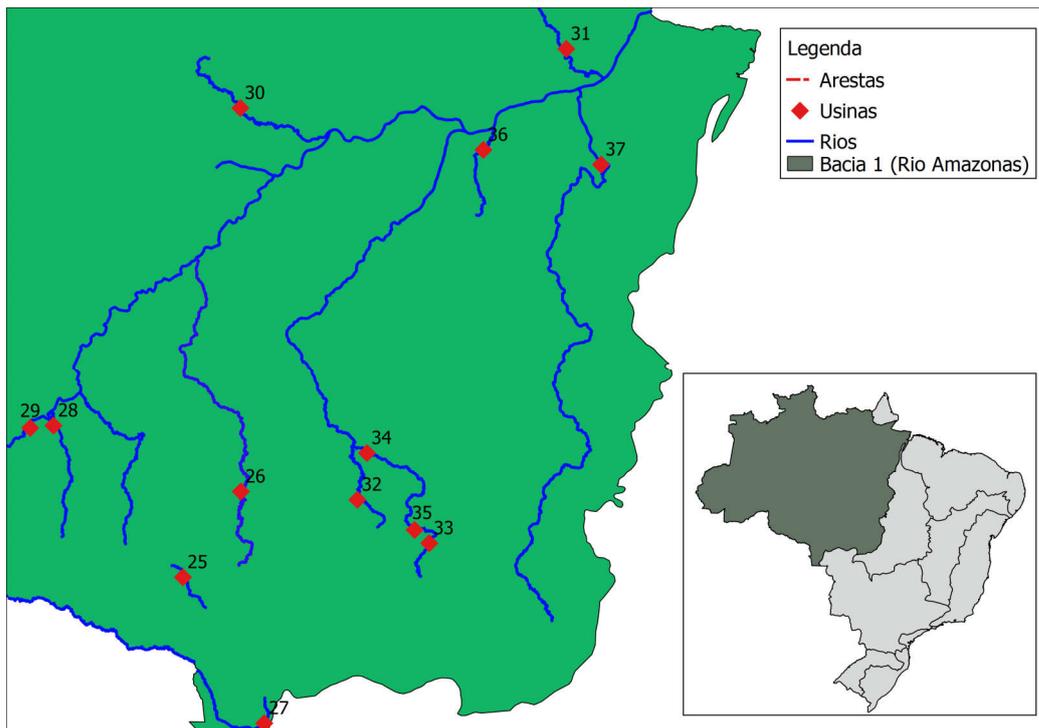


Figura B.2: Ilustração da árvore T_1 sobre a Bacia 1 (Rio Amazonas).

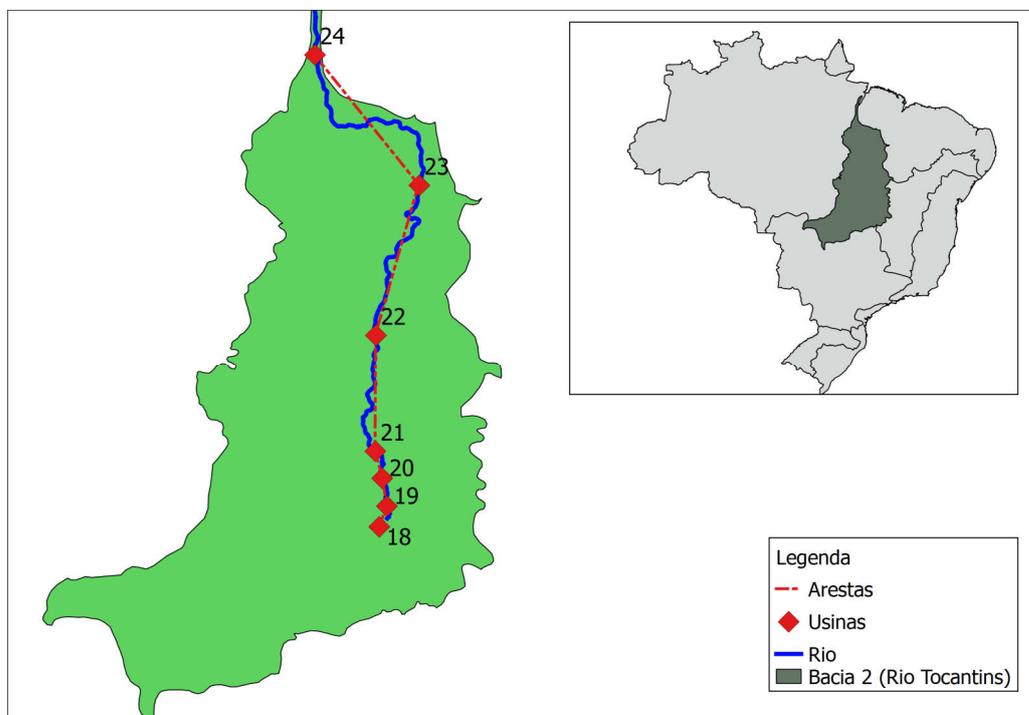


Figura B.3: Parte da árvore T_1 sobre a Bacia 2 (Rio Tocantins).

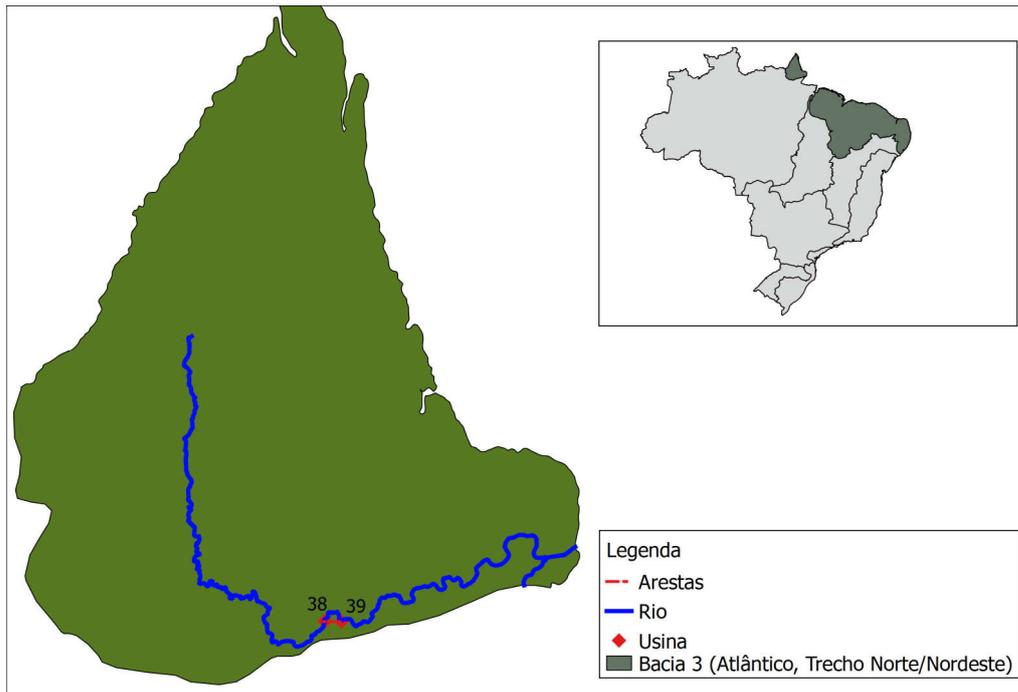


Figura B.4: Parte de T_1 sobre a Bacia 3 (Atlântico, Norte/Nordeste).

PUC-Rio - Certificação Digital Nº 1121533/CA

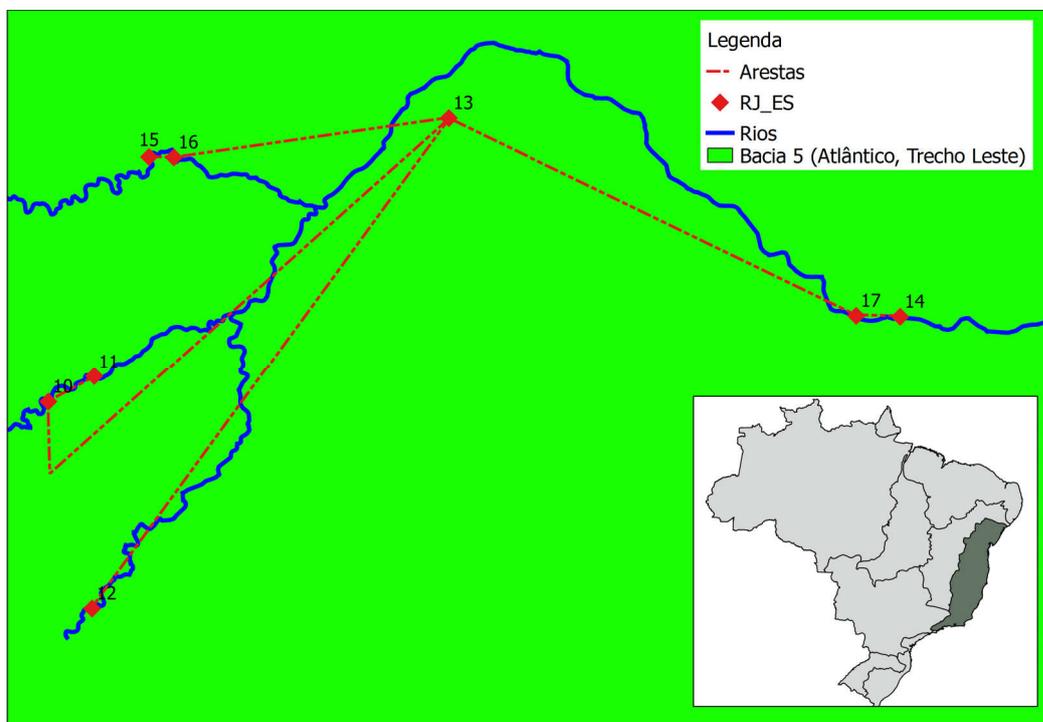


Figura B.5: Parte de T_1 sobre a Bacia 5 (Atlântico, Leste).

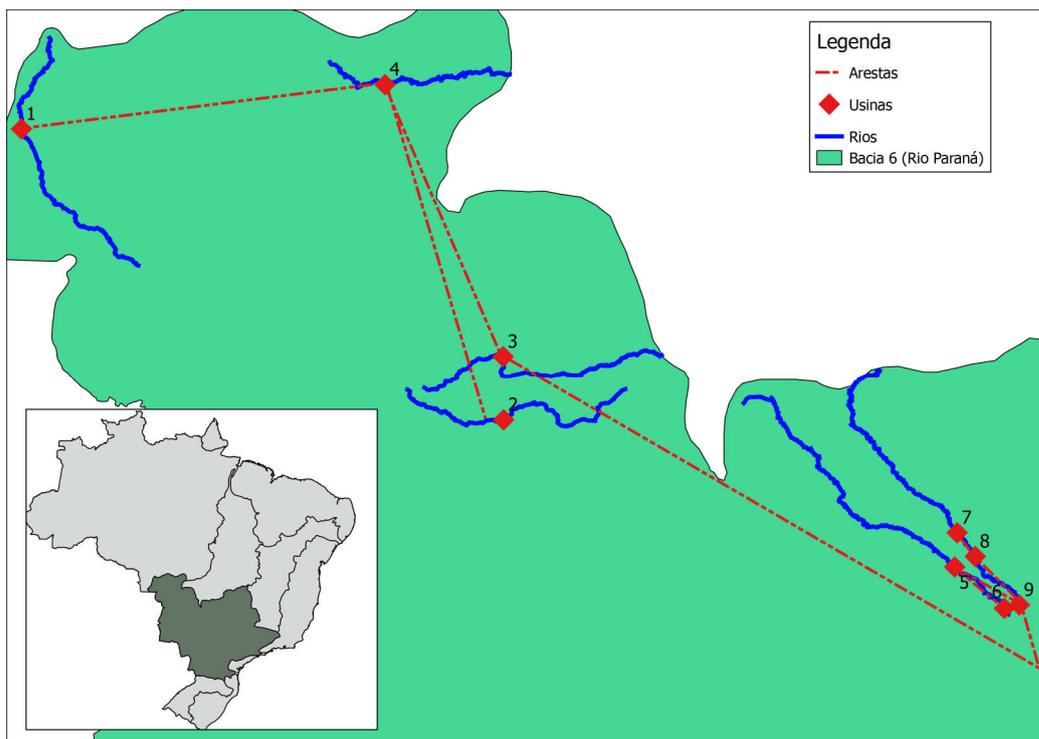


Figura B.6: Parte da árvore T_1 sobre a Bacia 6 (Rio Paraná).