

UNIVERSIDADE FEDERAL DE ALFENAS

VICTOR SILVEIRA GOMES

**MODELOS BASEADOS EM APRENDIZADO DE MÁQUINA PARA GERAÇÃO DE
SÉRIES SINTÉTICAS DO SETOR ELÉTRICO**

Alfenas/MG

2018

VICTOR SILVEIRA GOMES

**MODELOS BASEADOS EM APRENDIZADO DE MÁQUINA PARA GERAÇÃO DE
SÉRIES SINTÉTICAS DO SETOR ELÉTRICO**

Dissertação apresentada ao Programa de Pós-Graduação em Estatística Aplicada e Biometria, área de concentração em Estatística Aplicada e Biometria da Universidade Federal de Alfenas, MG, como parte dos requisitos para a obtenção do título de Mestre. Linha de Pesquisa: Modelagem Estatística e Estatística Computacional.
Orientador: Prof. Dr. Ricardo Menezes Salgado.
Coorientador: Prof. Dr. Luiz Alberto Beijo.

Alfenas/MG

2018

Dados Internacionais de Catalogação-na-Publicação (CIP)
Sistema de Bibliotecas da Universidade Federal de Alfenas

Gomes, Victor Silveira.

G633m Modelos baseados em aprendizado de máquina para geração de séries sintéticas do setor elétrico. / Victor Silveira Gomes – Alfenas/MG, 2018.
160 f.: il. –

Orientador: Ricardo Menezes Salgado.
Dissertação (Mestrado em Estatística Aplicada e Biometria) -
Universidade Federal de Alfenas, 2018.
Bibliografia.

1. Aprendizado de computador. 2. Modelo de Markov. 3. Hidrologia. 4.
Modelos em séries temporais. 5. Série sintética. I. Salgado, Ricardo Menezes.
II. Título.

CDD-519.232

Ficha Catalográfica elaborada por Fátima dos Reis Goiatá
Bibliotecária-Documentalista CRB/6-425



MINISTÉRIO DA EDUCAÇÃO
Universidade Federal de Alfenas / UNIFAL-MG
Programa de Pós-graduação em Estatística Aplicada e Biometria

Rua Gabriel Monteiro da Silva, 700. Alfenas - MG CEP 37130-000
Fone: (35) 3299-1392 (Secretaria) (35) 3299-1121 (Coordenação)
<https://www.unifal-mg.edu.br/ppgeab/>



VICTOR SILVEIRA GOMES

“MODELOS BASEADOS EM APRENDIZADO DE MÁQUINA PARA GERAÇÃO DE
SÉRIES SINTÉTICAS DO SETOR ELÉTRICO”

A Banca Examinadora, abaixo assinada, aprova a
Dissertação apresentada como parte dos requisitos para
a obtenção do título de Mestre em Estatística Aplicada
e Biometria pela Universidade Federal de Alfenas.
Área de Concentração: Estatística Aplicada e
Biometria

Aprovado em: 18 de junho de 2018.

Prof. Dr. Ricardo Menezes Salgado

Instituição: UNIFAL-MG

Assinatura: Ricardo Menezes Salgado

Prof. Dr. Flávio Barbieri Gonzaga

Instituição: UNIFAL-MG

Assinatura: Flávio Barbieri Gonzaga

Prof. Dr. Takaaki Ohishi

Instituição: UNICAMP

Assinatura: Takaaki Ohishi

*Dedico aos meus pais, amigos e
professores que me ajudaram nessa
caminhada.*

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Agradeço ao Operador Nacional do Sistema Elétrico (ONS) por liberar os dados para análise, a professora Dra. Dalia da *Duke University* por disponibilizar o SynTise para teste, ao Laboratório de Inteligência Computacional (LInC) pela estrutura disponibilizada para o trabalho, ao Programa de Pós-Graduação em Estatística Aplicada e Biometria (PPGEAB) da UNIFAL-MG pela oportunidade, meus amigos que fiz nesse período, aos meus pais, irmãos e aos orientadores Ricardo Menezes Salgado e Luiz Alberto Beijo por toda ajuda e paciência.

RESUMO

O Brasil tem como principal fonte de fornecimento de energia a hidroelétrica, devido principalmente à sua grande capacidade hídrica. Entender o comportamento das vazões de suas bacias é um fator fundamental para otimização da produção desse tipo energia, porém os dados históricos presentes são limitados, tornando-se um empecilho para o estudo, dado a importância dele no planejamento da produção de energia elétrica. Uma solução que vem sendo utilizada na literatura recente é a geração de série sintética. Neste trabalho, as seguintes técnicas foram utilizadas para geração sintética das vazões dos postos de Água Vermelha e Volta Grande: o SynTise, modelo apresentado em Denaxas et al. (2015), máquinas de vetores de suporte (SVM), redes neurais multicamadas (MLP), *random forest* (RF) e o modelo autorregressivo (AR). Foram geradas séries sintéticas equivalentes a 2000 anos em ambos reservatórios. O trabalho analisou quatro diferentes propostas para a seleção do componente aleatória dos modelos AR, MLP, SVM e RF que são: através de uma distribuição de probabilidade simétrica, através de uma distribuição de probabilidade assimétrica, pela ordem cronológica e através dos resíduos estimados. As novas propostas de componentes aleatórios e o método de seleção clássico, a seleção aleatória dos resíduos, foram avaliadas para os dois postos, assim como o SynTise, que foi ajustado para gerar séries sintéticas mensais para vazão dos reservatórios. Os resultados mostraram que, para os dois postos avaliados, modelos com componente aleatória ao longo do tempo foram opções melhores que o modelo clássico de componente aleatória selecionada aleatoriamente em todas as técnicas avaliadas. Na comparação entre os melhores resultados de cada técnica, obteve-se que para o posto de Volta Grande, o SVM apresentou os melhores resultados, enquanto para Água Vermelha, o MLP foi melhor entre todos os modelos.

Palavras-chave: Aprendizado de máquina. Modelos de Markov. Hidrologia. Modelos em séries temporais. Série sintética.

ABSTRACT

Brazil's main source of energy supply is hydroelectric, due mainly to its large hydro capacity. Understanding the flow behavior of its basins is a fundamental factor to optimize the production of this type of energy, but the present historical data are limited, becoming a hindrance to the study, given its importance in the planning of electric energy production. One solution that has been used in the recent literature is the generation of synthetic series. In this work, the following techniques were used for the synthetic generation of the flows of the Água Vermelha and Volta Grande stations: SynTise, model presented in [citeonline Denaxas \(2015\)](#), support vector machines (SVM), multilayer perceptron (MLP), random forest (RF) and the autoregressive model (AR). Synthetic series equivalent to 2000 years were generated for all these reservoirs. The work analyzed four different proposals for the selection of the random component of the AR, MLP, SVM and RF models, which are: through a symmetric probability distribution, through an asymmetric probability distribution, in chronological order and through the estimated residuals. The new random component proposals and the classical selection method, the random selection of the residues, were evaluated for the two stations, as well as SynTise, which was adjusted to generate synthetic monthly series for reservoir flow. The results showed that, for the two stations evaluated, models with random component over time were better options than the classic model of random component randomly selected in all the techniques evaluated. In the comparison between the best results of each technique, it was obtained that for the Volta Grande station, the SVM presented the best results, while for Água Vermelha, the MLP was better among all the models

Key-words: Hydrology. Machine learning. Markov models. Synthetic time series. Time series models.

LISTA DE TABELAS

Tabela 1 – RMSE da FAC e FDP mensal dos melhores modelos em cada uma das técnicas analisadas para usina de Água Vermelha	62
Tabela 2 – RMSE da FAC e FDP mensal dos melhores modelos em cada uma das técnicas analisadas para usina de Volta Grande	67
Tabela 3 – RMSE da FDP e FAC mensal dos modelos AR de CA ALEAT, DIST.N e DIST.SN	81
Tabela 4 – RMSE da FAC e FDP mensal dos modelos AR de CA ALEAT, TEMP e EST	85
Tabela 5 – RMSE da FAC e FDP mensal dos modelos SVM de CA ALEAT, DIST.N e DIST.SN	90
Tabela 6 – RMSE da FAC e FDP mensal dos modelos AR de CA ALEAT, TEMP e EST	94
Tabela 7 – RMSE da FAC e FDP mensal dos modelos MLP de CA ALEAT, DIST.N e DIST.SN	99
Tabela 8 – RMSE da FAC e FDP mensal dos modelos MLP de CA ALEAT, TEMPO e EST	103
Tabela 9 – Estruturas do modelo <i>Random Forest</i> para geração das séries sintéticas mensais na usina de Água Vermelha	104
Tabela 10 – RMSE da FAC e FDP mensal dos modelos RF de CA ALEAT, DIST.N e DIST.SN	108
Tabela 11 – RMSE da FAC e FDP mensal dos modelos RF de CA ALEAT, TEMP e EST	112
Tabela 12 – RMSE da FDP e FAC mensal dos modelos AR com CA ALEAT, DIST.N e DIST.SN	122
Tabela 13 – RMSE da FAC e FDP mensal dos modelos AR de CA ALEAT, TEMP e EST	126
Tabela 14 – RMSE da FAC e FDP mensal dos modelos SVM de CA ALEAT, DIST.N e DIST.SN	131
Tabela 15 – RMSE da FAC e FDP mensal dos modelos SVM de CA ALEAT, TEMP e EST	135

Tabela 16 – RMSE da FAC e FDP mensal dos modelos MLP de CA ALEAT, DIST.N e DIST.SN	140
Tabela 17 – RMSE da FAC e FDP mensal dos modelos MLP de CA ALEAT, TEMP e EST	144
Tabela 18 – Estruturas do modelo <i>Random Forest</i> para geração de séries sintéticas mensais na usina de Água Vermelha	145
Tabela 19 – RMSE da FAC e FDP mensal dos modelos RF de CA ALEAT, DIST.N e DIST.SN	149
Tabela 20 – RMSE da FAC e FDP mensal dos modelos RF de CA ALEAT, TEMP e EST	153

LISTA DE FIGURAS

Figura 1 – Capacidade energética do Brasil- 1971 a 2016(%)	15
Figura 2 – Usina Bernardo Mascarenhas, 1889, Juiz de Fora-MG	24
Figura 3 – Configuração da rede MLP	29
Figura 4 – Terminologias básicas de uma árvore de decisão	30
Figura 5 – Diagrama das usinas hidroelétricas pertencentes a bacia do Rio Grande	34
Figura 6 – Usina hidrelétrica de Água Vermelha	35
Figura 7 – Série histórica dividida por décadas da usina de Água Vermelha	36
Figura 8 – Estatísticas descritivas mensais da série de vazões mensais da usina de Água Vermelha	37
Figura 9 – Box-plots das vazões mensais da usina de Água Vermelha	38
Figura 10 – Série histórica da usina de Água Vermelha	39
Figura 11 – Usina hidrelétrica de Volta Grande	39
Figura 12 – Série histórica dividida por décadas da usina de Volta Grande	41
Figura 13 – Estatísticas descritivas mensais da usina de Água Vermelha	42
Figura 14 – Box-plots das vazões mensais da usina de Volta Grande	43
Figura 15 – Série histórica da usina de Volta Grande	44
Figura 16 – Histograma da série vazões mensais	45
Figura 17 – Geração de séries sintéticas	49
Figura 18 – Fluxograma da execução das séries sintéticas utilizadas nos modelos trabalho	51
Figura 19 – Exemplo de um box-plot	55
Figura 20 – Exemplo de um q-qplot	56
Figura 21 – Comparação entre as estatísticas descritivas das vazões mensais gera- das pelos melhores modelos em cada uma das técnicas analisadas para usina de Água Vermelha	58
Figura 22 – Comparação entre os box-plots das vazões geradas pelos melhores mo- delos da cada uma das técnicas analisadas para usina de Água Vermelha	60
Figura 23 – Comparação entre os q-qplots das vazões geradas pelos melhores mo- delos da cada uma das técnicas analisadas para usina de Água Vermelha	61

Figura 24 – Comparação das estatísticas descritivas mensais das vazões mensais geradas pelos melhores modelos em cada uma das técnicas analisadas para usina de Volta Grande	64
Figura 25 – Comparação entre os box-plots das vazões geradas pelos melhores modelos em cada uma das técnicas analisadas para usina de Volta Grande	65
Figura 26 – Comparação entre os q-qplots das vazões geradas pelos melhores modelos em cada uma das técnicas analisadas para usina de Volta Grande	66
Figura 27 – Comparação entre as estatísticas descritivas das séries de vazões mensais geradas pelos modelos AR com CA ALEAT, DIST.N e DIST.SN para usina de Água Vermelha	78
Figura 28 – Box-plots das vazões geradas pelos modelos AR para usina de Água Vermelha	79
Figura 29 – Q-qplots das vazões geradas pelos modelos AR para usina de Água Vermelha	80
Figura 30 – Comparação entre as estatísticas descritivas das séries de vazões mensais geradas pelos modelos AR de CA ALEAT, TEMP e EST para usina de Água Vermelha	82
Figura 31 – Box-plots das vazões geradas pelos modelos AR para usina de Água Vermelha	83
Figura 32 – Q-qplots das vazões geradas pelos modelos AR para usina de Água Vermelha	84
Figura 33 – Comparação entre as estatísticas descritivas das séries de vazões mensais geradas pelos modelos SVM de CA ALEAT, DIST.N e DIST.SN para usina de Água Vermelha	87
Figura 34 – Box-plots das vazões geradas pelos modelos SVM para usina de Água Vermelha	88
Figura 35 – Q-qplots das vazões geradas pelos modelos SVM para usina de Água Vermelha	89
Figura 36 – Comparação entre as estatísticas descritivas das vazões geradas pelos modelos SVM de CA ALEAT, TEMP e EST para usina de Água Vermelha	91

Figura 37 – Box-plots das vazões geradas pelos modelos SVM para usina de Água Vermelha	92
Figura 38 – Q-qplots das vazões geradas pelos modelos SVM para usina de Água Vermelha	93
Figura 39 – Comparação entre as estatísticas descritivas das vazões geradas pelos modelos MLP de CA ALEAT, DIST.N e DIST.SN para usina de Água Vermelha	96
Figura 40 – Box-plots das vazões geradas pelos modelos MLP para usina de Água Vermelha	97
Figura 41 – Q-qplots das vazões geradas pelos modelos MLP para usina de Água Vermelha	98
Figura 42 – Comparação entre as estatísticas descritivas das vazões mensais geradas pelos modelos MLP de CA ALEAT, TEMP e EST para usina de Água Vermelha	100
Figura 43 – Box-plots das vazões geradas pelos modelos MLP para usina de Água Vermelha	101
Figura 44 – Q-qplots das vazões geradas pelos modelos MLP para usina de Água Vermelha	102
Figura 45 – Comparação entre as estatísticas descritivas das vazões geradas pelos modelos RF de CA ALEAT, DIST.N e DIST.SN para usina de Água Vermelha	105
Figura 46 – Box-plots das vazões geradas pelos modelos RF para usina de Água Vermelha	106
Figura 47 – Q-qplots das vazões geradas pelos modelos RF para usina de Água Vermelha	107
Figura 48 – Comparação entre estatísticas descritivas das vazões geradas pelos modelos RF de CA ALEAT, TEMP e EST para usina de Água Vermelha	109
Figura 49 – Box-plots das vazões geradas pelos modelos RF para usina de Água Vermelha	110
Figura 50 – Q-qplots das vazões geradas pelos modelos RF para usina de Água Vermelha	111

Figura 51 – RMSE da FAC dos modelos de série possíveis no SynTise para usina de Água Vermelha	113
Figura 52 – RMSE da FDP dos modelos de série possíveis no SynTise para usina de Água Vermelha	114
Figura 53 – Comparação entre as estatísticas descritivas das vazões mensais geradas pelo modelo SynTise para usina de Água Vermelha	115
Figura 54 – Box-plots das vazões geradas pelo SynTise para usina de Água Vermelha	116
Figura 55 – Q-qplot das vazões geradas pelo SynTise.	117
Figura 56 – Comparação entre as estatísticas descritivas das vazões mensais geradas pelos modelos AR de CA ALEAT, DIST.N e DIST.SN para usina de Volta Grande	119
Figura 57 – Box-plots das vazões geradas pelos modelos AR para usina de Volta Grande	120
Figura 58 – Q-qplots das vazões geradas pelos modelos AR para usina de Volta Grande	121
Figura 59 – Comparação entre as estatísticas descritivas das vazões mensais geradas pelos modelos AR de CA ALEAT, TEMP e EST para usina de Volta Grande	123
Figura 60 – Box-plots das vazões geradas pelos modelos AR para usina de Volta Grande	124
Figura 61 – Q-qplots das vazões geradas pelos modelos AR para usina de Volta Grande	125
Figura 62 – Comparação entre as estatísticas descritivas das vazões mensais geradas pelos modelos SVM de CA ALEAT, DIST.N e DIST.SN para usina de Volta Grande	128
Figura 63 – Box-plots das vazões geradas pelos modelos SVM para usina de Volta Grande	129
Figura 64 – Q-qplots das vazões geradas pelos modelos SVM para usina de Volta Grande	130
Figura 65 – Comparação entre as estatísticas descritivas das vazões mensais geradas pelos modelos SVM de CA ALEAT, TEMP e EST para usina de Volta Grande	132

Figura 66 – Box-plots das vazões geradas pelos modelos SVM para usina de Volta Grande	133
Figura 67 – Q-qplots das vazões geradas pelos modelos SVM para usina de Volta Grande	134
Figura 68 – Comparação entre as estatísticas descritivas das vazões mensais geradas pelos modelos MLP de CA ALEAT, DIST.N e DIST.SN para usina de Volta Grande	137
Figura 69 – Box-plots das vazões geradas pelos modelos MLP para usina de Volta Grande	138
Figura 70 – Q-qplots das vazões geradas pelos modelos MLP para usina de Volta Grande	139
Figura 71 – Comparação entre as estatísticas descritivas das vazões mensais geradas pelos modelos MLP de CA ALEAT, TEMPO e EST para usina de Volta Grande	141
Figura 72 – Box-plots das vazões geradas pelos modelos MLP para usina de Volta Grande	142
Figura 73 – Q-qplots das vazões geradas pelos modelos MLP para usina de Volta Grande	143
Figura 74 – Comparação entre as estatísticas descritivas das vazões geradas pelos modelos RF de CA ALEAT, DIST.N e DIST.SN para usina de Volta Grande	146
Figura 75 – Box-plots das vazões geradas pelos modelos RF para usina de Volta Grande	147
Figura 76 – Q-qplots das vazões geradas pelos modelos RF para usina de Volta Grande	148
Figura 77 – Comparação entre as estatísticas descritivas das vazões mensais geradas pelos modelos RF de CA ALEAT, TEMP e EST para usina de Volta Grande	150
Figura 78 – Box-plots das vazões geradas pelos modelos RF para usina de Volta Grande	151
Figura 79 – Q-qplots das vazões geradas pelos modelos RF para usina de Volta Grande	152

Figura 80 – RMSE da FAC dos modelos de série possíveis no SynTise para usina de Volta Grande	154
Figura 81 – RMSE da FDP dos modelos de série possíveis no SynTise para usina de Volta Grande	155
Figura 82 – Comparação entre as estatísticas descritivas das vazões mensais geradas pelo modelo SynTise para usina de Volta Grande	156
Figura 83 – Box-plots das vazões geradas pelo SynTise para usina de Volta Grande	157
Figura 84 – Q-qplot das vazões geradas pelo SynTise	158

SUMÁRIO

1	INTRODUÇÃO	15
1.1	OBJETIVOS GERAIS	17
1.2	OBJETIVOS ESPECÍFICOS	17
1.3	ORGANIZAÇÃO DE TEXTO	18
2	REFERENCIAL TEÓRICO	19
2.1	REVISÃO BIBLIOGRÁFICA	19
2.2	SÉRIE DE VAZÕES	22
2.3	SÉRIES SINTÉTICAS	25
2.4	TÉCNICAS UTILIZADAS	26
2.4.1	Modelo autorregressivo (AR)	26
2.4.2	Máquinas de vetores e suporte (SVM)	27
2.4.3	Redes neurais multicamadas (MLP)	27
2.4.4	<i>Random forest</i>	29
2.4.5	SynTise	32
3	MATERIAL E MÉTODOS	34
3.1	DADOS DE VAZÕES	34
3.2	GERAÇÃO DE SÉRIES SINTÉTICAS	45
3.2.1	Modelos para geração baseado em séries temporais e em inteligência artificial	47
3.2.1.1	Técnicas para componente estacionária	47
3.2.1.2	Modelos para gerar o componente aleatório	47
3.2.2	Modelos baseados em cadeia de Markov	49
3.2.3	Fluxograma do modelo proposto	50
3.3	MÉTRICA DE AVALIAÇÃO	53
4	RESULTADOS	57
4.1	RECURSOS COMPUTACIONAIS	57
4.2	USINA DE ÁGUA VERMELHA	57
4.3	USINA DE VOLTA GRANDE	63
4.4	ANÁLISE GERAL	68
5	CONCLUSÃO E PERSPECTIVAS FUTURAS	70
	REFERÊNCIAS	71
	APÊNDICE A - MODELOS AR PARA USINA DE ÁGUA VERMELHA	77
	APÊNDICE B - MODELOS SVM PARA USINA DE ÁGUA VERMELHA	86
	APÊNDICE C - MODELOS MLP PARA USINA DE ÁGUA VERMELHA	95
	APÊNDICE D - MODELOS RF PARA USINA DE ÁGUA VERMELHA	104
	APÊNDICE E - MODELO SYNTISE PARA USINA DE ÁGUA VERMELHA	113
	APÊNDICE F - MODELOS AR PARA USINA DE VOLTA GRANDE	118
	APÊNDICE G - MODELOS SVM PARA USINA DE VOLTA GRANDE	127
	APÊNDICE H - MODELOS MLP PARA USINA DE VOLTA GRANDE	136
	APÊNDICE I - MODELOS RF PARA USINA DE VOLTA GRANDE	145
	APÊNDICE I - MODELO SYNTISE PARA USINA DE VOLTA GRANDE	154

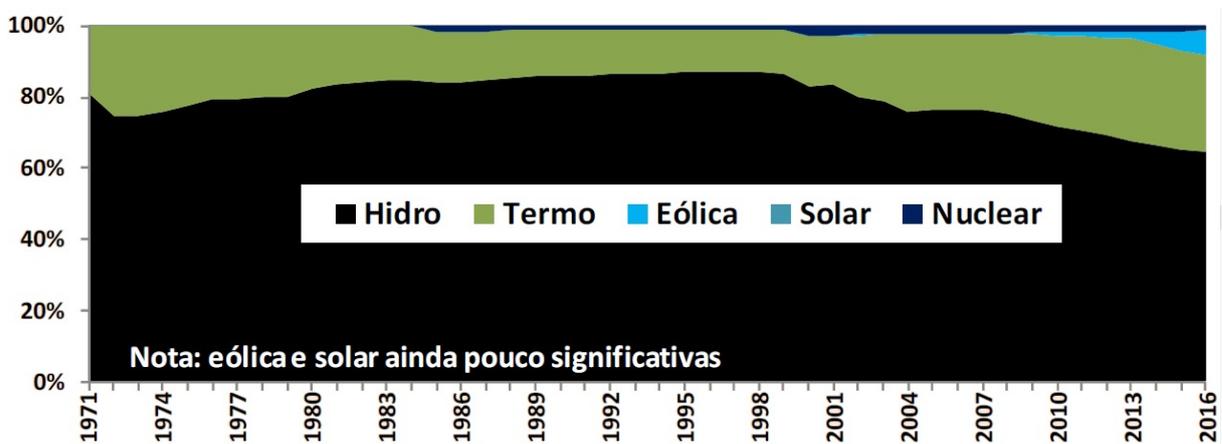
1 INTRODUÇÃO

Nos dias de hoje, a energia elétrica é um recurso indispensável para o desenvolvimento econômico de um país, fazendo com que haja uma grande preocupação que seu uso seja realizado de maneira a maximizar os benefícios proporcionados e minimizar os impactos negativos ao meio ambiente e à sociedade.

Dentre os tipos de produção de energia elétrica existente no Brasil, a que mais se destaca sem dúvida é a hidroelétrica. A geografia do território brasileiro, marcada em grande parte por um relevo suave, favoreceu a formação dos imenso recursos hídricos disponíveis. Desta forma, o Brasil possui uma das mais extensas redes fluviais do mundo, sendo um dos países de com maior potencial hidráulico, atrás apenas da China (ANEEL, 2016). Para se ter uma ideia mais concreta, o Brasil detém 12% das reservas de água doce do planeta, perfazendo 53% dos recursos hídricos da América do Sul. Grande parte das fronteiras do País é definida por corpos d'água – são 83 rios fronteirços e transfronteirços, além de bacias hidrográficas e de aquíferos. As bacias de rios transfronteirços ocupam 60% do território brasileiro (BRASIL, 2018).

Toda essa abundância se reflete na capacidade energética do país, que como mostra a Figura 1, desde 1971 sempre teve a energia hidrelétrica como a com maior percentual. Esse percentual da produção hidroelétrica foi crescendo até 1998, quando foi diminuindo gradativamente, principalmente com o advento de outras fontes de energia limpa, como a solar e eólica. Atualmente, a produção hidroelétrica é responsável por 64,5% da capacidade energética do país, sendo o menor em percentual desde 1971.

Figura 1 – Capacidade energética do Brasil- 1971 a 2016(%)



Fonte: BRASIL, 2017, p. 2.

Considerada uma energia limpa, visto que é uma fonte renovável e não emite poluentes, as hidroelétricas porém geram outros impactos tanto ecológicos quanto sociais como: a inundação de áreas para a construção de barragens que geram problemas de realocação das populações ribeirinhas, comunidades indígenas e pequenos agricultores, destruição de extensas áreas de vegetação natural, matas ciliares, o desmoronamento das margens, o assoreamento do leito dos rios, prejuízos à fauna e à flora locais, alterações no regime hidráulico dos rios, possibilidades da transmissão de doenças, entre outros (FRANCISCO, 2016).

A produção energética em usinas hidroelétricas depende diretamente da vazão dos seus afluentes. Entende-se como vazão o volume de água que atravessa uma seção transversal por unidade de tempo. As vazões médias representam a disponibilidade hídrica natural de uma bacia hidrográfica, sendo seu conhecimento de grande importância para um adequado planejamento do uso e da gestão compartilhada dos recursos hídricos, minimizando assim os conflitos entre diversos usuários (NOVAES, 2005). No Brasil, as usinas termoelétricas, que liberam CO₂ advinda da queima de combustível, sendo assim considerada uma energia suja por conta o impacto ambiental causado não chegam a representar nem 19% do fornecimento de energia no país. Além de utilizadas para abastecer indústrias, também funcionam como fonte alternativa em casos de crise energética ou para complementar a geração das usinas hidroelétricas, principalmente nos períodos de seca (MUNDO DA ELÉTRICA, 2016). Mesmo com a expansão de outras fontes em energia limpa no país, como eólica e solar por exemplo, atualmente elas ainda não são suficientes para suprir toda a demanda energética do país nesses períodos. Portanto, o gerenciamento dos recursos hídricos através da análise de registros históricos do país para produção de energia é algo de extrema importância.

Além do impacto ambiental, uma má gestão da produção energética também pode acarretar um grande prejuízo financeiro. Um grande exemplo dessa importância pode ser visto no relato em Bunn e Farmer (1985), onde é dito que em 1984, a diminuição do erro na previsão da demanda elétrica em 1% gerou uma economia na produção em torno de 10 milhões de libras ao ano para empresa de energia responsável, a *Central Electricity Generating Board* (CEGB), localizada no Reino Unido na cidade de Londres.

Dado toda importância vista quanto ao planejamento da produção energética do país, concluí-se que é de extrema importância que sua elaboração seja realizada de forma bem rigorosa. No entanto, os conjuntos de dados relacionados às vazões de rios geralmente possuem um número limitado de observações disponíveis para estudos de longo prazo, sendo insuficientes

para sua análise. Uma alternativa para aumentar o histórico de dados é o uso de séries sintéticas (SALGADO et al., 2012). Esta solução aliás já vem sendo utilizada por vários pesquisadores ao longo dos últimos anos, que desenvolveram modelos baseados em propriedades determinísticas e estocásticas, aplicados a várias escalas de tempo (anual, mensal, diária e horária), de acordo com o mérito de cada trabalho (DETZEL; MINE, 2009). Tratando-se de vazões de rios, o uso de séries sintéticas é de extrema importância, visto a quantidade limitada de dados presentes nesse tipo de estudo e o prejuízo financeiro que pode ser causado pelo mal planejamento da produção de energia elétrica. A vasta quantidade de trabalhos recentes relacionados ao tema tanto no Brasil (EVSUKOFF; DE LIMA; EBECKEN, 2011, SALGADO et al., 2012; ; SOUSA; GUEDES; SALDANHA, 2015) quanto no exterior (AKBARI; VAN OVERLOOP; AFSHAR, 2011; BORGOMEIO; FARMER; HALL, 2015; DENAXAS et al., 2015) refletem bem essa importância.

1.1 OBJETIVOS GERAIS

O objetivo principal deste trabalho é desenvolver modelos para gerar séries sintéticas para vazões médias mensais das usinas hidrelétricas de Volta Grande e Água Vermelha utilizando diferentes técnicas para o desenvolvimento desses modelos.

1.2 OBJETIVOS ESPECÍFICOS

Os objetivos específicos deste trabalho são:

- (a) propor metodologias de geração de série sintética para as vazões médias mensais das usinas utilizando os seguintes modelos: autorregressivo (AR), redes neurais artificiais (MLP), máquinas de vetor e suporte (SVM) e *random forest* (RF);
- (b) otimizar os resultados dos modelos citados utilizando diferentes abordagens para a inclusão dos resíduos nos modelos;
- (c) reprogramar o modelo SynTise, apresentado em Denaxas et al. (2015) para geração de séries sintéticas de velocidade e força dos ventos, de modo a gerar séries sintéticas de vazões médias mensais;

- (d) comparar os modelos levando em consideração as estatísticas descritivas e os erros da função densidade de probabilidade (FDP) e autocorrelação (FAC), com intuito principal de verificar se houve algum ganho significativo quanto ao modelo autorregressivo, que é um modelo de geração de séries sintética já consolidado.

1.3 ORGANIZAÇÃO DE TEXTO

O trabalho está organizado em cinco seções dispostas da seguinte forma: Na Seção 2 é apresentada uma revisão bibliográfica sobre geração de séries sintéticas, alguns conceitos importantes sobre esse tema. Ainda nesta mesma seção também é exibido definições, conceitos e histórico sobre série de vazões assim como a apresentação das técnicas utilizadas para sua geração nesse trabalho. A Seção 3 mostra os conjuntos de dados empregues, como é feita a modelagem para geração da série sintética e quais as métricas de avaliação. Na Seção 4, são exibidos os recursos computacionais utilizados, os resultados das análises nos postos de Água Vermelha e Volta Grande, além de uma análise geral desses resultados. Por fim, na Seção 5 são feitas conclusões sobre os resultados obtidos, além de propostas futuras para o trabalho.

2 REFERENCIAL TEÓRICO

Nesta seção será feita uma revisão bibliográfica acerca de diversos trabalhos sobre geração de série sintética. Também será apresentado um estudo sobre vazões de rio, séries sintéticas além das técnicas utilizadas nesse trabalho para geração dos modelos.

2.1 REVISÃO BIBLIOGRÁFICA

Na literatura, os modelos estocásticos autorregressivos são métodos já consolidados para geração de série sintética, tendo sido aplicado com sucesso em trabalhos mais clássicos. Blanchard e Desrochers (1984) utilizou o modelo autorregressivo AR(2) para geração de séries sintéticas de velocidade dos ventos na ilha de Madalena, localizada na província de Quebec no Canadá. A base utilizada nesse estudo era referentes aos valores mensais da velocidade dos ventos no período de janeiro de 1973 à dezembro de 1976. Mesmo sendo um modelo simples, conseguiu reproduzir bem características importantes da série como média, desvio padrão e autocorrelação.

No trabalho de Maceira, Penna e Damázio (2006) foi formulado modelo autorregressivo periódico PAR(P), um modelo multivariado que utiliza de diferentes características da série histórica para que assim sejam geradas as séries sintéticas. Uma peculiaridade desse modelo é que ele não utiliza nenhum tipo de transformação, sendo aplicado diretamente na série de vazões afluentes. Nesse trabalho o modelo foi aplicado na série histórica das Usinas de Água Vermelha e Três Marias, pertencentes às cascatas dos Rios Grande e São Francisco, respectivamente. Os dados utilizados são as vazões mensais entre os janeiro de 1931 e dezembro de 2003, sendo gerado 200 séries sintéticas com 70 anos de simulação cada. Os resultados obtidos conseguiram reproduzir bem importantes características da série como a média, desvio padrão e a autocorrelação, o modelo produziu secas mais severas do que as presentes na série histórica. Em trabalhos mais recentes, os modelos AR vem sendo utilizados como parâmetro de comparação dos resultados com novas propostas.

Já Ochoa-Rivera, García-Bartual e Andreu (2002) propuseram um modelo de redes neurais para geração de séries sintéticas das vazões de dois reservatórios localizados na bacia do rio Tangu, na Espanha, e compararam com o modelo AR(2) os erros obtidos na média, desvio padrão, assimetria, valores máximos e mínimos, capacidade dos reservatórios, entre outras

características. Concluiu-se que o modelo de redes neurais se mostrou uma opção promissora para geração de séries sintéticas.

Outro tipo de modelo que também é bastante difundido são os modelos de simulação baseados nas técnicas de Monte Carlo. Denaxas et al. (2015) apresentaram um modelo que permite o ajuste para geração de séries sintéticas utilizando múltiplos regimes e discretização no tempo por meio da técnica de Monte Carlo via Cadeia de Markov (MCMC). No trabalho o modelo de múltiplos regimes foi aplicado aos dados de produção de energia eólica por minuto na região do Texas, nos Estados Unidos, e seus resultados comparado ao mesmo modelo com regime único. Demonstrou-se que o método de geração de séries sintética para múltiplos regimes é mais eficiente do que o modelo de apenas um regime.

Houve também em Papaefthymiou e Klockl (2008) uma proposta de um modelo MCMC para geração de série sintética para velocidade de ventos, com dados obtidos do *Center for Energy Efficiency and Renewable Energy* (CEERE) que correspondiam a medidas de médias de dez minutos da velocidade do vento ao longo de dois anos. Os resultados apresentados pelos autores mostram que o modelo conseguiu reproduzir de forma consistente tanto função a densidade de probabilidade quanto a função de autocorrelação da série original.

Um terceiro exemplo desse tipo de método aplicado para geração de séries sintéticas é o elaborado em Nfaoui, Essiarab e Sayigh (2004). Nele os autores aplicam o método MCMC para gerar série sintética para velocidade média dos ventos por hora na cidade de Tangiers, no Marrocos. É utilizado uma matriz 12×12 (ou seja, a velocidade média dos ventos foi separada em 12 categorias distintas) que se mostrou a melhor opção entre as testadas. Os resultados obtidos conseguiram reproduzir bem as principais estatísticas descritivas (média e desvio padrão) e a distribuição densidade de probabilidade da série, porém o mesmo não ocorreu quanto a função da autocorrelação. O autor justifica esse fato devido a natureza intrínseca presente no processo de Markov.

Um outro conjunto de técnica de geração de séries sintéticas difundida atualmente são os modelos de inteligência computacional. O trabalho de Salgado et al. (2012) propuseram um modelo de redes neurais multicamadas (MLP) e um sistema baseado em regras fuzzy (FIS) para simulação de vazão mensal em dois reservatórios localizados na bacia do Rio Grande, no Brasil. Os dados correspondiam ao período de janeiro de 1931 a dezembro de 2008. As séries obtidas foram avaliadas quanto as suas médias, desvios padrão, assimetrias e curtoses. Uma análise gráfica via box plot e q-qplot também foi realizada. Os resultados mostraram que

ambos os modelos conseguiram replicar de forma bastante fiel as principais características da série histórica.

Sousa, Guedes e Saldanha (2015) também é um outro exemplo de trabalho que usa este tipo de técnica. Nele é utilizado um sistema denominado TEMPVAZ, onde através de uma base com dados de 146 cascatas afluentes contendo 83 anos sobre dados de chuvas. Utilizando um algoritmo evolucionário que busca uma aproximação não linear dos dados, obteve-se um conjunto de 600 anos cujos resultados mostraram uma boa representação da tendência e sazonalidade da série estudada.

Outro trabalho que utiliza modelos de inteligência computacional para geração de séries sintéticas é o de Evsukoff, De Lima e Ebecken (2011), cuja técnica principal utilizada é um modelo baseado em regra fuzzy (FIS). Ele teve como objetivo principal produzir séries sintéticas para a vazão diária de três bacias do rio Jordão, localizadas estado do Paraná, Brasil: Segredo, Salto Santiago e Salto Osório. Além do histórico da vazão do rio, utilizou-se também os dados históricos da precipitação diárias desses locais no período de 1996 à 2001. Além do modelo FIS, outros três foram gerados: ARMA, SMAP e redes neurais multicamadas (MLP). Gerada as séries, foram calculados os seguintes erro: *Root Mean Square Error* (RMSE), *Mean Absolute Percentage* (MAPE), coeficiente de Nash-Sutcliffe e o coeficiente de Nash-Sutcliffe com valores logaritmos. Comparando os erros, o modelo FIS se mostrou mais eficiente em todos os postos avaliados.

A fim de melhorar os resultados obtidos em modelos de predição, muitos trabalhos recentes usam a combinação entre esses modelos, técnica essa conhecida na literatura como método *ensemble*. Essa técnica tem como condição necessária e suficiente que o modelo combinado tenha um desempenho melhor que todos os modelos individuais (HANSEN; SALAMON, 1990). Atualmente já existem algoritmos de aprendizagem de máquinas cujos resultados são obtidos através de combinações de modelos feitos internamente no código, como o *Random Forest* e o *XGBoost* (FRIEDMAN, 2001; BREIMAN, 2001). Essa metodologia apesar de ser bastante difundida quando se tratando de previsões, como nos casos dos trabalhos de Yokoyama e Chiang (2012), Zhang et al. (2013), Zhang et al. (2015), Hassan, Khosravi e Jaafar (2015), em geração de série sintética é uma metodologia pouco utilizada. Em Jia e Culver (2006) três técnicas geração de séries sintéticas de vazão foram combinadas: redes neurais artificiais via bootstrap (BANN), *maintenance of variance extension* (MOVE) e *drainage area ratio* (DAR). Para comparar a performance dessas técnicas, foi conduzido um caso de estudo utilizando da-

dos da *Buck Mountain Run*, nos Estados Unidos. O *ensemble* entre esses modelos conseguiu satisfazer a condição suficiente dessa técnica.

2.2 SÉRIE DE VAZÕES

A vazão é uma representação do volume de água que atravessa uma seção transversal ao escoamento na unidade de tempo, ou seja, é a rapidez com o qual o volume e/ou massa escoam. Na literatura, essa representação é conhecida como vazão volumétrica, sendo escrita matematicamente conforme a Equação (1), onde V é o volume de água, t o tempo e Q a vazão volumétrica. A unidade de medida utilizada neste caso é o metro cúbico por segundo (m^3/s).

$$Q = \frac{V}{t} \quad (1)$$

No setor elétrico, adota-se o termo vazão natural, que se refere a vazão que ocorreria em uma seção do rio, se não houvesse as ações antrópicas na sua bacia contribuinte, tais como regularizações de vazões realizadas por reservatórios, desvios de água, evaporação dos reservatórios e usos consuntivos (irrigação, criação de animal, e abastecimento rural, urbano e industrial). A vazão natural é obtida por meio de um processo de reconstituição, que considera a vazão observada no local e as informações relativas às ações antrópicas na bacia (ONS, 2015). De acordo com Tucci, Porto e Barros (1995), são diversos fatores que influenciam no comportamento da vazão, sendo os mais importantes: o relevo, a cobertura da bacia, a duração e intensidade da precipitação, o tipo e natureza do solo e o nível de umidade presente nele.

O movimento das águas provavelmente tenha sido a primeira fonte de energia utilizada para diminuir o esforço no trabalho de homens e animais. Não se sabe exatamente quando a roda d'água foi inventada, mas sabe-se que os sistemas de irrigação existem há pelo menos 5.000 anos (REIS; SOUZA; JORGE FILHO, 2011). Diversos métodos eram utilizados na irrigação, um deles era o sistema conhecido como "roda Persa", que consistia numa grande roda em madeira com conchas em sua região periférica. Esta roda era instalada à beira do rio ou riacho de forma que pudesse se movimentar sobre seu eixo, impulsionada pelo movimento da água em contato com as conchas. Este movimento fazia com que a água dentro das conchas fossem elevadas (STANO JÚNIOR; BITENCOURT; TIAGO FILHO, 2007).

Com o passar dos tempos, notou-se que o movimento da roda poderia ser usado para outras finalidades, surgindo assim as primeiras rodas d'água que aproveitavam o movimento da água para movimentar moinhos d'água, diminuindo assim o esforço humano e de animais para moagem de grãos. A roda d'água evoluiu ao longo do tempo e inspirou o desenvolvimento de uma grande quantidade de máquinas hidráulicas que fazem a conversão de energia mecânica da água em outras formas de energia, como por exemplo as turbinas hidroelétricas usadas para geração de energia (GODFREY et al., 2004).

Por definição, as séries de vazões são uma série temporal, ou seja, um conjunto de observações realizadas em um período de tempo sucessivo no tempo, em que as observações vizinhas estão correlacionadas (MORETTIN; TOLOI, 2006). Estas observações podem ser coletadas em diferentes espaços de tempos, seja diária, semanal, mensal ou anual. As séries históricas são obtidas através de um processo baseado em medições e análises. Mais detalhes podem ser vistos em Fortunato et al. (1990).

No Brasil, a geração de energia hidroelétrica começou em 1889 com a usina hidroelétrica construída na cidade de Juiz de Fora, no estado de Minas Gerais, denominada usina Bernardo Mascarenhas, em homenagem ao seu idealizador, também foi a primeira hidroelétrica da América do Sul (Figura 2). Devido a grande capacidade hidrográfica do país, o aumento do uso desse tipo de energia se deu muito rápido, mas somente em 1937 que o país superou os 1.000 MW de capacidade de instalada de geração, e a partir de 1960 que a expansão de usinas teve um forte ritmo de crescimento, para atender ao desenvolvimento econômico da época. De 1970 a 2016, o Brasil passou de 11 GW de capacidade para 150,4 GW, com taxa de crescimento de 5,8% ao ano, indicador superior ao do PIB, que é de 3,5% ao ano.

Figura 2 – Usina Bernardo Mascarenhas, 1889, Juiz de Fora-MG



Fonte: ACURI, 2015.

Atualmente, o sistema hidrelétrico do Brasil é formado por seis grandes bacias que incluem: Paranaíba, Grande, Tietê, Paranapanema, Iguaçu e Paraguai. As usinas podem ser de três tipos: com reservatório, fio d'água ou por bombeamento. Nas usinas com reservatório ocorre a regularização do rio, ou seja, durante os períodos de cheia, além do fluxo da água utilizada para a produção de energia, ocorre acúmulo de água no reservatório. Durante os períodos secos, a água acumulada, além da decorrente do fluxo natural, é utilizada para gerar energia. Por isso, usinas com reservatórios de regularização têm como característica importante o fato que a produção de energia é mais constante (fator de capacidade elevado). Já as usinas a fio d'água possuem reservatório com dimensão reduzida, insuficiente para permitir a regularização do rio, ou seja, a produção de energia é inconstante dependendo da variação da vazão do rio. Nos períodos de cheia a usina produz muita energia e nos de seca a produção é bastante reduzida. Toda a água que chega na usina é utilizada para produção de energia, sem haver acumulação nos períodos de cheia, pois o reservatório não tem capacidade de acúmulo. A produção de energia é inconstante e a usina apresenta um baixo fator de capacidade (GRUPOS DE ESTUDOS TAPAJÓS, 2015). As usinas por bombeamento são aquelas que podem gerar energia elétrica, através da queda d'água de um reservatório localizado em nível mais elevado para outro em nível mais baixo, ou armazenar água em nível mais elevado, através do bombeamento da água

de um reservatório mais baixo para outro mais elevado (LIMA, 2014). Atualmente, o sistema hidrelétrico brasileiro possui 69 usinas com reservatório, 85 a fio d'água e 4 por bombeamento. Além disso, há mais 5 usinas em processo de construção até o momento (ONS, 2017).

Como dito anteriormente na Seção 1, no planejamento a longo prazo das usinas hidrelétricas brasileiras geralmente a quantidade limitadas de dados costuma ser um empecilho. Entre as soluções, uma bastante difundida recentemente são as séries sintéticas, que serão tratadas com mais detalhes na seção seguinte.

2.3 SÉRIES SINTÉTICAS

Os modelos de séries sintéticas são utilizados essencialmente quando a quantidade de dados presentes não são suficientes para que possa ser feita uma análise precisa sobre esse conjunto. Este tipo de técnica é utilizada em muitos trabalhos de diversas áreas do conhecimento como no estudo de precipitações diárias ou mensais (SEMENOV, 2008; BRISSETTE; KHALILI; LECONTE, 2007; SRIKANTHAN et al., 2005), velocidade e força de ventos (DENAXAS et al., 2015; SUOMALAINEN et al., 2012; BROKISH; KIRTLEY, 2009; PAPA-EFTHYMIU; KLOCKL, 2008), vazões mensais ou diárias de rios (SALGADO et al., 2012; MULLER; KAVISKI, 2007; NIJSSEN et al., 1997). Em todos esses casos, o uso das séries sintéticas foi de suma importância para o planejamento em suas respectivas áreas.

De acordo com Wilks (1999), uma série sintética é o resultado da aplicação de um método específico em um conjunto de dados observados, objetivando gerar sequências equiprováveis e que preservem as mesmas características estocásticas dos dados originais como a média, desvio padrão, assimetria, autocorrelação, entre outras.

Numa série sintética, os parâmetros estatísticos do modelo são obtidos a partir da série histórica. Os modelos de séries sintéticas podem ser de dois tipos: regressivos ou não regressivos. Os modelos regressivos são aqueles que levam em consideração a influência de períodos passados, enquanto os não regressivos consideram apenas a componente aleatória. No caso de séries sintéticas de vazões, esses modelos também podem ser cíclicos (geram séries com influências sazonais) ou acíclicos (séries sem influências sazonais, geralmente são séries anuais).

Quanto aos modelos não regressivos, esses são geralmente baseados na identificação da função distribuição de probabilidade que melhor se ajusta aos dados históricos presentes. Já nos modelos regressivos, devido a dependência temporal, é necessário o uso de métodos

estocásticos para que a correlação temporal e estrutura de dependência da série histórica seja mantida (SILVA; PORTELA, 2010).

2.4 TÉCNICAS UTILIZADAS

Nesta subseção será apresentada a base teórica dos modelos utilizados para geração de série sintética nesse trabalho. Serão exploradas as seguintes técnicas: modelo autorregressivo (AR), máquina de vetores e suporte (SVM), redes neurais multicamadas (MLP), *Random Forest* (RF) e SynTise.

2.4.1 Modelo autorregressivo (AR)

O modelo autorregressivo é um dos mais tradicionais e usuais no que tange a geração de séries sintéticas. Desenvolvido por Box e Jenkins (1976), os modelos autorregressivos são processos onde o valor da série em certo período do tempo é uma função dos seus valores anteriores mais um erro. Esse processo é caracterizado pelo parâmetro p , que é a ordem da função. Um modelo $AR(p)$ pode ser definido conforme a Equação (2), onde y_t representa o valor de saída no tempo t , a_i são parâmetros estimadas do modelo, a_0 é uma constante e ε_t é um erro independente que possui distribuição normal padrão.

$$y_t = a_0 + \sum_{i=1}^p a_i y_{t-i} + \varepsilon_t \quad (2)$$

Um modelo $AR(1)$ por exemplo, onde y_t representaria a vazão média do mês no t em função do valor com um período de atraso, seria representado conforme a Equação (3).

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t \quad (3)$$

Ou seja, y_t é uma função de uma porção de y_{t-1} mais um erro. Essa relação também pode ser expressa pela Equação (4), onde φ_1 é a porção da vazão no tempo $t - 1$ carregada sobre a vazão no tempo t , L é o operador de atraso que pode ser expresso de uma forma geral como

$L^b y_t = y_{t-b}$, onde b seria o número total de atrasos.

$$(1 - \varphi_1 L)y_t = c + \varepsilon_t \quad (4)$$

É importante ressaltar também que os modelos AR têm como condição que os dados de entrada sejam estacionários, ou seja, que a série se desenvolve no tempo aleatoriamente ao redor de uma média constante, refletindo alguma forma de equilíbrio estável (MONTGOMERY; JOHNSON; GARDINER, 1990).

2.4.2 Máquinas de vetores e suporte (SVM)

Fundamentado na teoria de aprendizagem estatística, o método conhecido como Máquina de Vetores de Suporte foi desenvolvido em Vapnik (1995) com o intuito de resolver problemas de classificação de padrões, que utiliza-se em seu aprendizado a minimização do risco estrutural (*structural risk minimization* (SRM)). Segundo Haykin (2007), a máquina de vetores e suporte é uma outra categoria das redes neurais alimentadas adiante, ou seja, redes cujas saídas dos neurônios de uma camada alimentam os neurônios da camada posterior, não ocorrendo a realimentação. Essa técnica possui uma alta capacidade de generalização e robustez quando confrontada com dados de grande dimensão, uma vez que não tende a ficar super (*overfitting*) ou sub (*underfitting*) ajustada, porém possui uma velocidade menor de treinamento se comparada a outras técnicas, dada sua alta complexidade computacional, sensibilidade e dificuldade na escolha dos parâmetros além de dificuldade de interpretação do conhecimento adquirido.

Entretanto, diversos estudos foram feitos para diminuir essas limitações, o que somado à sua alta capacidade de generalização, tornando o SVM uma técnica de aprendizado de máquinas bastante explorada e estudada (SMOLA; SCHÖLKOPF, 2004).

2.4.3 Redes neurais multicamadas (MLP)

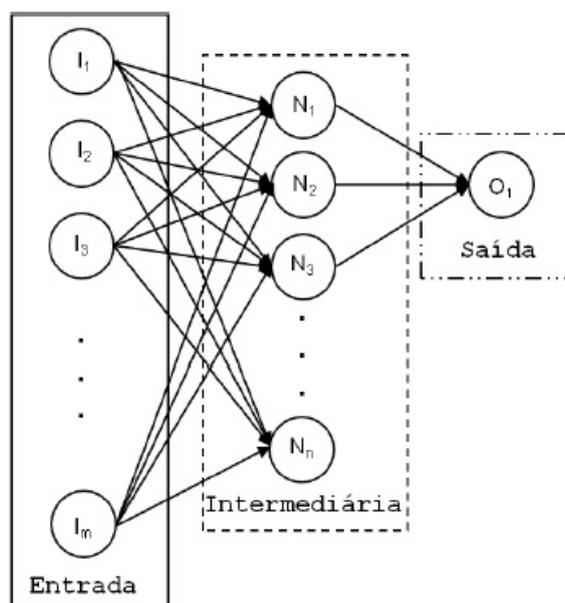
Redes neurais artificiais (RNAs) constituem sistemas de processamento distribuído paralelo que, por serem baseados no funcionamento do sistema nervoso biológico, possuem propensão natural de armazenar conhecimento experimental (HAYKIN, 2007). Embora existam

diversas arquiteturas de redes neurais, a arquitetura multicamadas, mais conhecida na literatura como *multilayer perceptron* (MLP), é o mais difundido em previsão de séries temporais (ZHANG; QI, 2005; AHMED; SARMA, 2006; OTHMAN; NASERI, 2011; SALGADO et al., 2012).

As redes neurais multicamadas são caracterizadas pela presença de múltiplas camadas de unidades básicas de processamento do tipo *Perceptron* (GUYON et al., 1991). O incremento de mais camadas neurais implica em um aumento da capacidade de processamento não linear e generalização da rede neural, superando assim as limitações do modelo *Perceptron* e Adaline apresentados em Minsky e Papert (1969). Esse tipo de modelo possui treinamento supervisionado, sendo que o algoritmo de treinamento visa ajustar os pesos da rede para, a partir de um conjunto conhecido de vetores de entrada, obter-se da rede o conjunto de saídas (AGUIAR, 2010).

Uma rede MLP com uma única camada escondida, geralmente é formada por n neurônios de entrada, h neurônios na camada escondida, e m neurônios na camada de saída conectados por meio de pesos sinápticos que devem ser ajustados durante o período de treinamento. Para o processo de ajuste dos pesos é utilizado o método de aprendizado conhecido como *back-propagation* (SALGADO et al., 2012). A Figura 3 ilustra esse bem esse modelo, onde I_i são todas as entradas do modelo, N_i todos neurônios na camada intermediária e O_1 a saída.

Figura 3 – Configuração da rede MLP



Fonte: SALGADO et al., 2012, p. 168.

2.4.4 *Random forest*

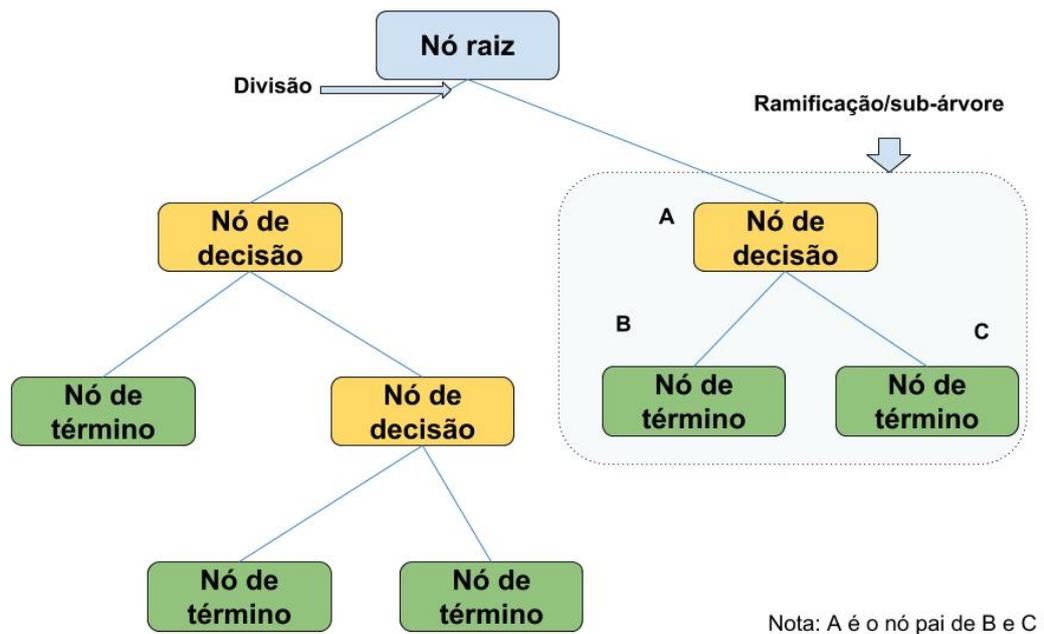
Antes de falar sobre o *Random Forest* (RF) é necessário entender o que são algoritmos de aprendizagem baseados em árvores de decisão. Árvore de decisão é um tipo de algoritmo de aprendizagem supervisionada (ou seja, possui uma saída pré-definida) muito utilizada em problemas de classificação. Ele funciona tanto para as variáveis categóricas quanto contínuas de entrada e saída. Na árvore de decisão, a população ou amostra é dividida em dois ou mais conjuntos homogêneos (ou sub-populações) com base nos divisores/diferenciadores mais significativos das variáveis de entrada. As árvores de decisão podem ser de dois tipos: de variável categórica (também chamada de árvore de classificação) ou de variável contínua (mais conhecida como árvore de regressão) (VOOO-INSIGHTS, 2016).

A terminologia básica utilizada em árvores de decisão é a seguinte (ver também a Figura 4 como forma de complemento):

- (a) **nó raiz:** representa a população inteira ou amostra, sendo ainda dividido em dois ou mais conjuntos homogêneos;
- (b) **divisão:** é o processo de dividir um nó em dois ou mais sub-nós;

- (c) **nó de decisão:** quando um sub-nó é dividido em sub-nós adicionais;
- (d) **folha ou nó de término:** os nós não divididos são chamados folha ou nó de término;
- (e) **poda:** o processo de remover sub-nós de um nó de decisão é chamado poda. Pode-se dizer que é o processo oposto ao de divisão;
- (f) **ramificação/sub-árvore:** uma sub-seção da árvore inteira é chamada de ramificação ou sub-árvore;
- (g) **nó pai e nó filho:** um nó que é dividido em sub-nós é chamado de nó pai. Os sub-nós são os nós filhos do nó pai.

Figura 4 – Terminologias básicas de uma árvore de decisão



Fonte: VOOO-INSIGHTS, 2016.

O *Random Forest* é um modelo de aprendizagem de máquina por *ensemble* bastante versátil, visto que empreende métodos de redução dimensional, trata valores em falta, valores anômalos, entre outras etapas essenciais na exploração de dados. Sendo capaz de executar tarefas de regressão e classificação baseado em combinações de árvores de decisão, que pode tanto ser usado para classificação quanto regressão. Trabalha com uma floresta de árvores formadas através de entradas aleatórias ou combinadas, com uma precisão comparada ao da técnica AdaBoost (BREIMAN, 2001).

No *Random Forest* ao invés de uma árvore, crescem múltiplas árvores, onde cada dá uma classificação, que é como uma espécie de "voto" para essa classe. Então é escolhida a classificação com mais votos entre todas as árvores da floresta e, no caso de regressão, toma-se a média das saídas.

A técnica funciona da seguinte maneira (VOOO-INSIGHTS,):

- (a) seja um conjunto de treinamento de tamanho N . Uma amostra desses N casos é escolhida aleatoriamente com reposição, sendo esse o conjunto de treinamento para o cultivo da árvore;
- (b) dado M variáveis de entrada, um número $m < M$ é especificado de modo que em cada nó, m variáveis de M sejam selecionadas aleatoriamente. A melhor divisão nestes m é utilizada para dividir o nó. O valor de m é mantido enquanto a floresta cresce;
- (c) cada árvore é cultivada na maior extensão possível e não há poda;
- (d) novas previsões são realizadas agregando as árvores (ou seja, votos majoritários para classificação, média para regressão).

Esse modelo possui grandes vantagens como: resolver problemas de classificação e regressão, conseguir lidar com dados em grande volume e com muitas dimensões, possuir um método eficaz para estimar dados faltantes sem grande perda de precisão, possuir métodos para equilibrar os erros em conjuntos de dados com classes desequilibradas, a estimativa de erro remove a necessidade de um conjunto de testes extra. Entretanto, existem também desvantagens nessa técnica, como: não ser tão boa em problemas de regressão, podendo sobre-ajustar o modelo em conjuntos de dados que tenham muita discrepância, possuir uma grande dificuldade para seleção dos melhores parâmetros, além de ser uma espécie de "caixa preta", pois o usuário tem pouco controle sobre o que faz o modelo.

Nesse trabalho, todas as parâmetros de programação serão feitas no software R (R Core Team, 2016). Entre os diversos parâmetros presentes neste software para a execução deste método, os seguintes serão explorados:

- (a) *n*: número total de árvores geradas;
- (b) *m*: número mínimo de nós;

- (c) *nperm*: número de vezes em que o erro *out-of-bag*(OOB) será permutado por árvore para avaliar a importância da variável;
- (d) *corr.bies*: total de correções do viés de regressão feito nas árvores.

2.4.5 SynTise

O SynTise é um método desenvolvido por Denaxas et al. (2015) para geração de séries sintéticas de velocidade e força dos ventos que utiliza modelos de Monte Carlo via Cadeia de Markov, também conhecidos como métodos MCMC.

Uma Cadeia de Markov é um processo estocástico onde a probabilidade de qualquer comportamento futuro do processo, quando o seu estado atual é conhecida exatamente, não é alterada pela conhecimento adicional sobre seu comportamento passado. Em outras palavras, o sistema não possui "memória" que lhe possibilite usar informações quanto ao seu comportamento atual. A Equação (5) formaliza matematicamente essa situação.

$$\mathbb{P}[Y_{t+1} = y_{t+1} | Y_0 = y_0, \dots, Y_t = y_t] = \mathbb{P}[Y_{t+1} = y_{t+1} | Y_t = y_t] = \mathbb{P}(y_{t+1}, y_t) \quad (5)$$

Esse tipo de processo estocástico é conhecido como processo sem memória, visto que o passado não influencia na probabilidade futura. As probabilidades $\mathbb{P}[Y_{t+1} = y_{t+1} | Y_t = y_t]$ são denominadas probabilidades de transição, e para representar todas as informações sobre os estados de transição, a notação matricial é bastante utilizada, como mostrada em (6), onde $\mathbb{P}_k(e_1, e_1)$ representa a probabilidade de estar no estado e_1 em t e continuar nesse estado em $t + 1$, $\mathbb{P}_k(e_1, e_2)$ a probabilidade de estar no estado e_1 em t e passar para estado e_2 em $t + 1$ e assim respectivamente.

$$\mathbb{P}_k = \begin{pmatrix} \mathbb{P}_k(e_1, e_1) & \mathbb{P}_k(e_1, e_2) & \mathbb{P}_k(e_1, e_3) & \cdots \\ \mathbb{P}_k(e_2, e_1) & \mathbb{P}_k(e_2, e_2) & \mathbb{P}_k(e_2, e_3) & \cdots \\ \mathbb{P}_k(e_3, e_1) & \mathbb{P}_k(e_3, e_2) & \mathbb{P}_k(e_3, e_3) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (6)$$

As Cadeias de Markov onde os valores passados tem influência são conhecidas na lite-

ratura como Cadeia de Markov de Alcance Variável (RISSANEN, 1983).

3 MATERIAL E MÉTODOS

Nesta seção será apresentada e analisada a base de dados utilizada nesse trabalho, também será explicado detalhadamente como é realizada a geração das séries sintéticas através das técnicas propostas na Seção 2.4 e por fim são mostradas as métricas utilizadas para avaliar e comparar cada uma das propostas sugeridas.

3.1 DADOS DE VAZÕES

Dentre as bacias hidrográficas presentes no país, a bacia do Grande mostrada na Figura 5) possui um total de quinze usinas hidrelétricas instaladas, sendo seis delas com reservatório e nove a fio d'água. Entre todas as usinas bacia, o banco de dados que será utilizado neste trabalho para aplicação do modelo proposto para geração de séries sintéticas são referentes as séries mensais de vazões das usinas de Água Vermelha e Volta Grande, no período de janeiro de 1931 à dezembro de 2010, totalizando 960 observações. Os dados são do Operador Nacional do Sistema Elétrico (ONS).

Figura 5 – Diagrama das usinas hidroelétricas pertencentes a bacia do Rio Grande



Fonte: Adaptado de ONS, 2017.

A usina hidrelétrica de Água Vermelha (Figura 6) foi construída em 1973, sendo concluída em 1979, faz parte da bacia do Rio Grande a 80 km da confluência com o rio Paranaíba (AES TIETE, 2014). Localizada na cidade de Oureste-SP, ela é uma usina com reservatório que possui 1396,2MW de potência instalado.

Figura 6 – Usina hidrelétrica de Água Vermelha

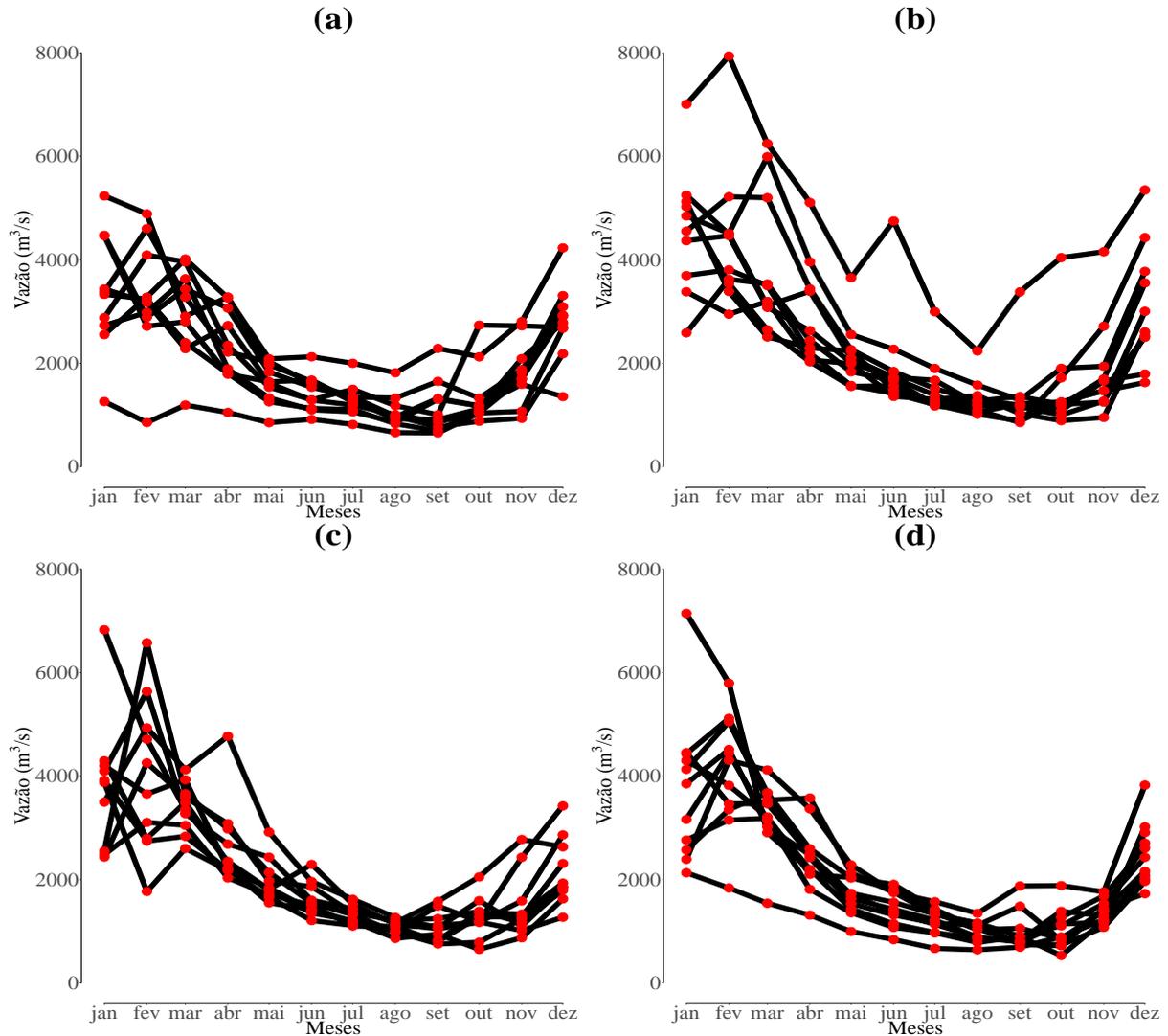


Fonte: AES BRASIL, 2016.

Separando a série histórica da usina de Água Vermelha por décadas no período de 1980 à 2010, conforme as Figuras 7a, 7b, 7c e 7d, vê-se que em todos os anos o comportamento de um modo geral é o mesmo, com a série decrescendo de fevereiro até agosto, voltando a crescer até dezembro, esse tipo de comportamento mostra um forte indício da presença de sazonalidade na série temporal como um todo. Os valores das séries mensais são em média os mesmos para as quatro décadas. Analisando cada uma individualmente, na Figura 7a nota-se que durante a década de 70 houve um ano onde as vazões tiveram valores praticamente constantes até o mês de agosto, onde após isso a série voltou apresentar o seu comportamento natural. Na Figura 7b é visto que nos anos 80 ocorreu um ano a vazão foi bem maior que todos os outros nos doze meses. No mês de março ocorreram três anos com vazões acima dos valores médios das demais décadas. Quanto a Figura 7c, mostra uma variação maior das vazões do mês de fevereiro quando comparada com as demais décadas. Outros dois pontos que chamam atenção nesse gráfico são que apresentam valores bem acima dos demais no mês de janeiro e abril. Por fim, na Figura 7d os meses de janeiro e fevereiro apresentam pontos nos quais a vazão destoa dos demais anos. Também destaca-se nesse gráfico a presença de um ano no qual a vazão foi abaixo da média

dos demais entre janeiro e julho, mesmo tendo apresentado o mesmo comportamento da série ao longo desse ano.

Figura 7 – Série histórica dividida por décadas da usina de Água Vermelha

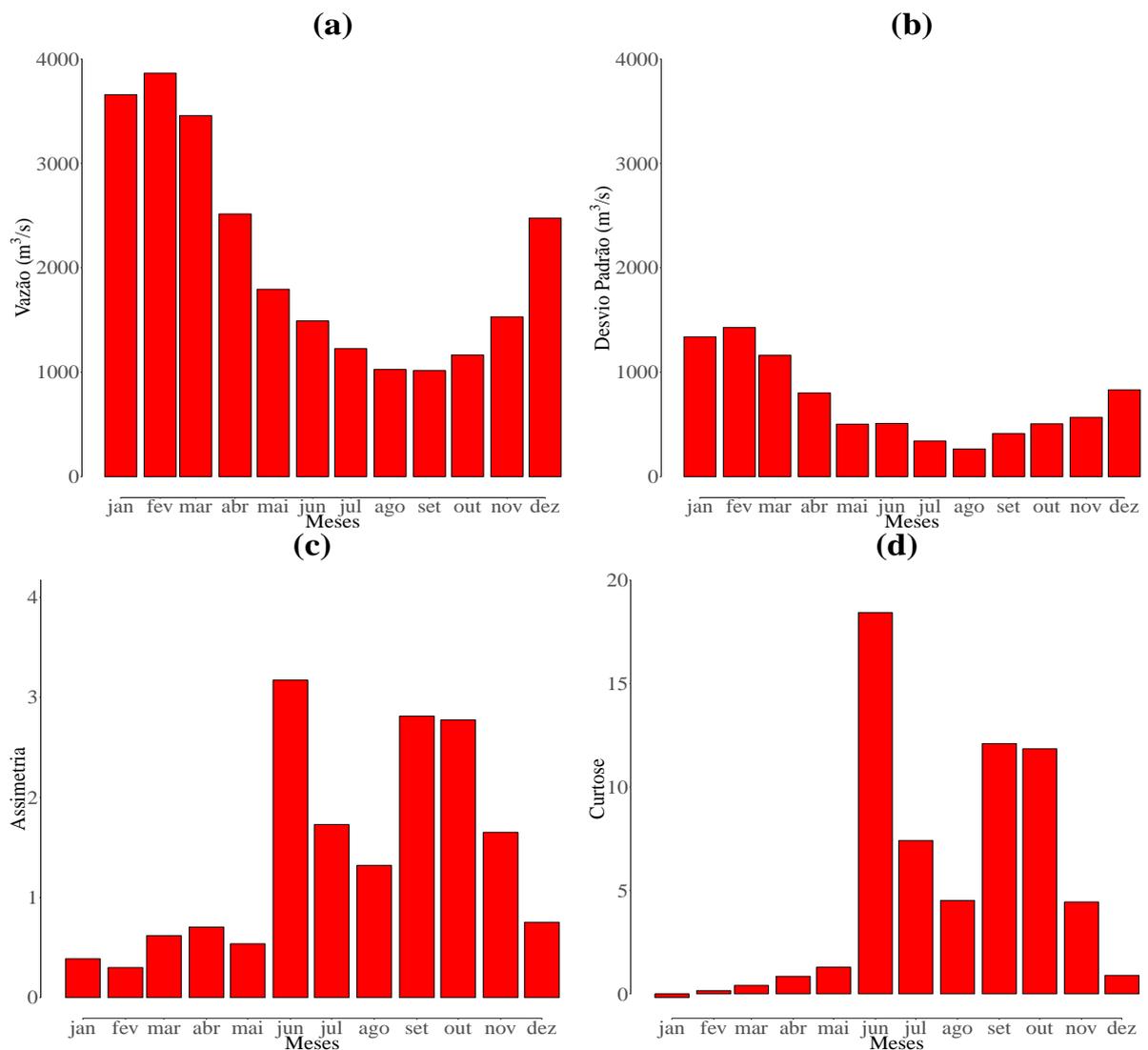


Legenda: a) Década de 70
 b) Década de 80
 c) Década de 90
 d) Década de 2000

Analisando as médias mensais das vazões da usina de Água Vermelha, através da Figura 8a, nota-se que o valor médio das vazões decresce entre fevereiro e agosto, sendo crescente a partir do mês de setembro. Nota-se também que os períodos de cheia acontecem predominantemente nos meses do verão (dezembro, janeiro e fevereiro), enquanto as secas ocorrem majoritariamente no inverno (junho, julho e agosto). Pela Figura 8b, verifica-se que o desvio padrão para esse conjunto de dados segue o mesmo comportamento da média, sendo menor nos períodos de seca e maior no de cheia. Entretanto, no que tange a assimetria desse conjunto

de dados (ver Figura 8c) janeiro, fevereiro e maio possuem uma assimetria positiva moderada, enquanto os demais meses essa assimetria positiva é mais significativa. Quanto a curtose (ver Figura 8d), ela pode ser considerada platicúrtica em janeiro e fevereiro (ou seja, a distribuição desses dados é mais achatada que a distribuição normal padrão nesses dois meses) e leptocúrtica nos demais meses (ou seja, a distribuição desses dados é mais afunilada que a distribuição normal padrão nos demais meses), em especial nos meses de junho, setembro e outubro, que possuem uma curtose acima de 10.

Figura 8 – Estatísticas descritivas mensais da série de vazões mensais da usina de Água Vermelha



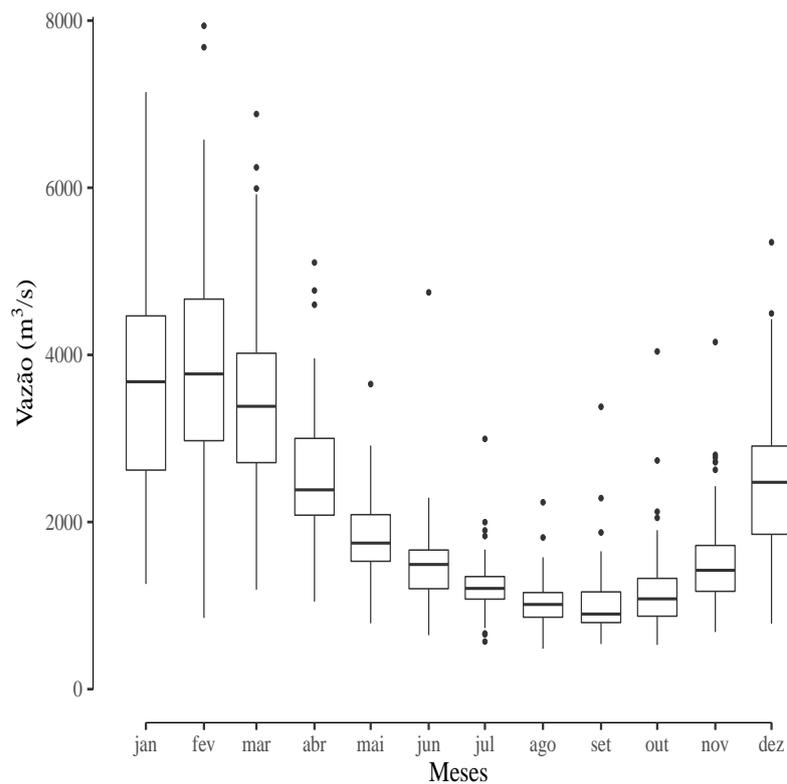
Legenda: a) Média
 b) Desvio Padrão
 c) Assimetria
 d) Curtose

Fonte: Elaborada pelo autor.

Pelos box-plots das vazões da Figura 9 percebe-se a ocorrência de *outliers* superiores em

todos os meses, com exceção de janeiro. Há também a presença de *outliers* inferiores no mês de julho, algo que não é muito comum na maioria das séries de vazões e deverá ser considerado na análise dos resultados apresentados nas técnicas a seguir. Quanto a amplitude presente nos meses também é algo a ser considerado, nota-se que os meses de cheias são os que possuem maior amplitude, enquanto nos meses de seca elas são menores, especialmente em julho.

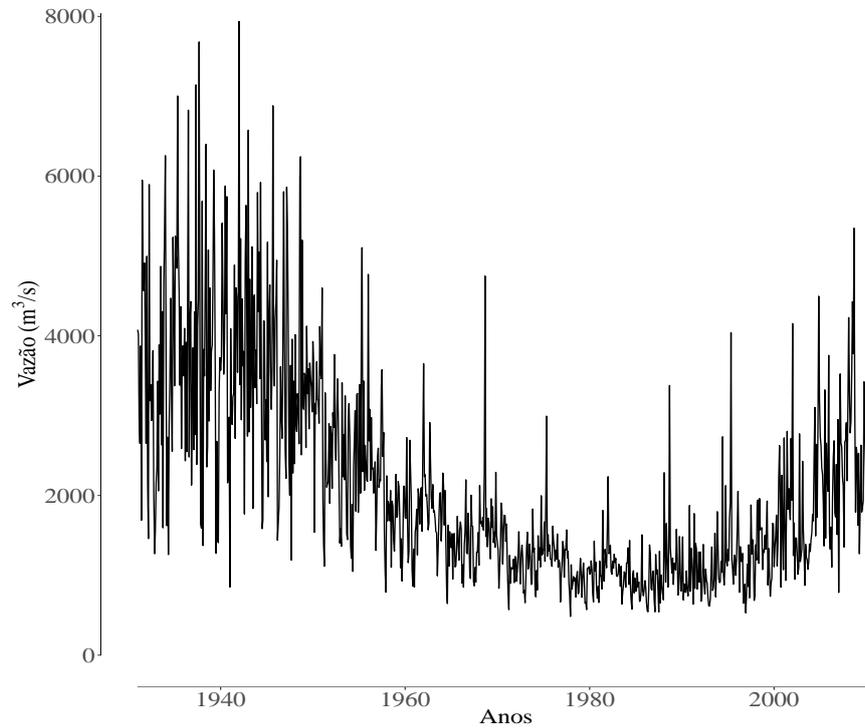
Figura 9 – Box-plots das vazões mensais da usina de Água Vermelha



Fonte: Elaborada pelo autor.

Já o gráfico da série histórica de vazões mensais da usina de Água Vermelha (ver Figura 10) mostra que como suspeitava-se, a série é sazonal, porém essa sazonalidade aparenta diminuir a partir de 1960. Também há presença de uma tendência negativa na série.

Figura 10 – Série histórica da usina de Água Vermelha



Fonte: Elaborada pelo autor.

O outro conjunto de dados a ser utilizado nesse trabalho é pertencente a usina hidrelétrica de Volta Grande (Figura 11) foi construída em 1974, as margens do Rio Grande, localizada na cidade de Conceição das Alagoas-MG. Ela é uma usina a fio d'água (ou seja, não possui reservatório) que tem 380 MW de potência instalada.

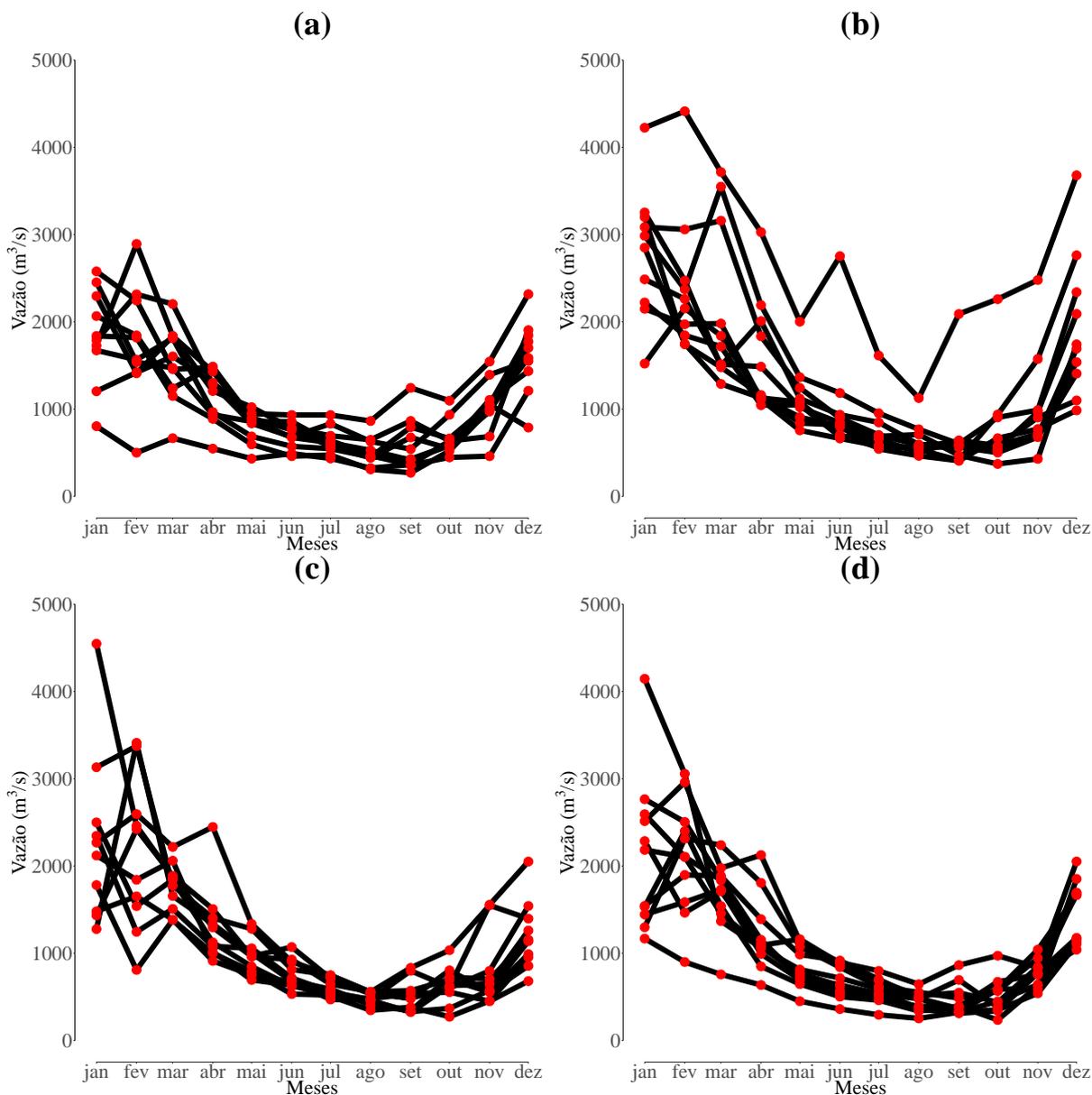
Figura 11 – Usina hidrelétrica de Volta Grande



Fonte: CEMIG, 2013.

Separando a série histórica da usina de Volta Grande por décadas no período de 1980 à 2010, conforme as Figuras 12a, 12b, 12c e 12d, vê-se que em todos os anos o comportamento de um modo geral é o mesmo, com a série decrescendo de fevereiro até agosto, voltando a crescer até dezembro, esse tipo de comportamento mostra um forte indício da presença de sazonalidade na série temporal como um todo. Os valores das séries mensais são em média os mesmos para as quatro décadas. Analisando cada uma individualmente, na Figura 7a nota-se que durante a década de 70 houve um ano onde as vazões tiveram valores praticamente constantes até o mês de agosto, onde após isso a série voltou apresentar o seu comportamento usual. Também aconteceu um valor que se destacou no mês de fevereiro por ser o maior da década. Na Figura 7b é visto que nos anos 80 ocorreu um ano a vazão foi bem maior que todos os outros nos doze meses. No mês de março ocorreram três anos com vazões acima dos valores médios das demais décadas. Quanto a Figura 7c, mostra uma variação maior das vazões do mês de fevereiro quando comparada com as demais décadas. Outros três pontos que chamam atenção nesse gráfico são que apresentam valores bem acima dos demais no mês de janeiro (que foi a maior vazão da década, bem acima das demais), abril e novembro. Por fim, na Figura 7d o mês de janeiro apresentou um ponto cuja vazão destoa dos demais anos, sendo o bem maior que os outros da mesma década, um efeito parecido ao ocorrido no da década de 90. Também destaca-se nesse gráfico a presença de um ano no qual a vazão foi abaixo da média dos demais entre janeiro e julho, mesmo tendo apresentado o mesmo comportamento da série ao longo desse ano.

Figura 12 – Série histórica dividida por décadas da usina de Volta Grande

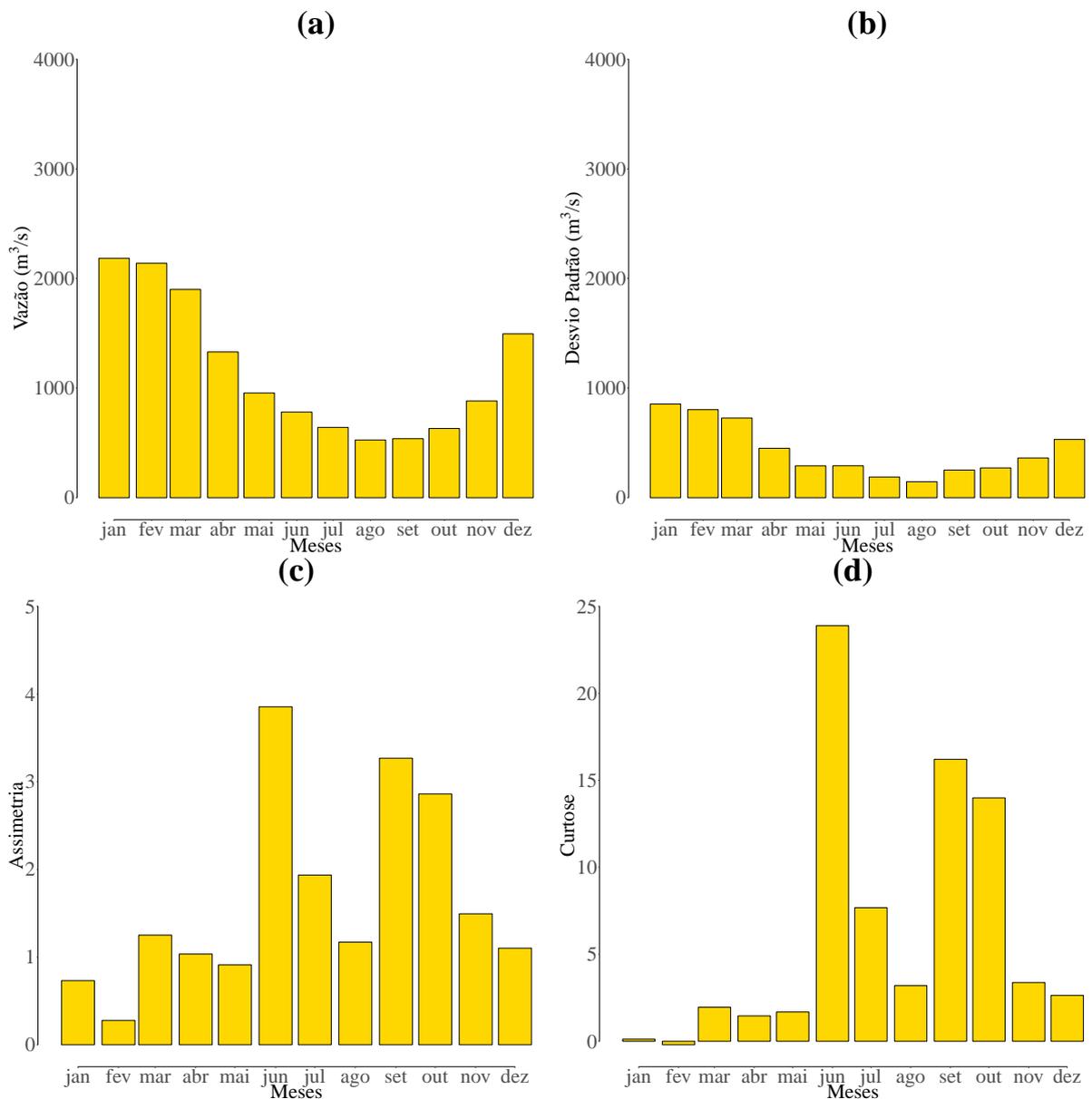


Legenda: a) Década de 70
 b) Década de 80
 c) Década de 90
 d) Década de 2000

Analisando as médias mensais das vazões da usina de Volta Grande (ver Figura 13ca), nota-se que o valor médio das vazões decresce entre fevereiro e agosto, voltando a crescer a partir do mês de setembro. Nota-se também que as cheias acontecem predominantemente no verão (dezembro, janeiro e fevereiro), já as secas ocorrem majoritariamente no inverno (junho, julho e agosto). Pelas estatísticas descritivas apresentadas na Figura 13b, nota-se que o desvio padrão para esse conjunto de dados segue o mesmo comportamento da média, com valores

menores nos períodos de seca e maiores no de cheia. Entretanto, no que tange a assimetria desse conjunto de dados (ver Figura 13c) metade dos meses possuem uma assimetria positiva moderada, no caso janeiro, fevereiro, março, abril, maio, junho e dezembro. Já nos demais meses, essa assimetria positiva é mais significativa. Enquanto na curtose (ver Figura 13d), ela é platicúrtica em janeiro e fevereiro e leptocúrtica nos demais meses em especial nos meses de junho, setembro e outubro, que possuem uma curtose acima de 10.

Figura 13 – Estatísticas descritivas mensais da usina de Água Vermelha

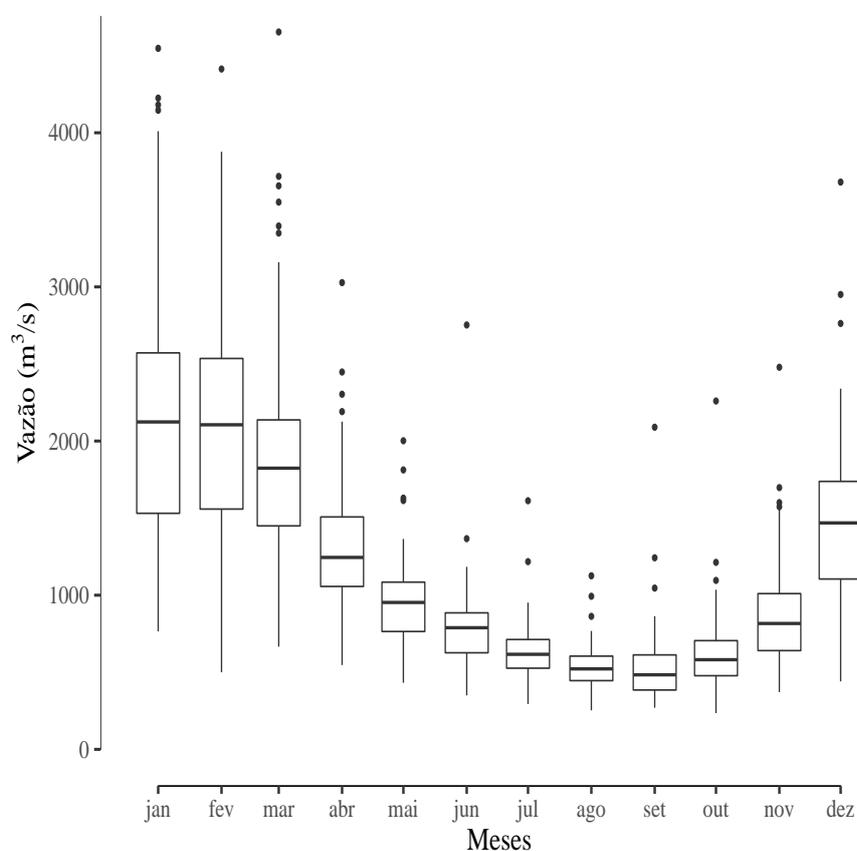


Legenda: a) Média
b) Desvio Padrão
c) Assimetria
d) Curtose

Fonte: Elaborada pelo autor.

Quanto aos box-plots dessa série, através da Figura 14, nota-se que os períodos de cheia ocorrem predominantemente nos meses do verão (dezembro, janeiro e fevereiro), já as secas acontecem majoritariamente no inverno (junho, julho e agosto). Nessa figura é possível perceber a ocorrência de *outliers* superiores em todos os meses. Não há presença de *outliers* inferiores nessa série. Quanto a amplitude, assim como no posto de Água Vermelha, nota-se que os meses de cheias possuem maior amplitude, enquanto nos meses de seca elas são menores, porém a menor amplitude nesse caso ocorre em agosto.

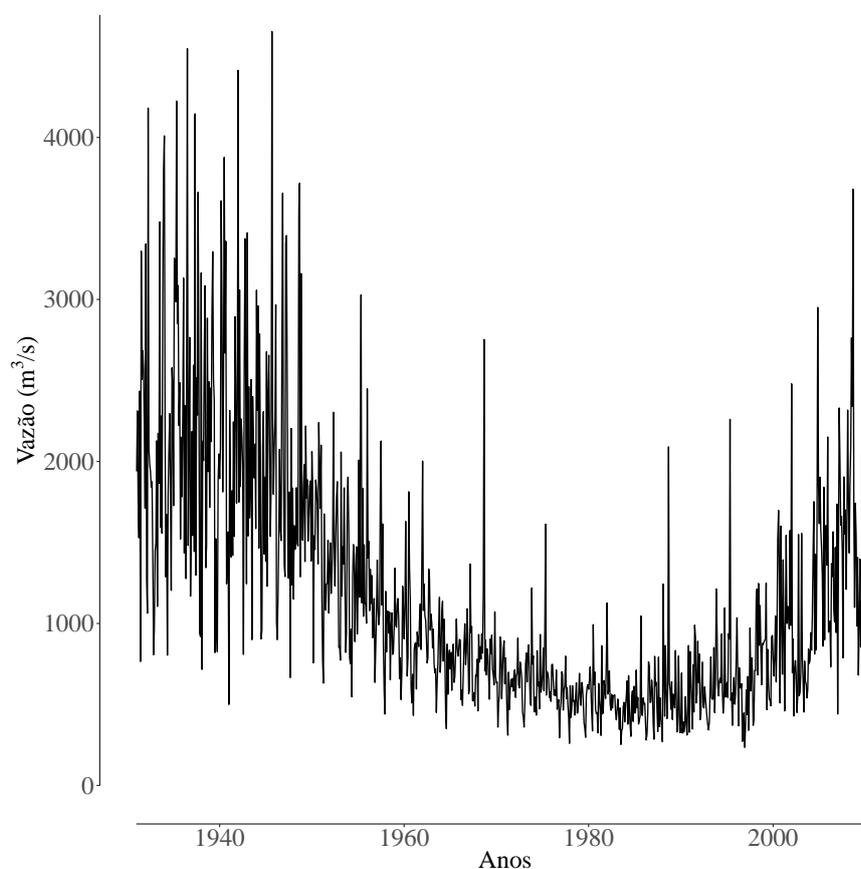
Figura 14 – Box-plots das vazões mensais da usina de Volta Grande



Fonte: Elaborada pelo autor.

Já gráfico da série histórica de vazões da usina de Volta Grande da Figura 15 mostra que a série é sazonal, como esperado, porém essa sazonalidade aparenta diminuir à partir de 1960. Também há a presença de uma tendência negativa na série. O comportamento é bastante semelhante ao posto de Água Vermelha.

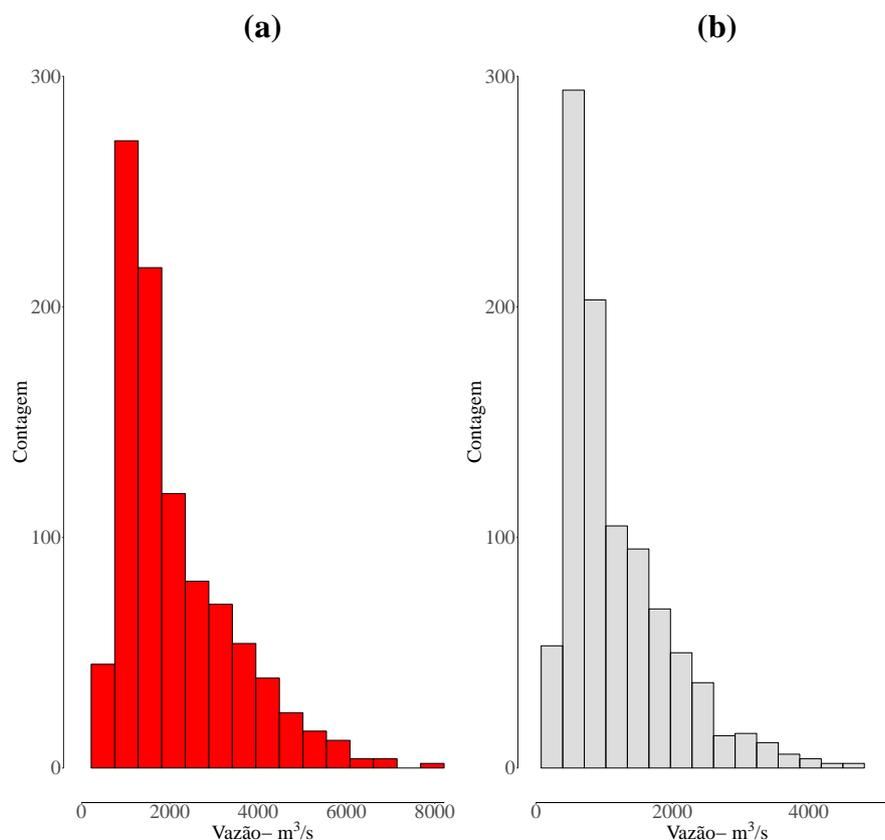
Figura 15 – Série histórica da usina de Volta Grande



Fonte: Elaborada pelo autor.

As séries de vazões como um todo geralmente são assimétricas devido às cheias e secas. Esse fato pode ser visto claramente nos histogramas das Figuras 16a e 16b, que mostram uma assimetria positiva nesses conjuntos de dados. Outro fato importante que verifica-se nesses gráficos é o comportamento semelhante de ambos os postos, mesmo sendo uma usina com reservatório e a outra fio d'água.

Figura 16 – Histograma da série vazões mensais



Legenda: a) Usina de Água Vermelha
b) Usina de Volta Grande

Fonte: Elaborada pelo autor.

Todas essas características da série serão importantes no que tange o processo de geração das séries sintéticas quanto nas métricas a serem utilizadas para sua avaliação.

3.2 GERAÇÃO DE SÉRIES SINTÉTICAS

Conforme dito na Seção 2.3, uma série sintética é o resultado da aplicação de um método específico em um conjunto de dados observados, objetivando gerar sequências equiprováveis e que preservem as mesmas características estocásticas dos dados originais como a média, desvio padrão, assimetria, autocorrelação, entre outras (WILKS, 1999). Para gerar um modelo de série sintética, é necessário saber como é composto uma série temporal, além disso, um pré-processamento dos dados também é importante de ser feito devido as características assimétricas geralmente presentes em séries de vazões mensais. Esses detalhes serão descritos a seguir.

Modelo de série temporal: Um modelo de série temporal é representado por dois componentes, um sazonal $s_{m,t}$ e outro estacionário $z_{m,t}$, além de um componente aleatório normal $a_{m,t}$ com média zero e variância constante σ_a^2 (t representa o ano e m o mês), como mostra a Equação (7)

$$y_{m,t} = s_{m,t} + z_{m,t} + a_{m,t} \quad (7)$$

Pré-processamento dos dados: Devido aos picos cheias e secas que acontecem na vazão de rios durante a série histórica, as séries de vazões mensais acabam por gerar um comportamento sazonal, e essa característica não pode ser desprezada quando afeta os resultados do modelo. A fim de ajustar o problema a essa característica, deve-se realizar uma transformação nos dados para que se remova o componente sazonal. Nesse trabalho, a solução encontrada foi dada pela Equação (8), onde $\hat{\mu}_m$ e $\hat{\sigma}_m$ são a média e o desvio padrão estimado para cada mês m , respectivamente.

$$\tilde{y}_{m,t} = \frac{y_{m,t} - \hat{\mu}_m}{\hat{\sigma}_m} \quad (8)$$

Portanto, o modelo de série temporal passa a ser representado pela Equação (9), onde $\tilde{z}_{m,t}$ e $\tilde{a}_{m,t}$ são os novos valores para o componente estacionário e aleatório da série, respectivamente, e $\tilde{y}_{m,t}$ o novo valor da série histórica após a transformação para retirada da sazonalidade na série.

$$\tilde{y}_{m,t} = \tilde{z}_{m,t} + \tilde{a}_{m,t} \quad (9)$$

Conseqüentemente, os valores estimados no momento geração da série sintética são $\tilde{z}_{m,t}$ e $\tilde{a}_{m,t}$. Para a geração da série de vazões sintéticas mensais são determinados doze modelos independentes, ou seja, as séries serão geradas utilizando somente a série histórica de cada mês.

Nas subseções seguintes serão descritos como esses componentes são estimados, sendo que existem técnicas com diferentes abordagens que serão discutidas nas Seções 3.2.1.1 e 3.2.1.2. Quanto ao SynTise, exibido na Seção 2.4.5, será apresentado em uma subseção à parte de seu singularidade. Também será mostrado um fluxograma exemplificando como todo processo é feito.

3.2.1 Modelos para geração baseado em séries temporais e em inteligência artificial

Esta subseção será focada nos modelos de série temporal e inteligência artificial (IA). Nas subseções seguintes serão descritos de quais maneiras serão gerados os componentes estacionário e o aleatório, ou seja, as variáveis $\tilde{z}_{m,t}$ e $\tilde{a}_{m,t}$ da Equação (9), respectivamente.

3.2.1.1 Técnicas para componente estacionária

Para o cálculo do componente estacionário (CE) é necessário a identificação do modelo para o ajuste, que é feito através do conjunto de treinamento.

Para o ajuste em determinado mês é calculado o erro de cada modelo variando o número de atrasos $p = 1, \dots, 11$, onde o valor p que apresentar o menor número índice é o escolhido para representar a série sintética. O método utilizado para o cálculo do erro será definido na Seção 3.3. Os parâmetros das técnicas apresentadas na Seção 2.4 também são definidos pelo mesmo método.

Portanto, após definido o modelo, são empregues para a estimação do CE as seguintes técnicas: modelo autorregressivo (AR), máquina de vetores e suporte (SVM), redes neurais multicamadas (MLP), *random forest* (RF).

3.2.1.2 Modelos para gerar o componente aleatório

Após calculado o CE da série sintética, o próximo passo é a estimação do componente aleatório (CA). O CA é estimado através do conjunto dos resíduos ($e_{m,t}$). Esse por sua vez é dado pela diferença entre o valor observado $\tilde{y}_{m,t}$, no mês m e ano t , e o valor estimado $\hat{y}_{m,t}$ pela técnica apresentada no CE, conforme Equação (10), sendo que $e_{m,t}$ é uma variável aleatória independente entre si para $m = 1, \dots, 12$.

$$\begin{bmatrix} e_{1,t} \\ e_{2,t} \\ \vdots \\ e_{12,t} \end{bmatrix} = \begin{bmatrix} \tilde{y}_{1,t} \\ \tilde{y}_{2,t} \\ \vdots \\ \tilde{y}_{12,t} \end{bmatrix} - \begin{bmatrix} \hat{y}_{1,t} \\ \hat{y}_{2,t} \\ \vdots \\ \hat{y}_{12,t} \end{bmatrix} \quad (10)$$

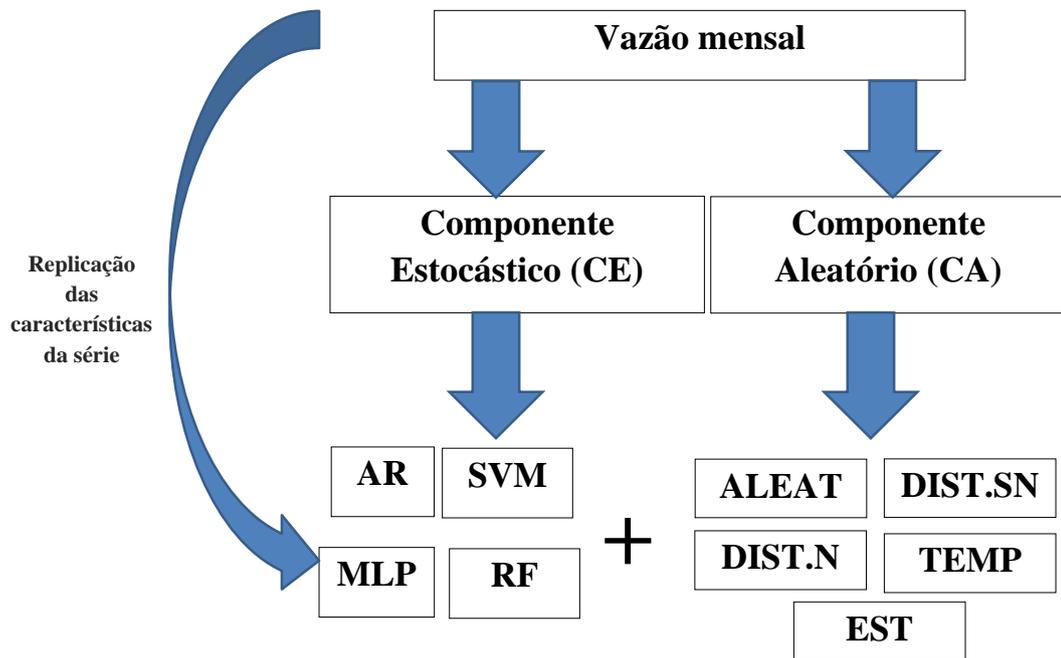
Um dos objetivos principais desse trabalho é o de apresentar novas propostas para ge-

ração do CA. Além do método mais comum, a seleção aleatória dos resíduos (ALEAT), feito em trabalhos como Ochoa-Rivera, García-Bartual e Andreu (2002) por exemplo, outras 4 novas propostas são sugeridas.

- (a) **seleção aleatória dos resíduos (ALEAT)**: a proposta mais usual, nela o resíduo $e_{m,t}$ é sorteado através de uma amostragem aleatória simples com reposição, tornando o valor sorteado CA;
- (b) **seleção através de uma distribuição de probabilidade simétrica (DIST.N)**: nessa proposta, o conjunto de resíduos $e_{m,t}$ é ajustado a uma distribuição de probabilidade normal. Após estimado, o CA é definido através da geração de valores aleatórios dessa distribuição;
- (c) **seleção através de uma distribuição de probabilidade assimétrica (DIST.SN)**: De maneira análoga a proposta anterior, nesse caso o conjunto de resíduos $e_{m,t}$ também é ajustado a uma distribuição de probabilidade, entretanto esse ajuste é feito para uma normal assimétrica (AZZALINI, 1985). Após estimado, o CA é definido através da geração de valores aleatórios dessa distribuição;
- (d) **seleção pela ordem cronológica (TEMP)**: nesta preposição, a seleção no conjunto de resíduos $e_{m,t}$ é feito na ordem cronológica com a qual foi estimado para se tornar o CA. Caso todos os valores dos resíduos tenham sido utilizados, a seleção é reiniciada no primeiro valor do conjunto;
- (e) **seleção através dos resíduos estimados (EST)**: a última proposta é a de se estimar os resíduos utilizando a mesma técnica usada para o cálculo do CE. Desse novo conjunto de resíduo estimado, o valor é selecionando por uma amostragem aleatória simples com reposição para ser o CA.

Calculado o CA, ele é somado ao CE, gerando o valor sintético da vazão. A Figura 17 ilustra todo o processo que foi descrito nesta e nas Seções 3.2 e 3.2.1.1.

Figura 17 – Geração de séries sintéticas



Fonte: Elaborada pelo autor.

Como dito anteriormente, a geração da série sintética pela técnica do SynTise, apresentada na Seção 2.4.5, é feita de uma forma distintas das apresentadas até o momento e será apresentada na seção seguinte.

3.2.2 Modelos baseados em cadeia de Markov

Conforme é dito em Denaxas et al. (2015), o SynTise é um modelo baseado em MCMC que permite uma discretização em 16 classes, denominadas nesse trabalho de regimes, onde cada um representa as estações do ano e as horas dos dias. Para adapta-ló ao problema desse trabalho, o código que define a sua estrutura foi alterada, sendo agora 12 regimes, cada qual relativo aos meses do ano, permitindo assim que o software produza as séries sintéticas para esse tipo de conjunto de dados.

No SynTise, inicialmente é feita uma discretização nos dados originais, onde é definido o número total de estados presentes na Cadeia de Markov, conforme apontado na Seção 2.4.5. Essa discretização é feita separando os dados em classes de maneira análoga a feita nas construções de histogramas, onde cada classe tem a mesma amplitude, sendo a responsável pelo

número de colunas da matriz de transição de probabilidade.

Já a ordem no modelo representa a quantidade de valores passado que terão influência sobre o resultado seguinte, valor de ordem 1 são referentes a Cadeia de Markov clássica, enquanto os valores de ordem superior são referentes as Cadeias de Markov de Alcance Variável. As linhas da matriz de transição aumentam exponencialmente em relação as suas colunas de acordo com a ordem da cadeia, permitindo assim representar todos os passos possíveis para casos de ordem superior a 1.

Os valores da ordem e estado do modelo são definidos antes da geração da série pelo usuário. Para definir qual a melhor estrutura para ser utilizada em cada aplicação, modelos com ordem de 1 até 5 e estados de 4 até 20 foram gerados e comparados quanto ao RMSE das FAC e FDP. Como os dados das vazões mensais são contínuos, a obtenção dos resultados das estimações da série sintética gerada é feita via método de Monte Carlo.

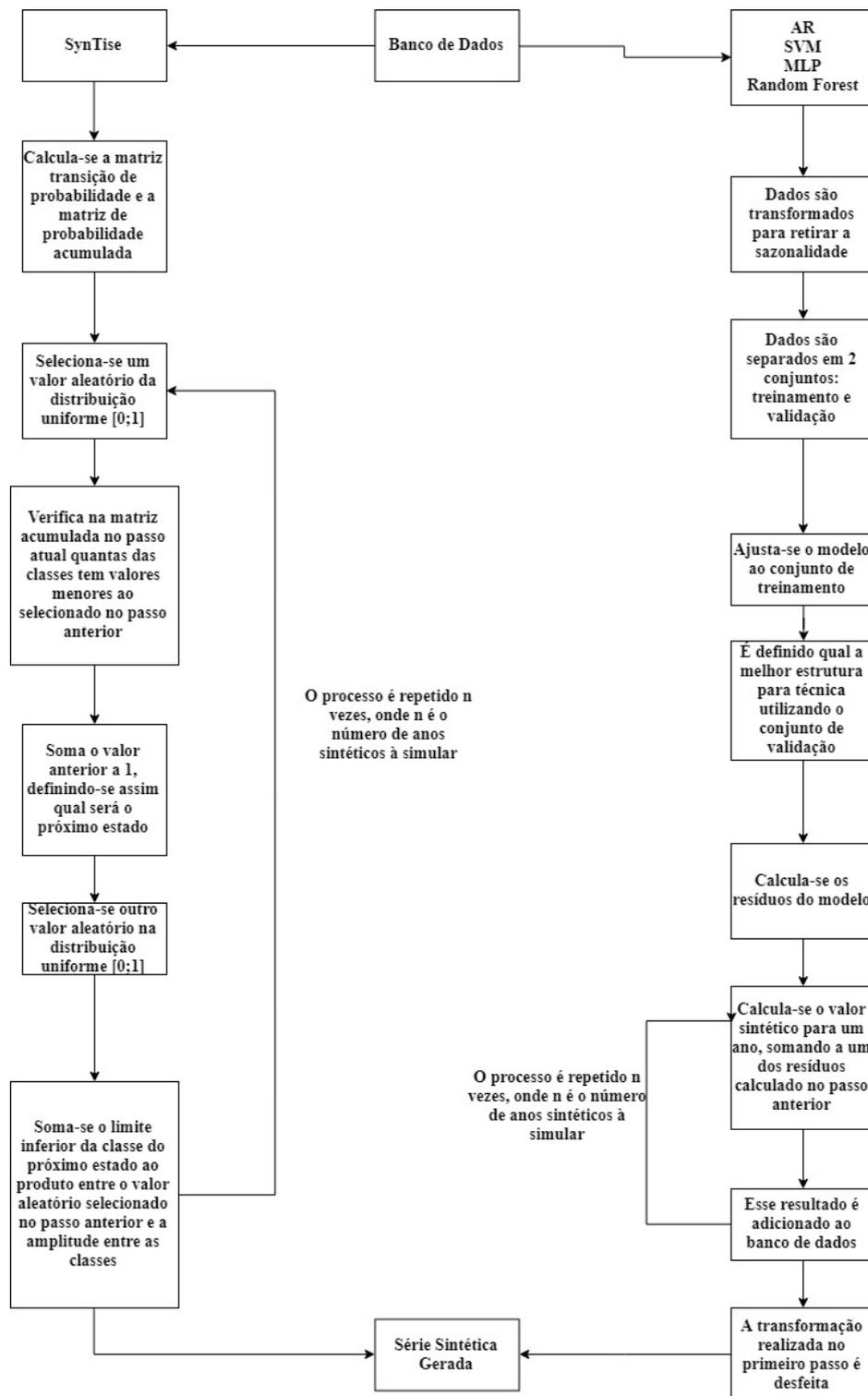
De forma mais detalhada, para o cálculo do valor da série sintética nesse caso, um valor aleatório da distribuição uniforme $[0; 1]$ é sorteado, utilizando a matriz de transição de probabilidade acumulada verifica-se em qual intervalo do passo atual esse valor se encaixará, definindo assim qual será o intervalo do passo seguinte. O valor final gerado é dado pela soma do limite inferior no intervalo o qual será o passo seguinte pelo produto da amplitude da discretização feita por um outro valor aleatório da distribuição uniforme $[0; 1]$. Portanto, o CE da série nesse caso é dado pelo limite inferior do intervalo do passo seguinte, enquanto o CA é o produto da amplitude da discretização feita por um outro valor aleatório da distribuição uniforme $[0; 1]$. Como todo o procedimento é feito baseado na matriz de transição de probabilidade, a transformação descrita na Seção 3.2 não é necessária para este modelo.

O modelo que originalmente foi programado em MatLab, teve todas suas funções originais reprogramadas no software R (R Core Team, 2016) e ajustada para o problema desse trabalho.

3.2.3 Fluxograma do modelo proposto

Portanto, de uma forma geral, pode-se dizer que existem duas possíveis formas de gerar as séries sintéticas nesse trabalho: através do modelo de série temporal e inteligência artificial ou por modelo baseado em cadeia de Markov. O fluxograma da Figura 18 apresenta de maneira resumida esse processo.

Figura 18 – Fluxograma da execução das séries sintéticas utilizadas nos modelos trabalho



Fonte: Elaborada pelo autor.

Nela nota-se que, caso opte por gerar a série sintética utilizando os modelos AR, SVM, MLP ou RF, o seguinte procedimento é realizado:

- (a) realiza-se uma transformação dos dados para retirar a sazonalidade presente na série;
- (b) separa-se os dados em dois conjuntos, sendo um para treinamento, outro para validação do modelo;
- (c) ajusta-se o modelo ao conjunto de treinamento;
- (d) utilizando o conjunto de validação, se escolhe a melhor estrutura para técnica utilizada;
- (e) definido os parâmetros do modelo, calculam-se os componentes aleatórios do modelo ajustado;
- (f) gera-se um valor para o ano seguinte utilizando o modelo ajustado, soma-o a um dos valores de resíduos calculados no passo anterior, sendo esse valor selecionado aleatoriamente;
- (g) esse valor é adicionado ao conjuntos de dados;
- (h) com o novo valor, os passos f) e g) são repetidos de acordo com o número de anos escolhidos pelo usuário para geração da série;
- (i) a transformação feita no passo a) é desfeita.

Entretanto, no caso do SynTise, o passo a passo é feito de distinta, sendo aplicado na seguinte forma:

- (a) seleciona-se um valor aleatória da distribuição Uniforme no intervalo $[0,1]$;
- (b) utilizando a linha da matriz de probabilidade acumulada no passo presente, verifica-se em quantas colunas o valor da probabilidade acumulada é menor que o selecionado anteriormente;
- (c) o valor obtido no passo anterior é somada a 1, sendo guardado numa variável chamada *nextState*;
- (d) um novo valor aleatório da distribuição Uniforme no intervalo $[0,1]$ é selecionado, guardado na variável *scale*;

- (e) por fim, o resultado da simulação é dado através produto entre a variável *scale* e o valor da amplitude entre as classes (chamada de *range*) somado do limite inferior da classe relativa a variável *nextState*;
- (f) esse processo é repetido de acordo com o número de anos desejados para série sintética.

3.3 MÉTRICA DE AVALIAÇÃO

Por fim, é necessário obter-se maneiras de se avaliar a performance dos modelos gerados. Duas importantes características importantes que são comumente avaliadas em diversos trabalhos da área como Montanari, Rosso e Taqqu (1997), Brokish e Kirtley (2009) e Denaxas et al. (2015) são as funções de autocorrelação (FAC) e densidade de probabilidade (FDP).

A FAC representa o valor da correlação entre diferentes períodos de tempo numa série. Ela pode ser definida como razão entre a autocovariância, dada pela Equação (11) e a variância dos dados, conforme mostra Equação (12).

$$c_k = \frac{1}{n-k} \sum_{t=1}^{n-k} (y_t - \bar{y})(y_{t+k} - \bar{y}) \quad (11)$$

$$ACF(k) = \frac{c_k}{c_0} \quad (12)$$

onde n é o total de valores da série, k o número de passos seguintes a t , \bar{y} a média dos valores da série e c_0 o valor de c para $k = 0$.

Para a comparação da FAC da série histórica com os modelos gerados, deve-se definir um valor para o número de passos k . O cálculo desse valor será feito conforme em Ripley (2002) através do cálculo $10 \log(N)$, onde N é o tamanho da série histórica.

Já a FDP, representa a probabilidade da variável assumir certo valor. Uma maneira de se verificar o comportamento da FDP de determinado banco de dados é através da análise do seu histograma. Portanto, para tal comparação, os erros serão calculados entre a média das classes da série sintética e da série histórica. O número de classes a ser utilizado será definido conforme Ferreira (2005), através da fórmula $5 \log(N)$. Nas duas fórmulas apresentadas, resultados não inteiros serão arredondados ao inteiro superior.

Uma medida muito utilizada para avaliar o erro entre a FAC e FDP da série sintética

gerada em comparação a série histórica é o *Root Mean Square Error* (RMSE), dado pela Equação (13). Essa medida será a principal forma utilizada nesse trabalho para fins de comparações entre os modelos.

$$RMSE = \sqrt{\frac{\sum_{t=1}^k (\hat{y}_t - y_t)^2}{k}} \quad (13)$$

onde \hat{y}_t é o valor gerado pela série sintética no tempo t , y_t é o valor da série histórica em t e k o número de passos definido.

O RMSE da FAC é calculado através do erro entre as FACs da série sintética gerada e da série histórica nos k passos que são definidos previamente. Quanto a FDP, para o cálculo do RMSE são geradas um certo número de classes de mesma amplitude para as séries históricas e sintéticas. Após isso, as médias entre os limites de cada classe dessas duas séries são calculadas e os seus dados são realocados nas novas classes. Por fim, o RMSE é calculado sobre a diferença entre a frequência relativa das novas classes das duas séries.

O processo de seleção do melhor modelo entre as técnicas geradas será feito essencialmente utilizando esses erros. A média entre o RMSE da FDP de todos os doze meses será o primeiro critério para decisão, em caso de empate, a média entre RMSE da FAC para os doze meses será utilizado como critério para desempate.

Além dessas funções, outras medidas importantes para avaliação de uma série sintética que é bastante utilizada em outros trabalhos como Ochoa-Rivera, García-Bartual e Andreu (2002), Carta, Ramirez e Velazquez (2009), Salgado et al. (2012) e Detzel et al. (2014) são as estatísticas descritivas dos dados tais como: média, desvio padrão, assimetria e curtose, além dos gráficos de box-plot e q-q plot.

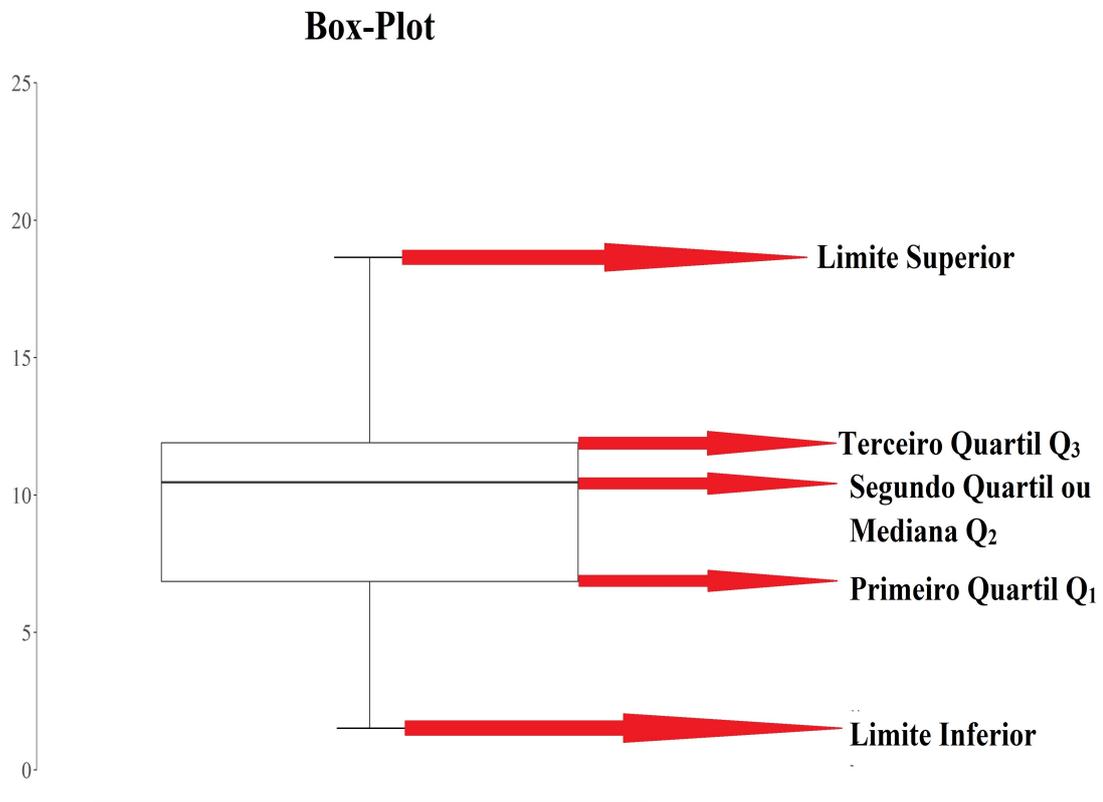
Os box-plots desse trabalho serão formados pelo primeiro e terceiro quartil além da mediana. As hastes inferiores e superiores se estendem, respectivamente, do quartil inferior até o menor valor que não está abaixo do limite inferior e do quartil superior até o maior valor que não está acima do limite superior. Os limites inferior e superior são calculados pelas fórmulas 14 e 15, respectivamente, onde Q_1 é o primeiro quartil dos dados e Q_3 o terceiro quartil.

$$LI = Q_1 - 1,5(Q_3 - Q_1) \quad (14)$$

$$LS = Q_3 + 1,5(Q_3 - Q_1) \quad (15)$$

Os pontos fora destes limites são chamados *outliers* e são denotados por ponto. A Figura 19 mostra um exemplo de como será construído os box-plots.

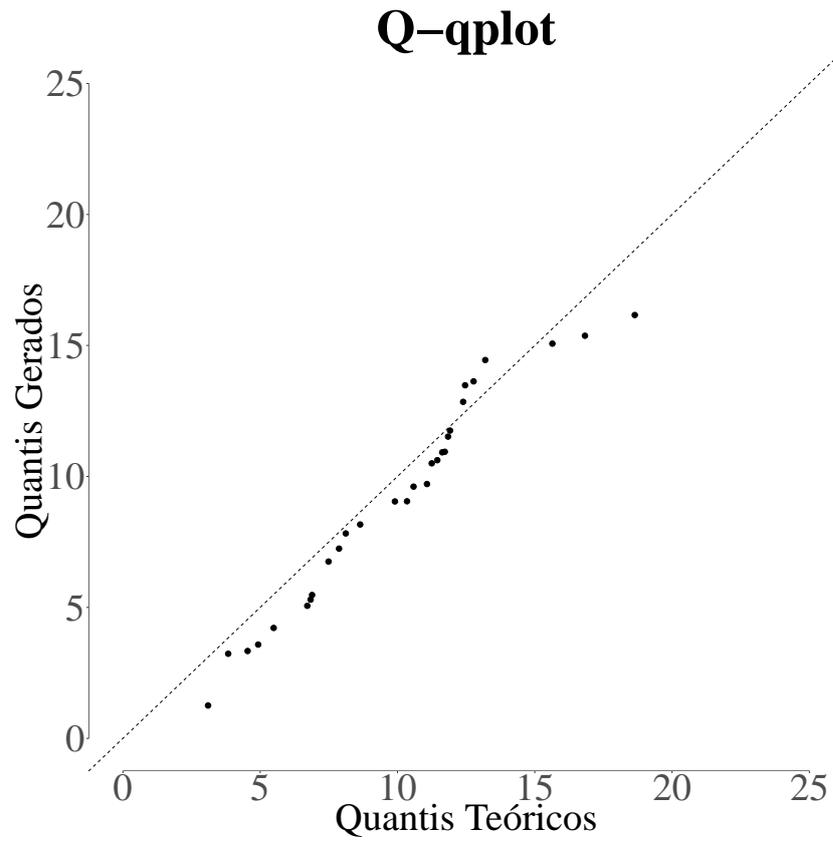
Figura 19 – Exemplo de um box-plot



Fonte: Elaborada pelo autor.

Por fim, o q-qplot é um gráfico no qual verifica-se se a distribuição dos quantis gerados é a mesma do quantil teórico. Como mostra o exemplo da Figura 20, quanto mais próximo os pontos estiverem da reta diagonal, mais perto da distribuição teórica os dados gerados estarão. Essa ferramenta terá bastante utilidade para que seja verificado se a distribuição dos modelos de série sintéticas geradas estarão próximos da série de dados reais.

Figura 20 – Exemplo de um q-qplot



Fonte: Elaborada pelo autor.

4 RESULTADOS

Nesta seção é realizada uma comparação entre os melhores resultados para cada uma das técnicas apresentadas na Seção 2.4 para as usinas de Água Vermelha e Volta Grande. A seleção do melhor tipo de modelagem entre as propostas na Seção 3.2.1 para as técnicas baseadas em inteligência artificial e série temporal podem ser vistas em detalhes nos Apêndices A, B, C, D, F, G, H e I. Todos os modelos que serão apresentados irão gerar um total de 2000 anos em séries sintéticas. Os recursos computacionais empregues também serão exibidos nessa seção.

4.1 RECURSOS COMPUTACIONAIS

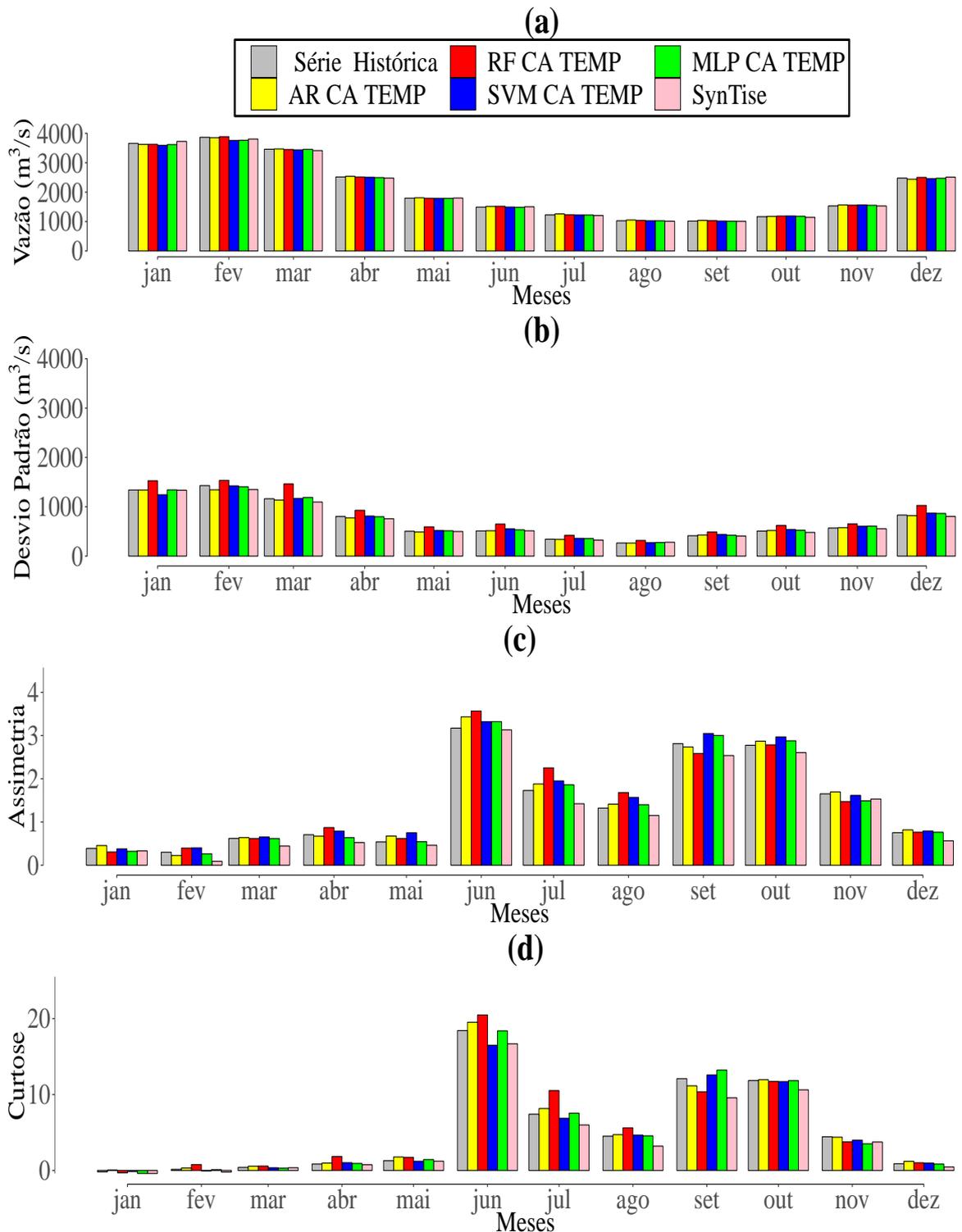
Todas as rotinas de análise dos dados, geração de modelos de séries sintéticas e avaliação dos resultados foram realizadas utilizando o software R versão 3.3.1 (R Core Team, 2016). Os procedimentos foram realizados no Laboratório de Inteligência Computacional (LInC), no CPU Intel(R) Core(TM)2 Quad, com 8GB de memória RAM no sistema operacional Microsoft Windows 10 Pro.

4.2 USINA DE ÁGUA VERMELHA

Para usina de Água Vermelha, entre as modelagens propostas para geração do CA nas técnicas de inteligência artificial e série temporal, em todas elas o CA TEMP obteve os melhores resultados (ver Apêndices A, B, C e D). Os resultados desses modelos serão comparados entre si e o SynTise (ver Apêndice E) a fim de selecionar qual deles obteve os melhores resultados.

Na comparação das estatísticas descritivas desses modelos, nota-se que todos obtiveram valores de médias bem próximos ao da série histórica. (ver Figura 21a). Já quanto ao desvio padrão, o RF acabou por superestimar o valor da série histórica, diferentemente dos demais modelos (ver Figura 21b). Quanto a assimetria e curtose, as Figuras 21c e 21d mostram que o SynTise apesar de conseguir replicar bem essas características, não obteve uma eficiência tão grande quanto o AR, SVM e MLP, principalmente nos meses com valores maiores, como junho, setembro e outubro.

Figura 21 – Comparação entre as estatísticas descritivas das vazões mensais geradas pelos melhores modelos em cada uma das técnicas analisadas para usina de Água Vermelha

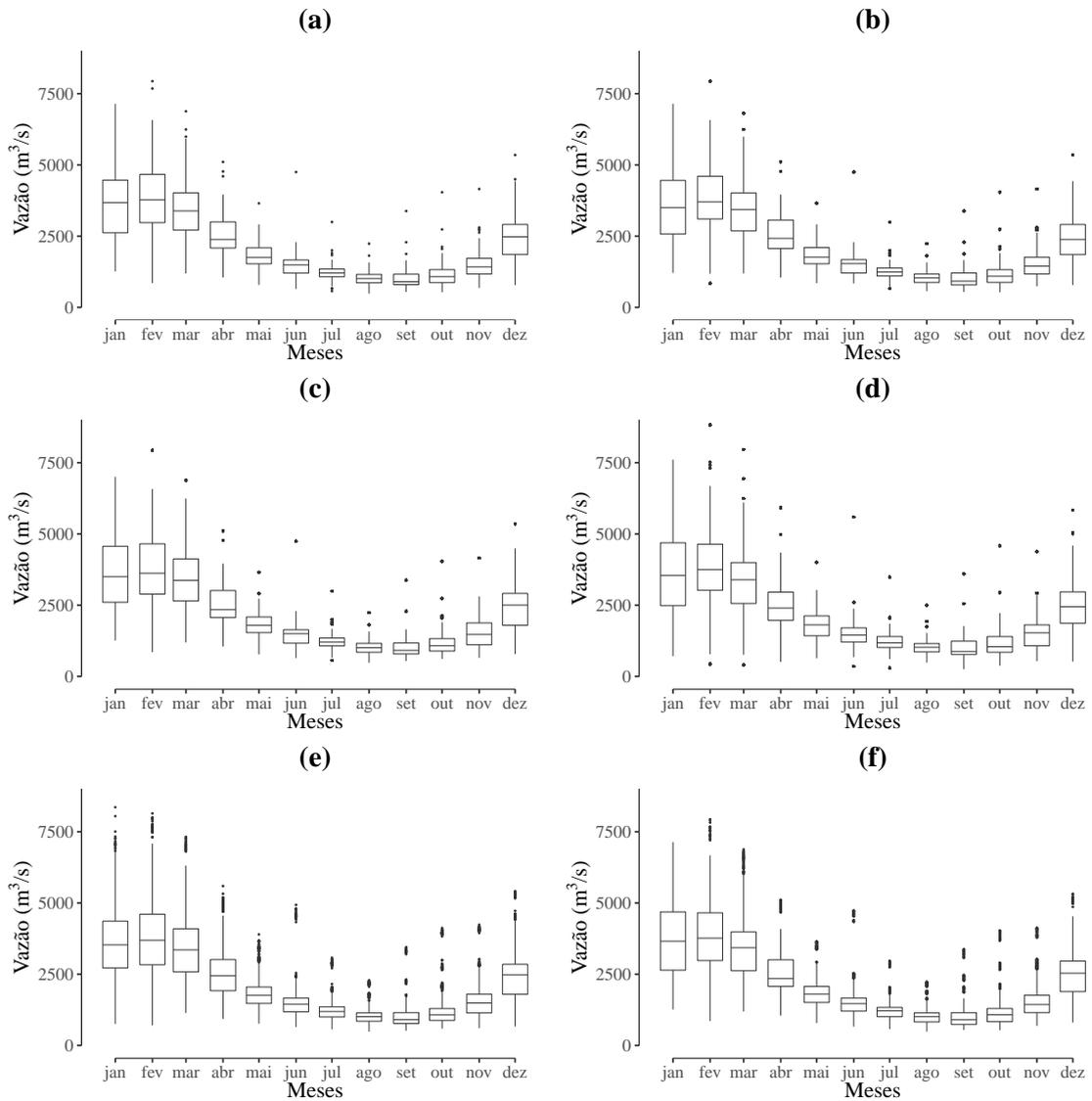


Legenda: a) Média
 b) Desvio Padrão
 c) Assimetria
 d) Curtose

Fonte: Elaborada pelo autor.

Nos box-plots (ver Figuras 22b, 22c , 22d, 22e, e 22f) e q-qplots (ver Figuras 23a, 23b, 23c, 23d, e 23e) apresentados desses modelos, vê-se que o modelo AR apresenta dificuldade quanto a representação dos *outliers* da série histórica, apresentando inclusive um valor negativo no mês de fevereiro, valor esse inexistente na série histórica. Já o SVM e MLP conseguiram replicar muito bem esses valores, inclusive os *outliers* inferiores de julho. O SynTise não apresentou nenhum *outlier* inferior, porém conseguiu representar de maneira eficaz os demais valores. Quanto ao RF, apresentou a mesma dificuldade que o modelo AR quanto a estimativa de *outliers* inferiores, visto que estimou valores para fevereiro, março e junho, meses nos quais não há presença desse tipo de dado, porém conseguiu uma boa representação dos *outliers* superiores e amplitude das séries mensais.

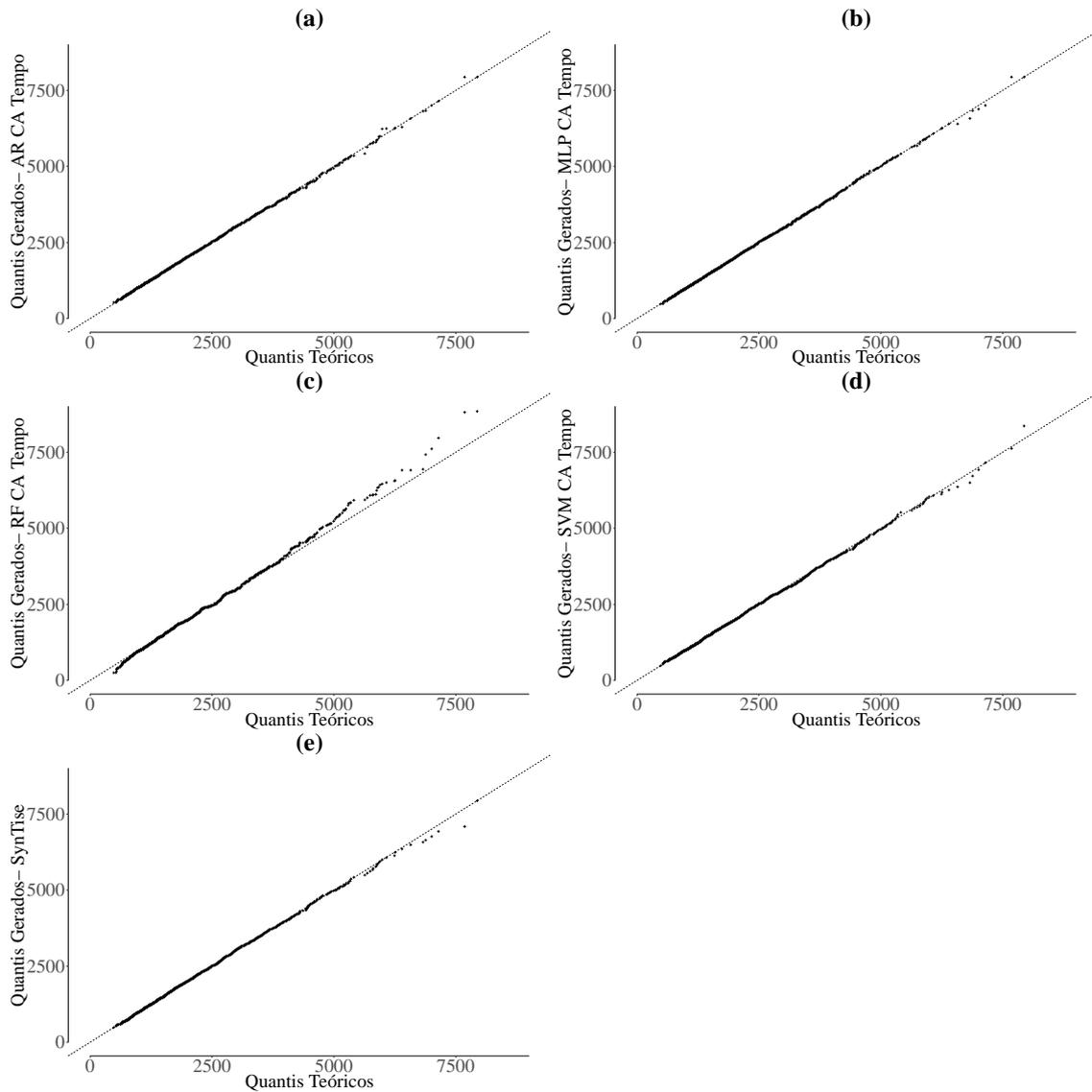
Figura 22 – Comparação entre os box-plots das vazões geradas pelos melhores modelos de cada uma das técnicas analisadas para usina de Água Vermelha



Legenda: a) Série Histórica
 b) AR CA TEMP
 c) MLP CA TEMP
 d) RF CA TEMP
 e) SVM CA TEMP
 f) SynTise

Fonte: Elaborada pelo autor.

Figura 23 – Comparação entre os q-plots das vazões geradas pelos melhores modelos de cada uma das técnicas analisadas para usina de Água Vermelha



Legenda: a) AR CA TEMP
 b) MLP CA TEMP
 c) RF CA TEMP
 d) SVM CA TEMP
 e) SynTise

Fonte: Elaborada pelo autor.

Quanto aos erros da FAC e FDP, pela Tabela 1, nota-se que o erro das FACs dos modelos AR, MLP e RF foram significativamente menores que os demais na maior parte dos meses, quanto a FDP, SynTise e SVM obtiveram erros próximos aos desses modelos, em especial nos meses de junho e maio, enquanto o RF também ficou aquém nesse quesito, obtendo resultados do nível dos demais modelos apenas em abril, maio e julho.

Tabela 1 – RMSE da FAC e FDP mensal dos melhores modelos em cada uma das técnicas analisadas para usina de Água Vermelha

RMSE da FAC					
Mês	AR CA TEMP	RF CA TEMP	SVM CA TEMP	MLP CA TEMP	SynTise
janeiro	0,065	0,061	0,081	0,055	0,079
fevereiro	0,082	0,061	0,095	0,076	0,085
março	0,076	0,056	0,112	0,029	0,113
abril	0,061	0,070	0,081	0,060	0,082
maio	0,047	0,041	0,081	0,046	0,075
junho	0,030	0,064	0,060	0,032	0,069
julho	0,052	0,080	0,069	0,039	0,076
agosto	0,055	0,040	0,110	0,045	0,111
setembro	0,036	0,036	0,102	0,050	0,110
outubro	0,016	0,046	0,085	0,042	0,085
novembro	0,023	0,066	0,081	0,032	0,072
dezembro	0,051	0,047	0,112	0,030	0,097
Média dos meses	0,049	0,056	0,089	0,045	0,088
RMSE da FDP					
Mês	AR CA TEMP	RF CA TEMP	SVM CA TEMP	MLP CA TEMP	SynTise
janeiro	0,021	0,016	0,026	0,016	0,020
fevereiro	0,016	0,014	0,025	0,020	0,019
março	0,013	0,024	0,034	0,017	0,025
abril	0,012	0,019	0,033	0,016	0,018
maio	0,021	0,016	0,036	0,019	0,007
junho	0,017	0,012	0,029	0,008	0,011
julho	0,020	0,013	0,034	0,010	0,027
agosto	0,012	0,011	0,017	0,011	0,021
setembro	0,006	0,009	0,015	0,012	0,021
outubro	0,009	0,012	0,027	0,005	0,016
novembro	0,015	0,014	0,021	0,016	0,017
dezembro	0,013	0,023	0,016	0,014	0,018
Média dos meses	0,015	0,015	0,026	0,014	0,018

Fonte: Elaborada pelo autor.

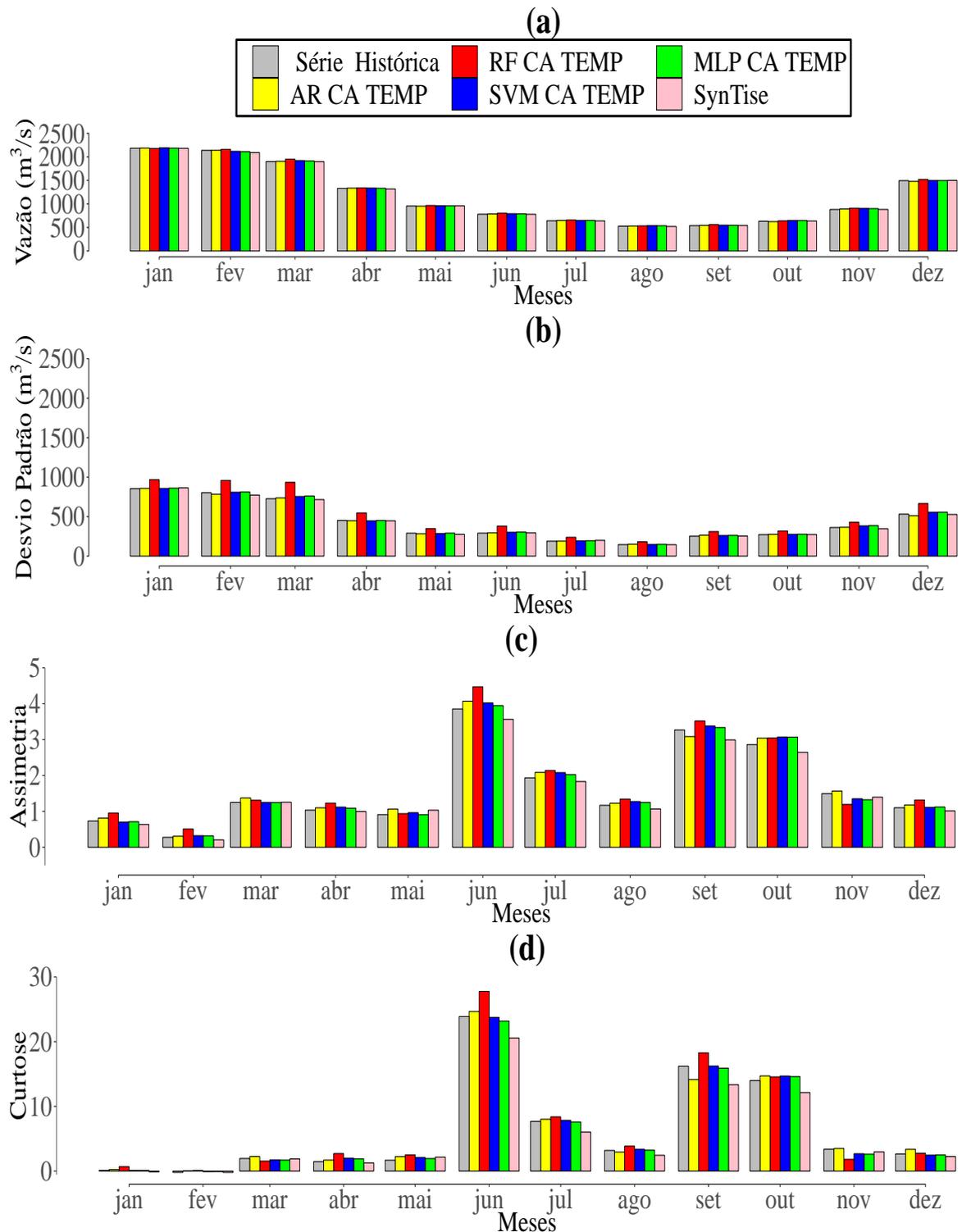
Por fim, comparando as médias dos RMSE da FDP apresentado na Tabela 1, nota-se que o modelo MLP foi o que apresentou menor erro. Sendo assim, dentre todas as propostas apresentadas para geração de série sintéticas, considera-se o modelo que possui um melhor resultado é o MLP CA TEMP.

4.3 USINA DE VOLTA GRANDE

Para usina de Volta Grande, entre as modelagens propostas para geração do CA nas técnicas de inteligência artificial e série temporal, em todas elas o CA TEMP obteve os melhores resultados (ver Apêndices F, G, H e I). Os resultados desses modelos serão comparados entre si e o SynTise (ver Apêndice J) a fim de selecionar qual deles obteve os melhores resultados.

Na comparação das estatísticas descritivas desses modelos, nota-se que todos conseguiram obtiveram valores de médias bem próximos ao da série histórica (ver Figura 24a). Já quanto ao desvios padrão, o RF acabou por superestimar o valor da série histórica, diferentemente dos demais modelos (ver Figura 24b). Quanto a assimetria e curtose, as Figuras 24c e 24d mostram que o SynTise apesar de conseguir replicar bem essas características, não obteve uma eficiência tão grande quanto o AR, SVM e MLP, principalmente nos meses com valores maiores, como junho, setembro e outubro. O RF de um modo geral também conseguiu replicar bem essas características, porém assim como o SynTise, nos meses de maiores valores, o modelo não foi tão bem, mas nesse caso superestimando o resultado dessas estatísticas.

Figura 24 – Comparação das estatísticas descritivas mensais das vazões mensais geradas pelos melhores modelos em cada uma das técnicas analisadas para usina de Volta Grande

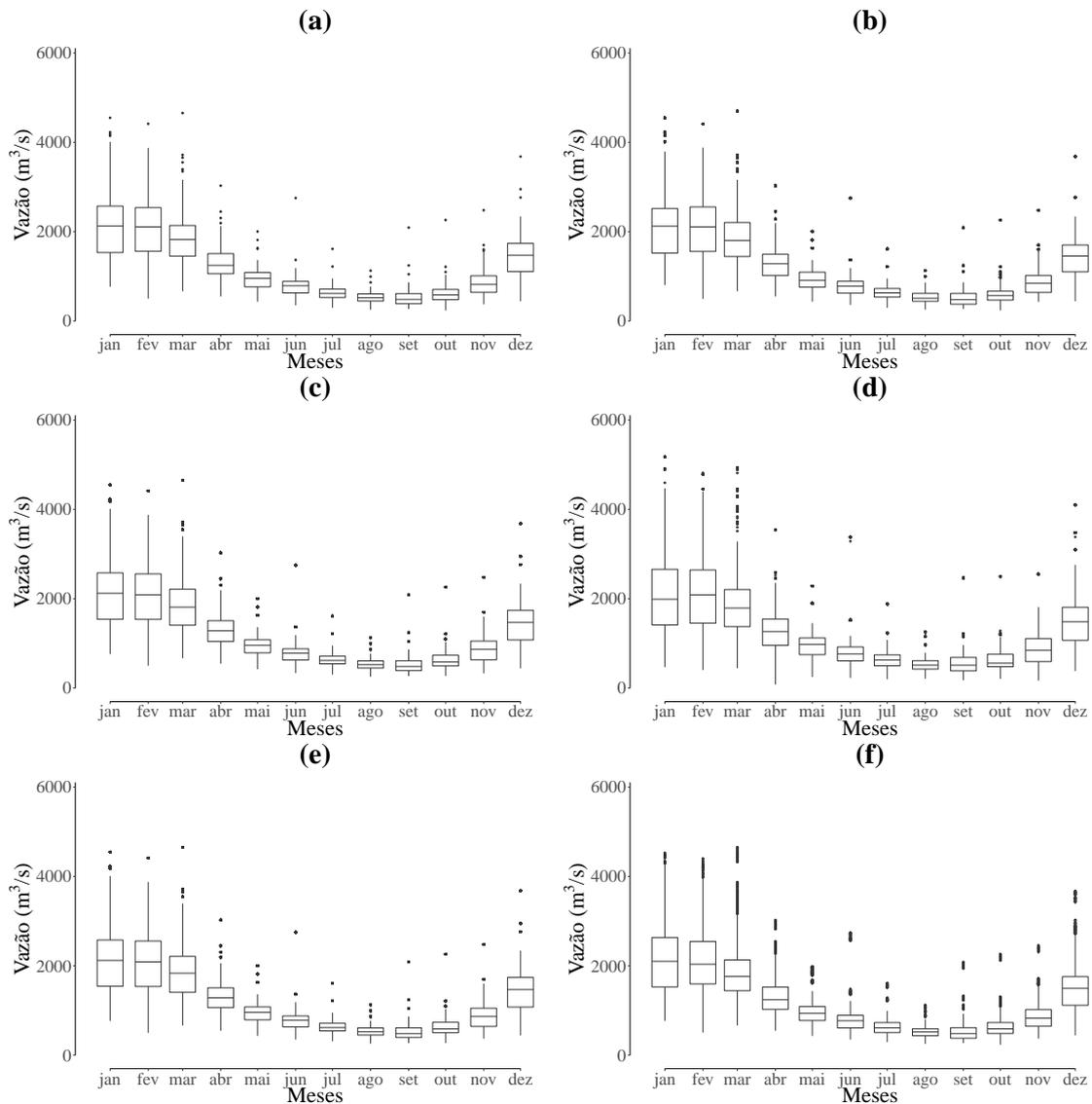


Legenda: a) Média
 b) Desvio Padrão
 c) Assimetria
 d) Curtose

Fonte: Elaborada pelo autor.

Através dos box-plots (ver Figuras 25b, 25c, 25d, 25e, e 25f) e q-qplots (ver Figuras 26a, 26b, 26c, 26d, e 26e), vê-se que o modelo AR apresenta dificuldade quanto a representação dos *outliers* da série histórica, diferentemente do SVM, MLP e SynTise que consegue replicar muito bem esses valores. Quanto ao RF, apresentou dificuldades na representação dos quantis mais elevados e *outliers*, porém representou bem a amplitude e as características descritivas da série, assim como os demais modelos.

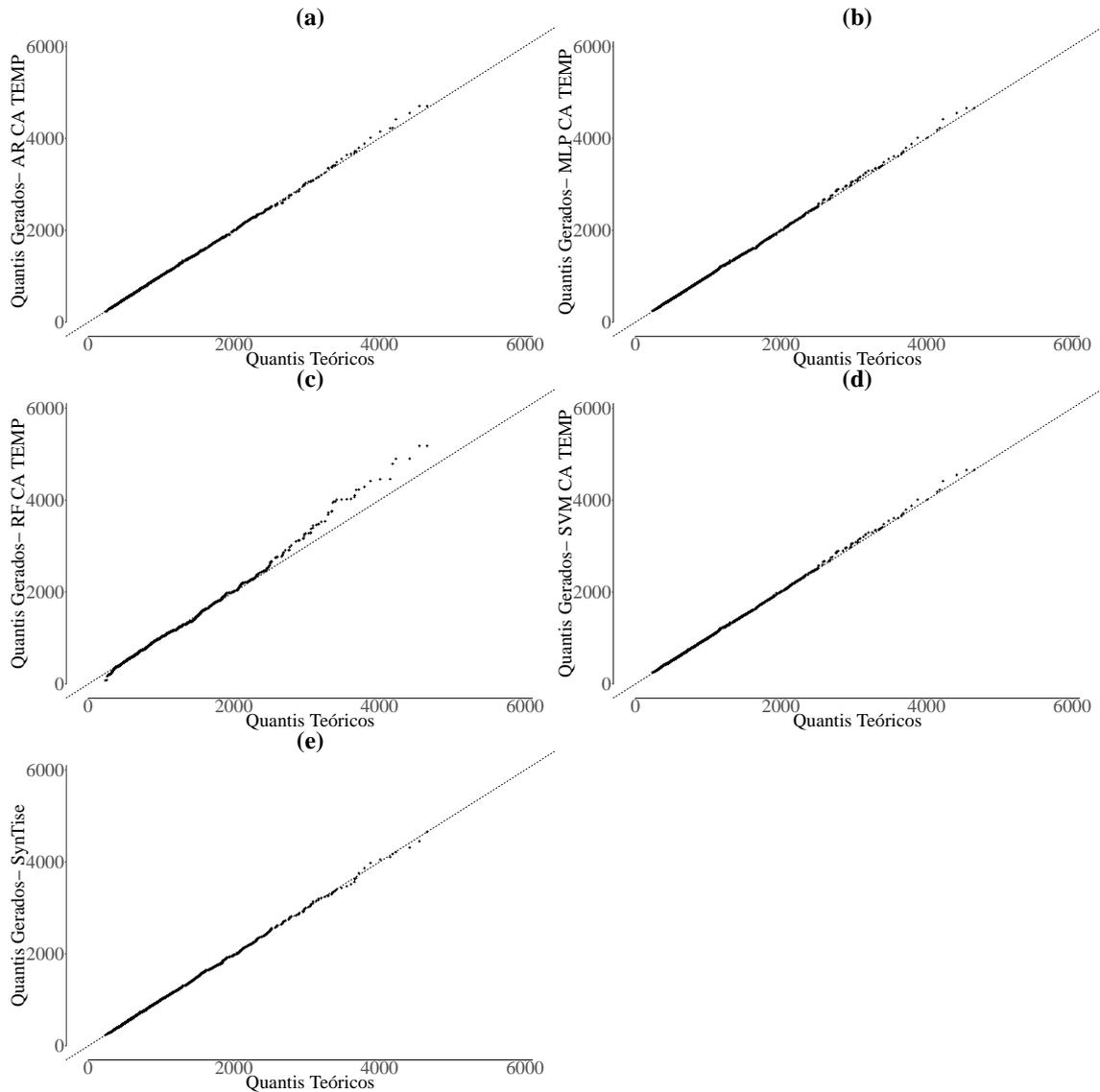
Figura 25 – Comparação entre os box-plots das vazões geradas pelos melhores modelos em cada uma das técnicas analisadas para usina de Volta Grande



Legenda: a) Série Histórica
 b) AR CA TEMP
 c) MLP CA TEMP
 d) RF CA TEMP
 e) SVM CA TEMP
 f) SynTise

Fonte: Elaborada pelo autor.

Figura 26 – Comparação entre os q-qplots das vazões geradas pelos melhores modelos em cada uma das técnicas analisadas para usina de Volta Grande



Legenda: a) AR CA TEMP
 b) MLP CA TEMP
 c) RF CA TEMP
 d) SVM CA TEMP
 e) SynTise

Fonte: Elaborada pelo autor.

Quanto aos erros da FAC e FDP, pela Tabela 2, nota-se que o erro das FACs dos modelos AR, SVM e MLP foram significativamente menores que os demais, quanto a FDP, o SynTise obteve erros próximos aos desses modelos, sendo o menor nos meses de março e fevereiro. O RF foi o melhor no mês de outubro nesse quesito, porém ficou aquém dos demais modelos nos meses de janeiro, fevereiro e dezembro. Quanto a FAC, ele não conseguiu erros tão baixos como AR, SVM e MLP, mas ainda sim seus erros foram menores que o SynTise.

Tabela 2 – RMSE da FAC e FDP mensal dos melhores modelos em cada uma das técnicas analisadas para usina de Volta Grande

RMSE da FAC					
Mês	AR CA TEMP	RF CA TEMP	SVM CA TEMP	MLP CA TEMP	SynTise
janeiro	0,045	0,080	0,063	0,061	0,097
fevereiro	0,070	0,064	0,062	0,063	0,118
março	0,048	0,052	0,026	0,029	0,107
abril	0,052	0,063	0,042	0,048	0,082
maio	0,048	0,060	0,032	0,038	0,086
junho	0,022	0,041	0,028	0,026	0,069
julho	0,029	0,045	0,029	0,027	0,081
agosto	0,039	0,062	0,042	0,035	0,104
setembro	0,032	0,045	0,031	0,038	0,103
outubro	0,025	0,034	0,039	0,036	0,063
novembro	0,027	0,031	0,030	0,034	0,065
dezembro	0,044	0,043	0,031	0,030	0,074
Média dos meses	0,040	0,052	0,038	0,039	0,087
RMSE da FDP					
Mês	AR CA TEMP	RF CA TEMP	SVM CA TEMP	MLP CA TEMP	SynTise
janeiro	0,007	0,026	0,014	0,012	0,012
fevereiro	0,022	0,022	0,011	0,019	0,015
março	0,012	0,025	0,009	0,010	0,009
abril	0,018	0,017	0,013	0,016	0,017
maio	0,017	0,016	0,012	0,012	0,013
junho	0,007	0,015	0,012	0,013	0,010
julho	0,010	0,021	0,010	0,005	0,018
agosto	0,015	0,017	0,012	0,013	0,020
setembro	0,004	0,022	0,013	0,007	0,009
outubro	0,012	0,008	0,022	0,022	0,009
novembro	0,019	0,020	0,019	0,021	0,019
dezembro	0,014	0,021	0,012	0,012	0,011
Média dos meses	0,013	0,019	0,013	0,014	0,014

Fonte: Elaborada pelo autor.

Sendo assim, através do RMSE médio dos meses da Tabela 2, nota-se que os modelos AR e SVM CA TEMP são os que possuem menor erro. Como critério de desempate, o RMSE médio da FAC é analisado e o modelo SVM CA TEMP possui menor erro, sendo assim considerado o melhor modelo para geração de série sintética para usina de Volta Grande.

4.4 ANÁLISE GERAL

Analisando os dois postos, viu-se que os modelos com CA TEMP foram os mais eficientes em todas as técnicas. Nota-se também que em relação as estatísticas descritivas, todos os modelos de um modo geral conseguiram representar bem as estatísticas descritivas, com exceção do *random forest* que superestimou o desvio padrão nos dois casos em todos os meses. Nos dois postos, todos os modelos apresentaram dificuldades em replicar os valores dos quartis maiores, em especial o *random forest*. Já quanto aos RMSE da FAC e FDP, é visto que o SynTise obteve erros significativamente maiores que os demais modelos quanto a FAC nos dois postos, já na FDP, os erros foram próximos dos demais modelos em ambos os casos. O modelo SVM foi o que teve os menores erros para o posto de Volta Grande, porém esse comportamento não acontece na para usina de Água Vermelha, onde seus erros foram maiores que os demais modelos. Comportamento esse que não ocorreu no modelo MLP, que teve bons resultados em ambos os casos analisados, sendo considerado a opção mais viável para o caso de Água Vermelha. Numa comparação, vendo que o modelo MLP mesmo não sendo tão eficiente quanto o SVM no posto de Volta Grande também conseguiu produzir bem próximos, chega-se a conclusão que dentre todas os modelos avaliado, o MLP CA TEMP se mostrou a opção mais viável para geração de séries sintéticas de vazão mensal.

Em relação a outros trabalhos, em Salgado et al. (2012) a mesma análise das estatísticas descritivas, box-plots e q-qplots foram utilizadas para a seleção do melhor modelo para as usinas de Furnas e Peixoto, que também pertencem a bacia do Rio Grande, entretanto não houve uma avaliação da função densidade de probabilidade e autocorrelação das séries sintéticas geradas. Papaefthymiou e Klockl (2008) foram feitas análises dessas características para produção de séries sintéticas da velocidade média dos ventos na cidade de Amsterdã, porém somente gráficas, com o intuito de mostrar que a série gerada mantém essas características dos dados. Ochoa-Rivera, García-Bartual e Andreu (2002) avaliou duas das principais estatísticas descritivas dos dados, média e desvio padrão, pelas séries sintéticas geradas para as vazões mensais dos reservatórios de Buendia e Entrepeñas, na Espanha. Além da vazão dos rios, foram geradas séries sintéticas pelo mesmo modelo de redes neurais para características como a capacidade do reservatório, o tamanho, magnitude, frequência e intensidade da vazão média e anual. Entre os trabalhos que avaliaram as funções de probabilidade e autocorrelação, está o de Denaxas et al. (2015), que calculou o erro dessas funções através do RMSE para comparar o modelo com

múltiplos regimes com o de regime único, para comprovar que o modelo de múltiplos regimes apresenta menores erros e selecionar qual melhor estrutura para geração da série sintética.

5 CONCLUSÃO E PERSPECTIVAS FUTURAS

O trabalho apresentou cinco técnicas para geração de série sintética para as usinas de Volta Grande e Água Vermelha: modelo autorregressivo (AR), máquinas de vetores e suporte (SVM), redes neurais multicamadas (MLP), *random forest* (RF) e SynTise. Para as quatro primeiras técnicas, foram propostos quatro novas abordagens para seleção do componente aleatório (CA) do modelo: por ordem cronológica (TEMP), através dos resíduos estimados (EST,) através de uma distribuição de probabilidade simétrica (DIST.N) e através de uma distribuição de probabilidade assimétrica (DIST.SN). Essas propostas foram comparadas em cada uma das técnicas com a abordagem mais usual, a seleção aleatória dos resíduos (ALEAT).

Os resultados apresentados mostraram que entre as propostas de novos modelos, o uso do CA TEMP obteve melhores resultados do que o CA ALEAT em todas as técnicas testadas, quanto a estimação dos *outliers* presentes nas séries, a replicação das principais estatísticas descritivas e aos erros da FAC e FDP. As demais propostas de geração dos CA não se mostraram muito eficazes ao uso do CA ALEAT, principalmente no que tange a reprodução da assimetria e curtose da série histórica. Esse comportamento foi verificado nos dois postos analisados, Água Vermelha e Volta Grande, mostrando assim que esse efeito ocorre tanto em usinas fio d'água, como nas que possuem reservatório. O modelo SynTise proposto em (DENAXAS et al., 2015) para geração de séries sintéticas de vento, originalmente programado em Matlab foi reprogramado no software R para gerar dados de vazões sintéticas e obteve bons resultados quanto a reprodução das principais características das séries analisadas. Quando comparados com o modelo clássico AR, todos os modelos de inteligência artificial quanto o SynTise se mostraram uma opção viável, principalmente em relação aos *outliers*, cujo modelo AR apresentou uma certa dificuldade de estimá-los. Numa comparação entre os melhores modelos selecionados, concluiu-se que o modelo MLP CA TEMP se mostrou a melhor opção dentre todas as apresentadas para geração de séries sintéticas de vazão mensal.

Como trabalhos futuros uma proposta seria o uso de combinação desses modelos em um *ensemble* com intuito de otimizar os resultados já obtidos. Quanto ao SynTise, outras maneiras de realização do Método de Monte Carlo para a realização do MCMC do processo de geração das série sintéticas podem ser avaliados a fim de obter-se uma melhora nesse modelo, visto que a seleção do passo seguinte foi selecionado de maneira aleatória.

REFERÊNCIAS

- ACURI, R. **Usina de Marmelos**. [s.l.:s.n.], 2015. Disponível em: <<http://www.ricardoarcuri.com.br/jfora/usina/>>. Acesso em: 24 nov. 2017.
- AES BRASIL. **Eficiência em Água Vermelha**. [s.l.:s.n.], 2016. Disponível em: <<http://www.aesbrasilsustentabilidade.com.br/pt/noticias/item/eficiencia-em-agua-vermelha>>. Acesso em: 13 maio 2018.
- AES TIETE. **Situação dos principais reservatórios do Brasil**. [s.l.:s.n.], 2014. Disponível em: <www.aestiete.com.br>. Acesso em: 03 maio 2018.
- AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. **Matriz de energia elétrica**. [s.l.:s.n.], 2016. Disponível em: <<http://www.aneel.gov.br/>>. Acesso em: 12 maio 2018.
- AGUIAR, F. G. **Utilização de redes neurais artificiais para detecção de padrões de vazamento em dutos**. 2010. 95 f. Dissertação (Mestrado em Engenharia Mecânica) — Escola de Engenharia de São Carlos, Universidade de São Paulo, São Paulo, 2010.
- AHMED, J. A.; SARMA, A. K. Artificial neural network model for synthetic streamflow generation. **Water Resources Management**, Reston, v. 21, n. 6, p. 1015–1029, July 2006.
- AKBARI, M.; VAN OVERLOOP, P. J.; AFSHAR, A. Clustered k nearest neighbor algorithm for daily inflow forecasting. **Water Resources Management**, Reston, v. 25, n. 5, p. 1341–1357, Jan. 2011.
- AZZALINI, A. A class of distributions which includes the normal ones. **Scandinavian Journal of Statistics**, [s.l.], v. 12, n. 2, p. 171–178, [1985].
- BLANCHARD, M.; DESROCHERS, G. Generation of autocorrelated wind speeds for wind energy conversion system studies. **Solar Energy**, Amsterdã, v. 33, n. 6, p. 571–579, June 1984.
- BORGOMEIO, E.; FARMER, C. L.; HALL, J. W. Numerical rivers: A synthetic streamflow generator for water resources vulnerability assessments. **Water Resources Research**, Reston, v. 51, n. 7, p. 5382–5405, July 2015.
- BOX, G. E.; JENKINS, G. M. **Time series analysis: forecasting and control**. revised ed. New Jersey: Holden-Day, 1976. 553 p.
- BRASIL. Ministério das Relações Exteriores. **Recursos Hídricos**. Brasília, 2018. Disponível em: <<http://www.itamaraty.gov.br/pt-BR/politica-externa/desenvolvimento-sustentavel-e-meio-ambiente/176-recursos-hidricos>>. Acesso em: 12 maio 2018.
- BRASIL. Ministério de Minas e Energia. **Capacidade instalada de geração elétrica Brasil e mundo (2016)**. Brasília, 2017. Disponível em: <<https://goo.gl/cgL4QS>>. Acesso em: 27 abr.

2017.

BREIMAN, L. Random forests. **Machine learning**, Amsterdã, v. 45, n. 1, p. 5–32, Apr. 2001.

BRISSETTE, F.; KHALILI, M.; LECONTE, R. Efficient stochastic generation of multi-site synthetic precipitation data. **Journal of Hydrology**, [s.l.], v. 345, n. 3-4, p. 121–133, June 2007.

BROKISH, K.; KIRTLEY, J. Pitfalls of modeling wind power using markov chains. In: IEEE/PES POWER SYSTEM CONFERENCE AND EXPOSITION, 9., 2009. **Proceedings...** Seattle: IEEE, 2009. p. 1–6.

BUNN, D.; FARMER, E. D. **Comparative models for electrical load forecasting**. New York: John Wiley and Sons, 1985. 232 p.

CARTA, J. A.; RAMIREZ, P.; VELAZQUEZ, S. A review of wind speed probability distributions used in wind energy analysis: Case studies in the canary islands. **Renewable and sustainable energy reviews**, Amsterdã, v. 13, n. 5, p. 933–955, May 2009.

COMPANHIA ENERGÉTICA DE MINAS GERAIS. **Programa proximidade**. [s.l.:s.n.], 2013. Disponível em: <http://www.cemig.com.br/pt-br/A_Cemig_e_o_Futuro/sustentabilidade/nossos_programas/sociais/Paginas/plano_integracao_comunidades.aspx>. Acesso em: 13 maio 2018.

DENAXAS, E. A. et al. Syntise: A modified multi-regime mcmc approach for generation of wind power synthetic time series. In: ANNUAL IEEE SYSTEMS CONFERENCE (SYSCON) PROCEEDINGS, 9., 2015. **Proceedings...** Vancouver: IEEE, 2015. p. 668–674.

DETZEL, D. H. M.; MINE, M. R. M. Geração estocástica de séries sintéticas de precipitações em escala diária: revisão dos principais modelos e métodos. In: SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, 18., 2009. **Anais...** Campo Grande: ResearchGate, 2009. p. 1–18.

DETZEL, D. H. M. et al. Cenários sintéticos de vazões para grandes sistemas hídricos através de modelos contemporâneos e amostragem. **RBRH: revista brasileira de recursos hídricos**, Rio de Janeiro, v. 19, n. 1, p. 17–28, mar. 2014.

EVSUKOFF, A. G.; DE LIMA, B. S.; EBECKEN, N. F. F. Long-term runoff modeling using rainfall forecasts with application to the iguaçu river basin. **Water resources management**, Reston, v. 25, n. 3, p. 963–985, Jan. 2011.

FERREIRA, D. F. **Estatística básica**. 2. ed. Lavras: UFLA, 2005. 663 p.

FORTUNATO, L. A. et al. **Introdução ao planejamento da expansão e operação de sistemas de produção de energia elétrica**. Niterói: EDUFF, 1990. 232 p.

FRANCISCO, W. D. C. **Energia hidrelétrica**. [s.l.:s.n.], 2016. Disponível em: <<http://brasilecola.uol.com.br/geografia/energia-hidreletrica.htm>>. Acesso em: 08 nov. 2017.

- FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine. **Annals of Statistics**, Beachwood, v. 29, n. 5, p. 1189–1232, Oct. 2001.
- GODFREY, B. et al. **Renewable energy: power for a sustainable future**. Oxford: The Open University Oxford, 2004. 157 p.
- GRUPOS DE ESTUDOS TAPAJÓS. **Diferenças entre usinas com reservatório de regularização e fio d'água**. [s.l.:s.n.], 2015. Disponível em: <<http://www.grupodeestudostapajos.com.br/diferencas-entre-usinas-com-reservatorio-de-regularizacao-e-fio-dagua/>>. Acesso em: 19 maio 2018.
- GUYON, I. et al. Design of a neural network character recognizer for a touch terminal. **Pattern Recognition**, Londres, v. 24, n. 2, p. 105–119, July 1991.
- HANSEN, L. K.; SALAMON, P. Neural network ensembles. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Toronto, v. 12, p. 993–1001, Oct. 1990.
- HASSAN, S.; KHOSRAVI, A.; JAAFAR, J. Examining performance of aggregation algorithms for neural network-based electricity demand forecasting. **International Journal of Electrical Power & Energy Systems**, Manchester, v. 64, p. 1098–1105, Sept. 2015.
- HAYKIN, S. **Redes neurais: princípios e prática**. 2. ed. [s.l.]: Bookman Editora, 2007.
- JIA, Y.; CULVER, T. B. Bootstrapped artificial neural networks for synthetic flow generation with a small data sample. **Journal of Hydrology**, [s.l.], v. 331, n. 3, p. 580–590, Dec. 2006.
- LIMA, A. G. G. **Geração hidrelétrica**. [s.l.:s.n.], 2014. Disponível em: <<http://www.antonioguilherme.web.br.com/Arquivos/hidro.php>>. Acesso em: 19 maio 2018.
- MACEIRA, M. E. P.; PENNA, D. D. J.; DAMÁZIO, J. M. Geração de cenários sintéticos de energia e vazão para o planejamento da operação energética. **Cadernos do IME-Série Estatística**, Rio de Janeiro, v. 21, n. 2, p. 11, dez. 2006.
- MINSKY, M.; PAPERT, S. **Perceptrons: an introduction to computational geometry**. expanded ed. Cambridge: [s.n.], 1969. 258 p.
- MONTANARI, A.; ROSSO, R.; TAQQU, M. S. Fractionally differenced arima models applied to hydrologic time series: Identification, estimation, and simulation. **Water resources research**, Reston, v. 33, n. 5, p. 1035–1044, May 1997.
- MONTGOMERY, D. C.; JOHNSON, L. A.; GARDINER, J. S. **Forecasting and time series analysis**. Michigan: McGraw-Hill Companies, 1990. 381 p.
- MORETTIN, P. A.; TOLOI, C. **Análise de séries temporais**. 2. ed. São Paulo: Edgard Blucher, 2006. 538 p.
- MULLER, F.; KAVISKI, E. Geração de vazões diárias ii: Preservação da forma dos hidrogramas em grandes bacias. In: SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS,

17., 2007. **Anais...** São Paulo: ResearchGate, 2007. p. 1–19.

MUNDO DA ELÉTRICA. **Usinas termelétricas. O que são e como funcionam?** [s.l.:s.n.], 2016. Disponível em: <<https://www.mundodaeletrica.com.br/usinas-termeletricas-o-que-sao-e-como-funcionam/>>. Acesso em: 05 dez. 2018.

NFAOUI, H.; ESSIARAB, H.; SAYIGH, A. A stochastic markov chain model for simulating wind speed time series at Tangiers, Morocco. **Renewable Energy**, Lemosos, v. 29, n. 8, p. 1407–1418, July 2004.

NIJSSEN, B. et al. Streamflow simulation for continental-scale river basins. **Water Resources Research**, Reston, v. 33, n. 4, p. 711–724, Apr. 1997.

NOVAES, L. F. **Modelo para a quantificação da disponibilidade hídrica na bacia do Paracatu**. 2005. 115 f. Dissertação (Mestrado em Engenharia Agrícola) — Departamento de Engenharia Agrícola, Universidade Federal de Viçosa, Viçosa, 2005.

OCHOA-RIVERA, J.; GARCÍA-BARTUAL, R.; ANDREU, J. Multivariate synthetic streamflow generation using a hybrid model based on artificial neural networks. **Hydrology and Earth System Sciences Discussions**, Paris, v. 6, n. 4, p. 641–654, Jan. 2002.

OPERADOR NACIONAL DO SISTEMA ELÉTRICO. **Diagrama esquemático das usinas hidroelétricas do SIN**. [s.l.:s.n.], 2017. Disponível em: <<https://goo.gl/V4CG9e>>. Acesso em: 27 abr. 2018.

_____. **Séries Históricas de Vazões, 2015**. [s.l.:s.n.], 2015. Disponível em: <<http://www.ons.org.br/operacao/vazoesnaturais.aspx>>. Acesso em: 03 maio 2018.

OTHMAN, F.; NASERI, M. Reservoir inflow forecasting using artificial neural network. **International Journal of Physical Sciences**, Sapele, v. 6, n. 3, p. 434–440, Feb. 2011.

PAPAEFTHYMIU, G.; KLOCKL, B. Mcmc for wind power simulation. **IEEE Transactions on Energy Conversion**, Columbia, v. 23, n. 1, p. 234–240, Feb. 2008.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2016. Acesso em: 01 jul. 2016. Disponível em: <<https://www.R-project.org/>>.

REIS, J. F.; SOUZA, W. L. C. D.; JORGE FILHO, S. L. D. O. **Medição da vazão da usina hidrelétrica de Roncador**. 2011. 93 f. Monografia (Bacharelado em Engenharia Industrial Elétrica) — Departamento Acadêmico de Eletrotécnica, Universidade Tecnológica Federal do Paraná, Curitiba, 2011.

RIPLEY, B. D. **Modern applied statistics with S**. New York: Springer, 2002. 495 p.

RISSANEN, J. A universal data compression system. **IEEE Transactions on Information Theory**, Maryland City, v. 29, n. 5, p. 656–664, Sept. 1983.

SALGADO, R. M. et al. Modelos de inteligência computacional para geração de séries

sintéticas de vazões médias mensais. **Learning and Nonlinear Models**, [s.l.], v. 10, n. 3, p. 166–174, jan. 2012.

SEMENOV, M. A. Simulation of extreme weather events by a stochastic weather generator. **Climate Research**, Oldendorf, v. 35, n. 3, p. 203–212, Feb. 2008.

SILVA, A. T.; PORTELA, M. M. Geração de séries sintéticas de escoamentos anuais e mensais. aplicação extensiva a portugal continental. **Recursos Hídricos**, Lisboa, v. 31, n. 2, nov. 2010.

SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on support vector regression. **Statistics and computing**, Coventry, v. 14, n. 3, p. 199–222, Aug. 2004.

SOUSA, B. F.; GUEDES, L. S.; SALDANHA, R. R. Tempvaz: sistema para estudo de séries temporais de vazões de rios para o planejamento da operação de sistemas hidrotérmicos. **Proceeding Series of the Brazilian Society of Computational and Applied Mathematics**, São Carlos, v. 3, n. 1, p. 631–632, [2015].

SRIKANTHAN, R. et al. Comparison of two approaches for generation of daily rainfall data. **Stochastic Environmental Research and Risk Assessment**, Berlim, v. 19, n. 3, p. 215–226, Feb. 2005.

STANO JÚNIOR, Â.; BITENCOURT, V. A.; TIAGO FILHO, G. L. **Energias Renováveis**. Itajubá: Centro Nacional de Referência em Pequenas Centrais Hidroelétricas- CERPCH, 2007. 44 p.

SUOMALAINEN, K. et al. Synthetic wind speed scenarios including diurnal effects: Implications for wind power dimensioning. **Energy**, Aalborg, v. 37, n. 1, p. 41–50, Jan. 2012.

TUCCI, C. E.; PORTO, R.; BARROS, M. T. **Drenagem urbana: coleção ABRH de recursos hídricos**. Porto Alegre: UFRGS/ABRH, 1995. 5 v.

VAPNIK, V. **The nature of statistical learning**. New York: Springer, 1995. 188 p.

VOOO-INSIGHTS. **Um tutorial completo sobre modelagem baseada em árvores de decisão (códigos R e Python)**. [s.l.:s.n.], 2016. Disponível em: <<https://www.vooo.pro/insights/um-tutorial-completo-sobre-a-modelagem-baseada-em-tree-arvore-do-zero-em-r-python/>>. Acesso em: 01 set. 2018.

WILKS, D. Simultaneous stochastic simulation of daily precipitation, temperature and solar radiation at multiple sites in complex terrain. **Agricultural and Forest Meteorology**, New Haven, v. 96, n. 1, p. 85–101, Aug. 1999.

YOKOYAMA, J.; CHIANG, H.-D. Short term load forecasting improved by ensemble and its variations. In: IEEE POWER AND ENERGY SOCIETY GENERAL MEETING, 1., 2012. **Proceedings...** San Diego: IEEE, 2012. p. 1–6.

ZHANG, G. P.; QI, M. Neural network forecasting for seasonal and trend time series. **European Journal of Operational Research**, Poznan, v. 160, n. 2, p. 501–514, Jan. 2005.

ZHANG, R. et al. Short-term load forecasting of Australian National Electricity Market by an ensemble model of extreme learning machine. **Generation, Transmission & Distribution**, [s.l.], v. 7, n. 4, p. 391–397, Apr. 2013.

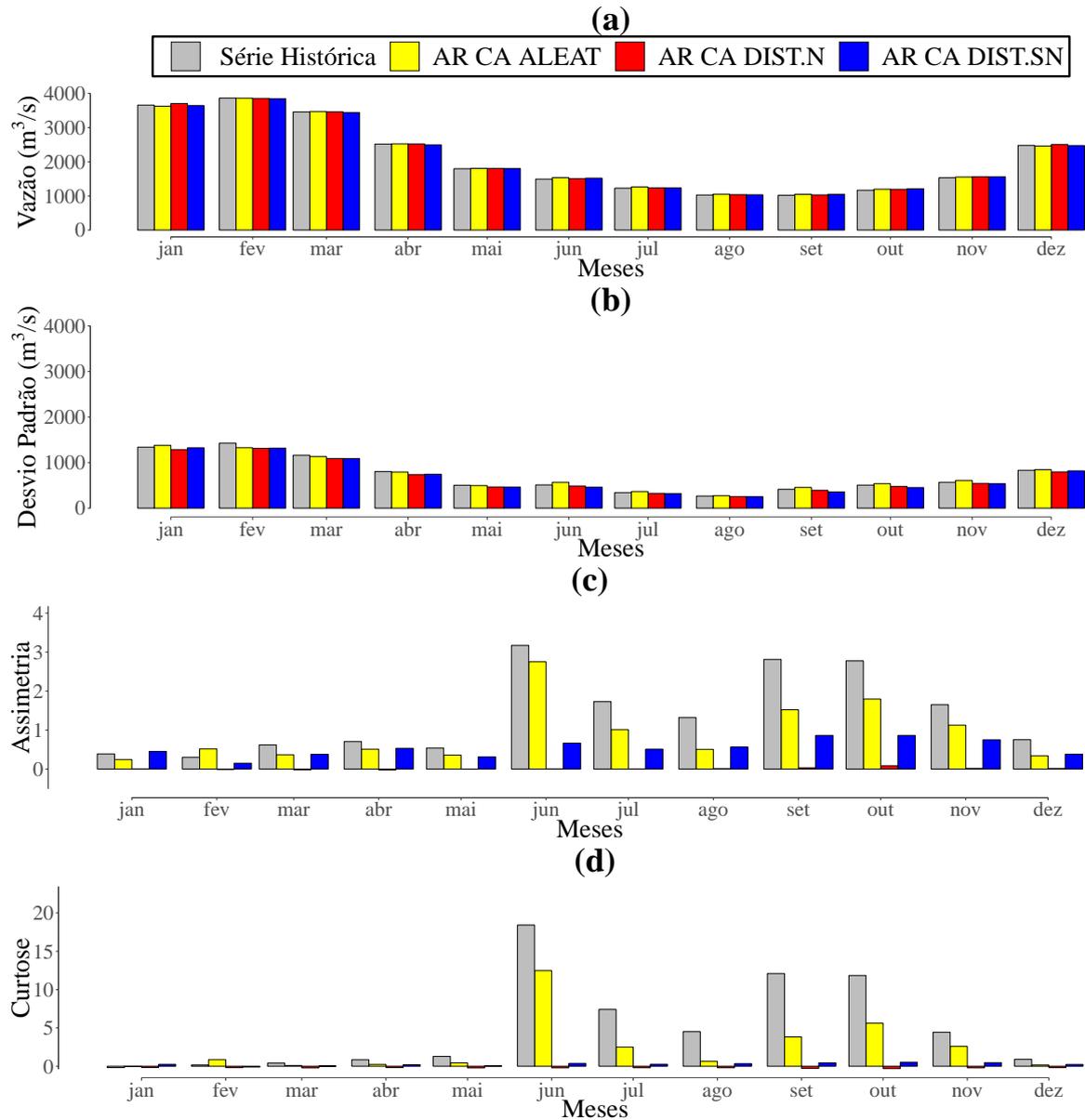
ZHANG, X. et al. Short-term load forecasting based on big data technologies. **Journal of Power and Energy Systems**, [s.l.], v. 1, n. 3, p. 59 – 67, Sept. 2015.

APÊNDICE A - MODELOS AR PARA USINA DE ÁGUA VERMELHA

Para o modelo AR, o parâmetro p que foi considerado para a geração da série sintética para usina de Água Vermelha foi o AR(2), ou seja, a predição do dados foi feita utilizando os dois últimos anos da série histórica.

Inicialmente, foi executada a comparação entre o modelo clássico de CA ALEAT com os de CA DIST.N e DIST.SN. Pelas Figuras 27a e 27b, nota-se que os três modelos conseguiram de certa forma replicar bem tanto a média, quanto o desvio padrão da série histórica do dados em todos os meses. Porém, analisando as Figuras 27c e 27d, vê-se que os modelos com CA DIST.N e DIST.SN não conseguiram replicar bem essas características descritivas dos dados, principalmente nos meses que possuíam valores maiores para assimetria e curtose, como junho, setembro e outubro. O modelo de CA ALEAT apresentou resultados razoáveis para esses meses, mas melhores que esses dois modelos.

Figura 27 – Comparação entre as estatísticas descritivas das séries de vazões mensais geradas pelos modelos AR com CA ALEAT, DIST.N e DIST.SN para usina de Água Vermelha



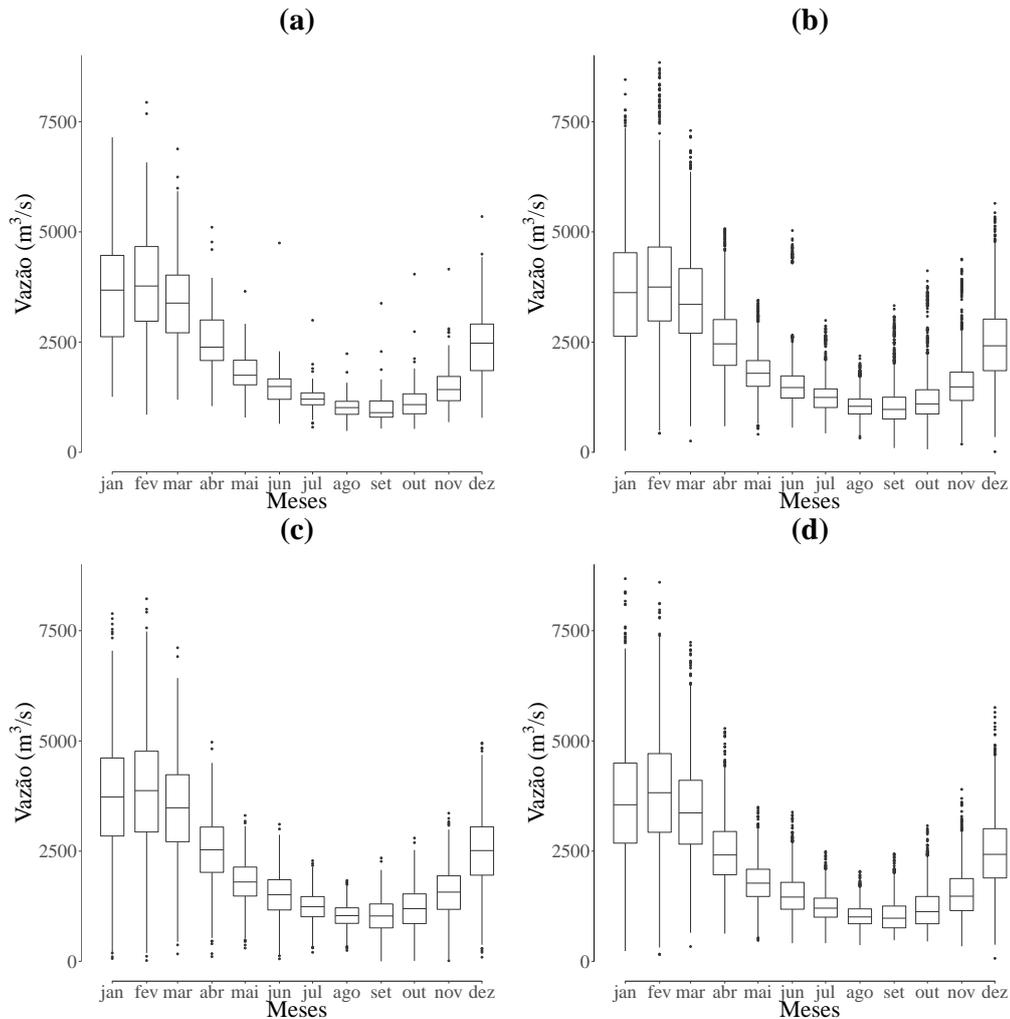
Legenda: a) Média
 b) Desvio Padrão
 c) Assimetria
 d) Curtose

Fonte: Elaborada pelo autor.

Pelos box-plots das vazões geradas nos modelos mostrados nas Figuras 28b, 28c e 28d, é exposta a dificuldade desses modelo quanto aos *outliers*. Mesmo o modelo com CA ALEAT, que conseguiu representar melhor a série histórica, nota-se que *outliers* inferiores foram gerados para os meses de fevereiro, março, maio, agosto, novembro e dezembro, enquanto na série histórica da Figura 28a, apenas julho isso está presente. Os modelos com CA DIST.N até gerou

outliers inferiores para julho, porém se equivocou gerando valores desse tipo em vários outros meses, além de possuir amplitude bastante distinta da série histórica, principalmente nos meses de janeiro, fevereiro e novembro, comportamento semelhante ao apresentado pelo modelo com CA DIST.SN.

Figura 28 – Box-plots das vazões geradas pelos modelos AR para usina de Água Vermelha

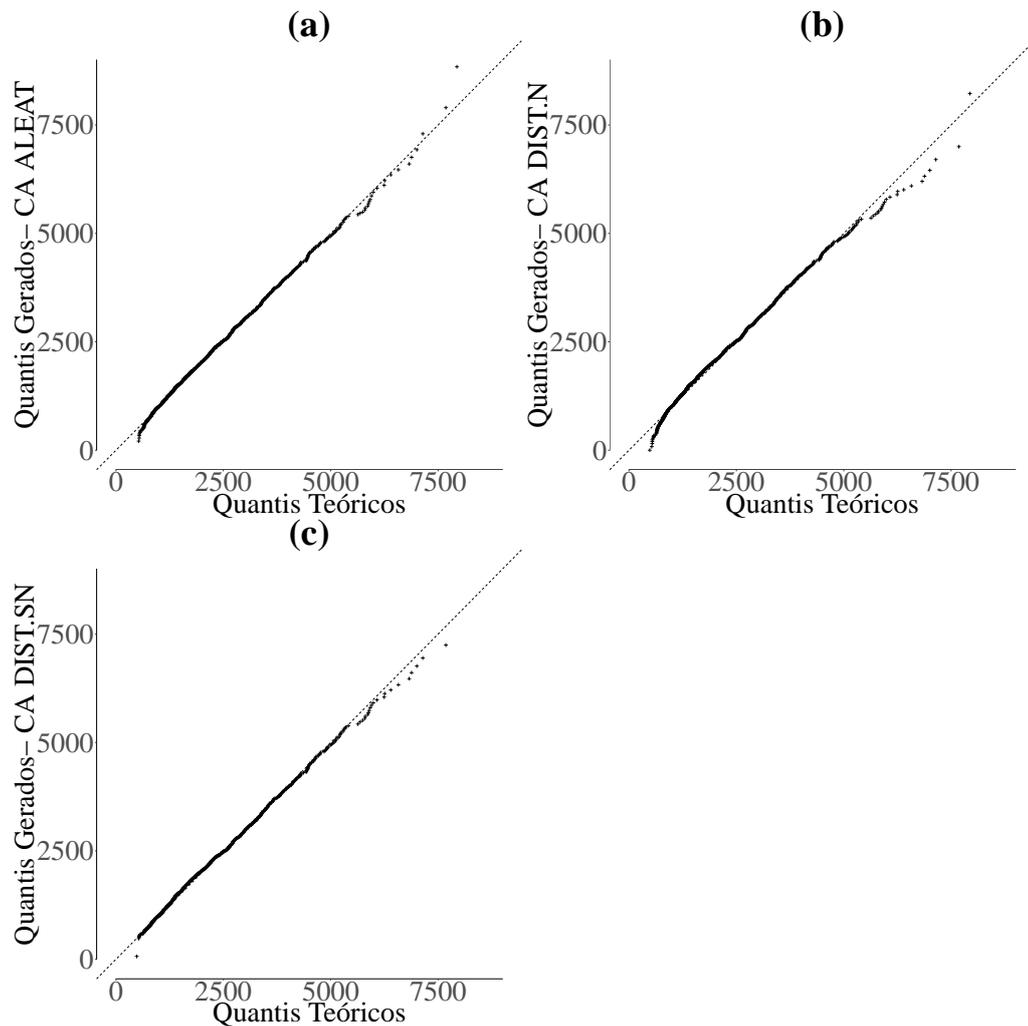


Legenda: a) Série Histórica
 b) CA ALEAT
 c) CA DIST.N
 d) CA DIST.SN

Fonte: Elaborada pelo autor.

Analisando os q-qplots das Figuras 29a, 29b e 29c fica mais evidente a dificuldade que esses modelos apresentam para replicar os quantis mais elevados e os menores da série histórica, já que os quantis gerados não apresentaram um comportamento semelhante aos quantis presentes nos dados. Mesmo assim, o modelo com CA ALEAT apresentou uma capacidade bem superior de representar quantis da série histórica quando comparado aos outros dois modelos.

Figura 29 – Q-qplots das vazões geradas pelos modelos AR para usina de Água Vermelha



Legenda: a) CA ALEAT
 b) CA DIST.N
 c) CA DIST.SN

Fonte: Elaborada pelo autor.

Comparando os erros da FAC e FDP conforme a Tabela 3, pode-se concluir então que os modelos AR CA DIST.N e DIST.SN não foram uma opção mais viável que o modelo AR CA ALEAT, visto que como dito anteriormente, além de não conseguir replicar bem os *outliers*, a curtose e assimetria presente, também apresentou um erro maior tanto na FAC quanto FDP em todos os meses.

Tabela 3 – RMSE da FDP e FAC mensal dos modelos AR de CA ALEAT, DIST.N e DIST.SN

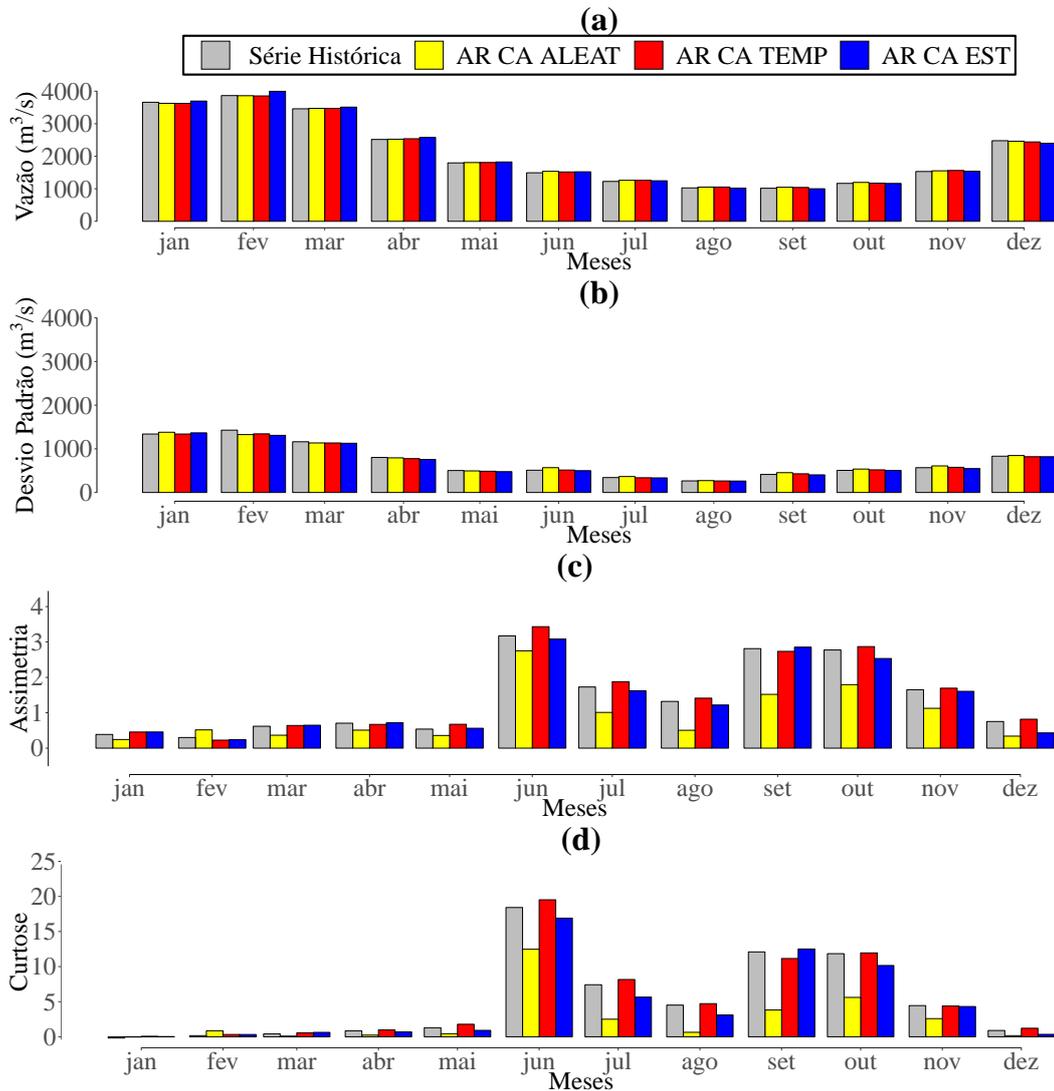
RMSE da FAC			
Mês	AR CA ALEAT	AR CA DIST.N	AR CA DIST.SN
janeiro	0,056	0,072	0,076
fevereiro	0,055	0,09	0,092
março	0,06	0,096	0,104
abril	0,066	0,081	0,093
maio	0,049	0,071	0,082
junho	0,048	0,054	0,071
julho	0,039	0,074	0,082
agosto	0,056	0,106	0,115
setembro	0,045	0,113	0,116
outubro	0,072	0,091	0,099
novembro	0,054	0,073	0,077
dezembro	0,049	0,078	0,089
Média dos meses	0,055	0,083	0,091
RMSE da FDP			
Mês	AR CA ALEAT	AR CA DIST.N	AR CA DIST.SN
janeiro	0,021	0,032	0,037
fevereiro	0,014	0,022	0,029
março	0,036	0,038	0,029
abril	0,03	0,044	0,025
maio	0,025	0,035	0,030
junho	0,015	0,057	0,039
julho	0,030	0,054	0,049
agosto	0,025	0,031	0,036
setembro	0,057	0,078	0,039
outubro	0,044	0,072	0,055
novembro	0,039	0,057	0,028
dezembro	0,032	0,029	0,032
Média dos meses	0,031	0,046	0,036

Fonte: Elaborada pelo autor.

Outras duas propostas apresentadas nesse trabalho a serem analisadas são as CA TEMP e EST.

A análise inicial através das estatísticas descritivas às Figuras 30a e 30b mostram que os três modelos conseguiram reproduzir bem tanto a média e o desvio padrão da série histórica do dados em todos os meses. Já na assimetria e curtose, as duas novas propostas de CA acabaram reproduzindo melhor essas estatísticas do que o modelo de CA ALEAT (ver Figuras 30c e 30d), em especial nos meses cujos valores são maiores, como julho, setembro e outubro.

Figura 30 – Comparação entre as estatísticas descritivas das séries de vazões mensais geradas pelos modelos AR de CA ALEAT, TEMP e EST para usina de Água Vermelha



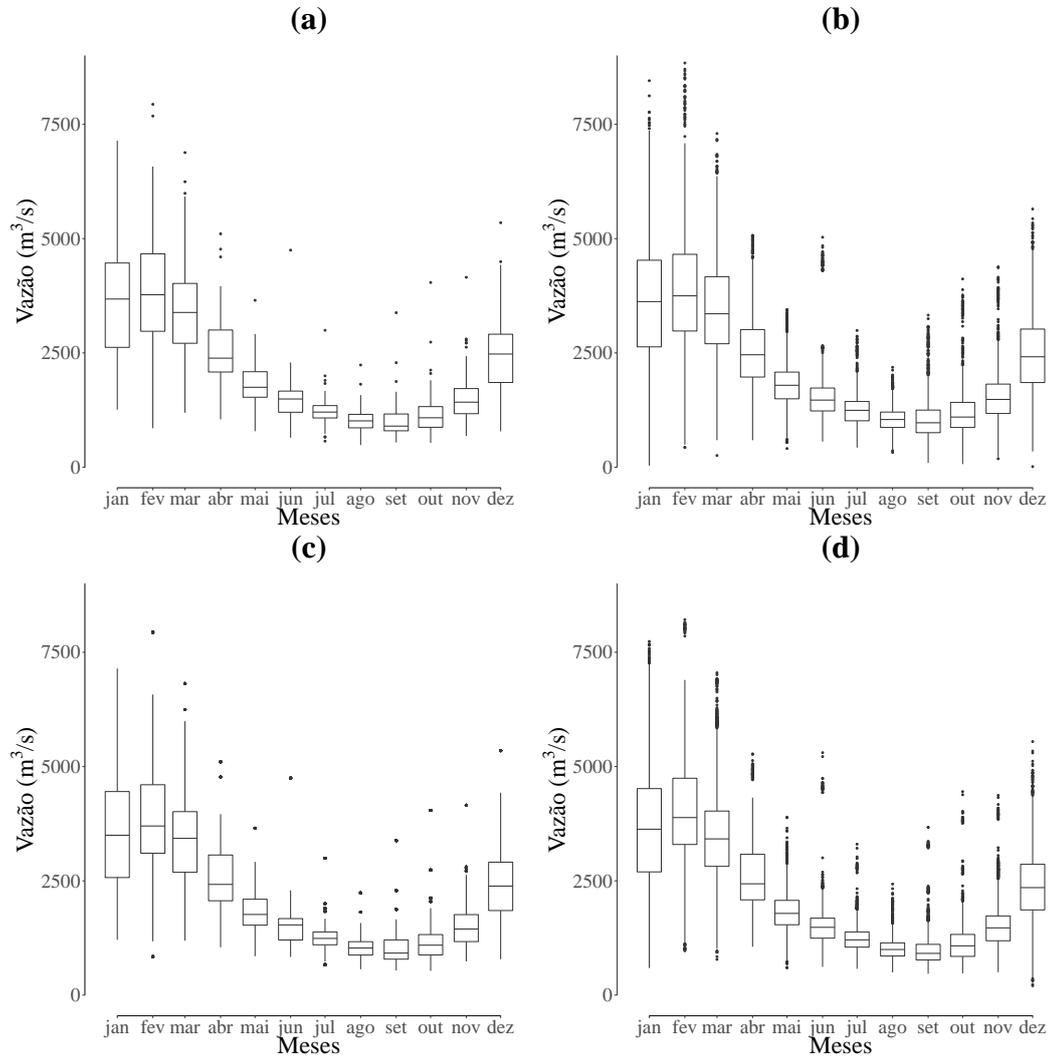
Legenda: a) Média
b) Desvio Padrão
c) Assimetria
d) Curtose

Fonte: Elaborada pelo autor.

Através dos box-plots das vazões geradas pelos modelos exibidos nas Figuras 31c e 31d, pode-se notar que o modelo com CA TEMP proporcionalmente gera poucos *outliers*, porém os valores gerados por ele são bem próximos presentes na série histórica, inclusive com um *outlier* inferior no mês de julho, assim como na série histórica. Já o modelo com CA EST possui *outliers* superiores com valores próximos aos presentes na série histórica, porém assim como o modelo de CA ALEAT (ver Figura 31b), ele gerou *outliers* inferiores em meses onde isso não ocorre na série histórica como fevereiro, abril, junho e dezembro. Já o modelo com CA TEMP

cometeu esse equívoco apenas no mês de fevereiro.

Figura 31 – Box-plots das vazões geradas pelos modelos AR para usina de Água Vermelha

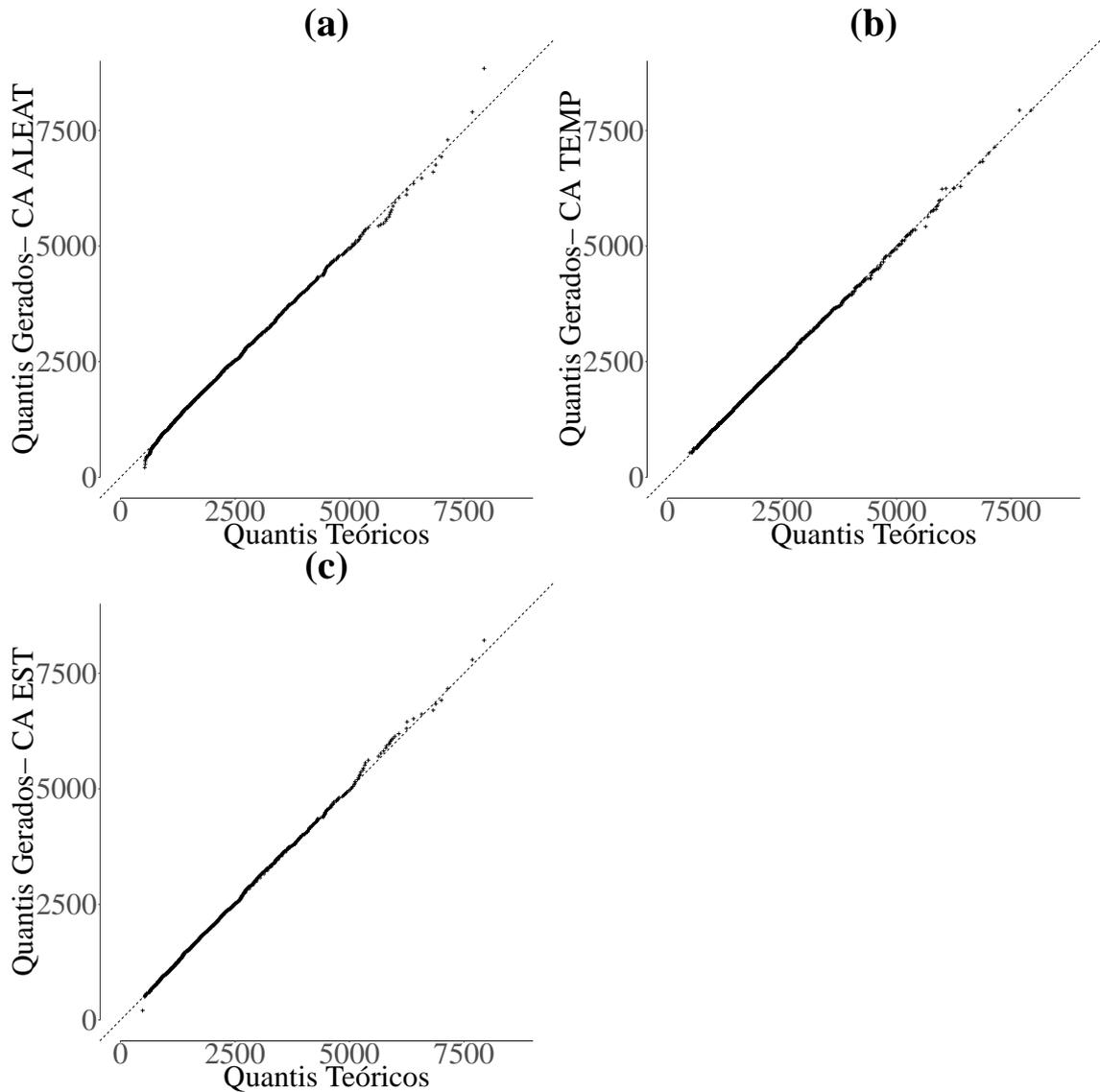


Legenda: a) Série Histórica
 b) CA ALEAT
 c) CA TEMP
 d) CA EST

Fonte: Elaborada pelo autor.

Os q-qplots apresentados nas Figuras 32b e 32c deixam mais claro que ambas as propostas de modelagem para os CA se ajustaram bem aos dados, mostrando-se superior ao modelo de CA ALEAT, principalmente quando se observa o q-qplot da Figura 32a.

Figura 32 – Q-qplots das vazões geradas pelos modelos AR para usina de Água Vermelha



Legenda: a) CA ALEAT
b) CA TEMP
c) CA EST

Fonte: Elaborada pelo autor.

Por fim, explorando os RMSE da FAC e FDP desses três modelos (ver Tabela 4), constata-se que a FAC do modelo de CA EST apresentou um erro maior da FAC em todos os meses, enquanto o modelo de CA TEMP obteve um erro menor nos meses de abril, maio, junho, agosto, setembro, outubro e novembro, sendo superado pelo modelo de CA ALEAT nos demais meses. Já quanto a FDP, o modelo com CA TEMP obteve os menores erros em praticamente todos os meses, sendo que onde isso não ocorreu a diferença foi muito pequena. Destaca-se o fato de nos meses de outubro e setembro, o erro da FDP do modelo de CA TEMP foi tão baixo que atingiu a terceira casa decimal.

Tabela 4 – RMSE da FAC e FDP mensal dos modelos AR de CA ALEAT, TEMP e EST

RMSE da FAC			
Mês	AR CA ALEAT	AR CA TEMP	AR CA EST
janeiro	0,056	0,065	0,073
fevereiro	0,055	0,082	0,087
março	0,06	0,076	0,104
abril	0,066	0,061	0,093
maio	0,049	0,047	0,086
junho	0,048	0,03	0,063
julho	0,039	0,052	0,083
agosto	0,056	0,055	0,114
setembro	0,045	0,036	0,118
outubro	0,072	0,016	0,084
novembro	0,054	0,023	0,069
dezembro	0,049	0,051	0,093
Média dos meses	0,054	0,049	0,089
Mês	AR CA ALEAT	AR CA TEMP	AR CA EST
RMSE da FDP			
janeiro	0,021	0,021	0,016
fevereiro	0,014	0,016	0,032
março	0,036	0,013	0,012
abril	0,03	0,012	0,033
maio	0,025	0,021	0,027
junho	0,015	0,017	0,011
julho	0,03	0,02	0,022
agosto	0,025	0,012	0,026
setembro	0,057	0,006	0,01
outubro	0,044	0,009	0,018
novembro	0,039	0,015	0,021
dezembro	0,032	0,013	0,036
Média dos meses	0,030	0,015	0,022

Fonte: Elaborada pelo autor.

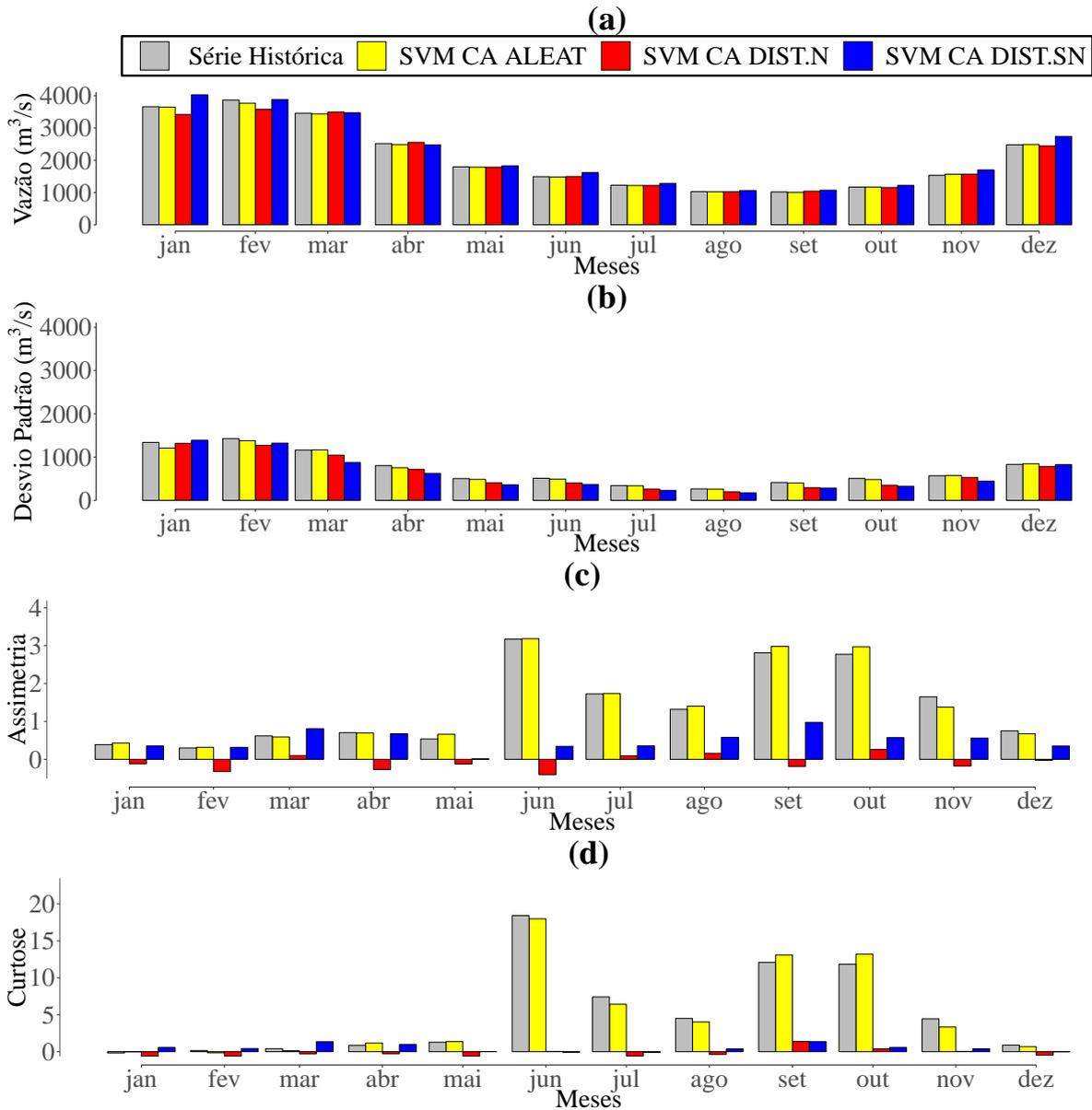
Dada todas essas análises feitas sobre as propostas de modelos apresentadas, conclui-se que para o modelo AR, o uso de CA TEMP aparenta ser a alternativa mais adequada dentre todos os modelos apresentados nesta seção.

APÊNDICE B - MODELOS SVM PARA USINA DE ÁGUA VERMELHA

Os modelos SVM gerados utilizaram seis passos anteriores, ou seja, a predição dos dados foi feita utilizando os últimos seis anos da série histórica.

A primeira das comparações executadas foi entre o modelo clássico de CA ALEAT com os de CA DIST.N e DIS.SN. Pelas Figuras 33a e 33b, nota-se que os três modelos conseguiram de certa forma replicar de forma razoável tanto a média, quanto o desvio padrão da série histórica dos dados em todos os meses, com exceção de setembro, outubro e novembro, onde os modelos que utilizam CA DIST.N e DIST.SN acabaram obtendo resultados consideravelmente menores que o modelo com CA ALEAT. Porém, analisando as Figuras 33c e 33d, vê-se que os modelos com CA DIST.N e DIST.SN não conseguiram replicar bem tanto a curtose quanto a assimetria em praticamente todos os meses, sendo que em alguns casos, até assimetrias negativas em meses com alta assimetria positiva acabaram por serem gerados pelo modelo com CA DIST.N.

Figura 33 – Comparação entre as estatísticas descritivas das séries de vazões mensais geradas pelos modelos SVM de CA ALEAT, DIST.N e DIST.SN para usina de Água Vermelha



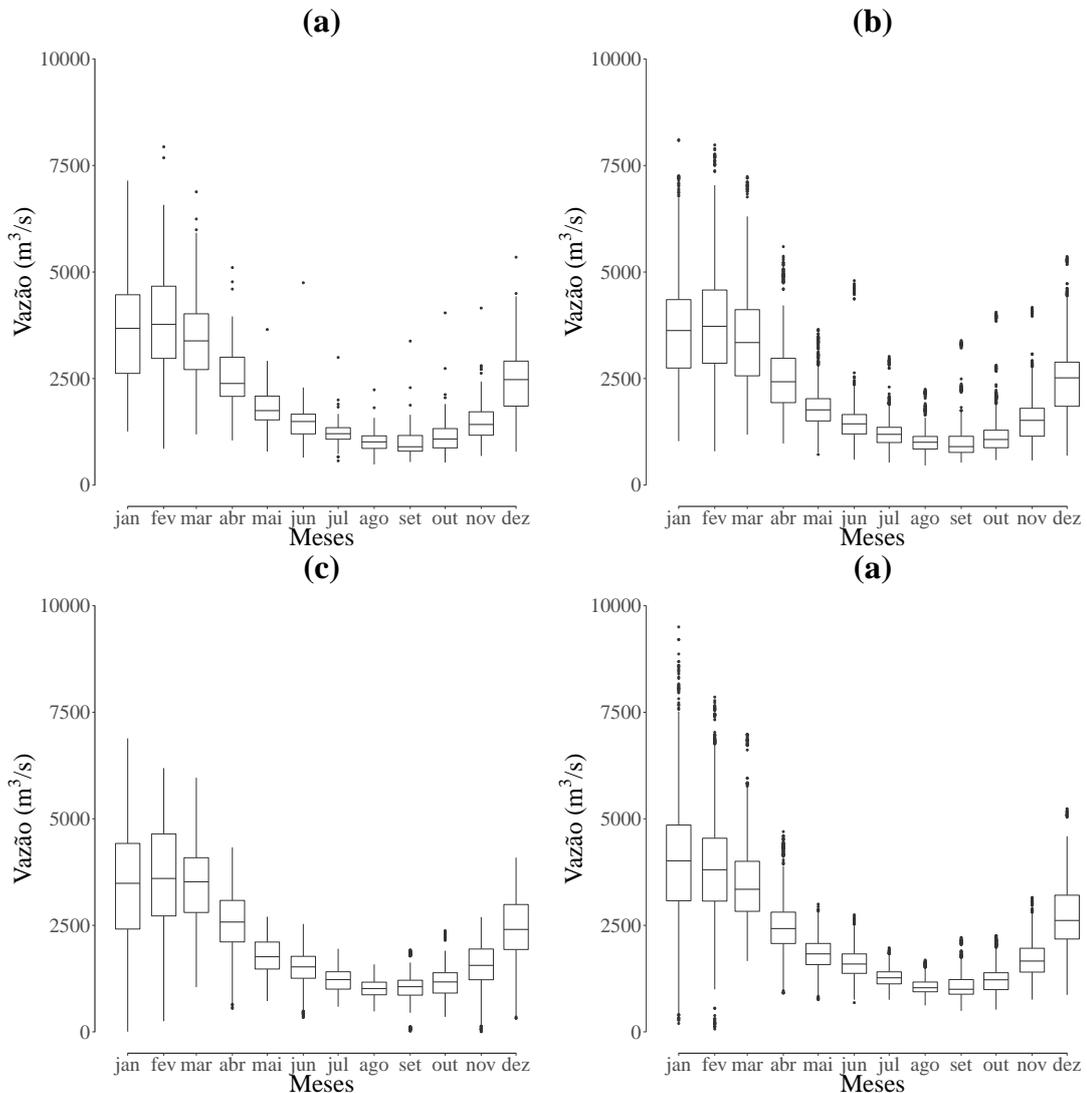
Legenda: a) Média
 b) Desvio Padrão
 c) Assimetria
 d) Curtose

Fonte: Elaborada pelo autor.

Nos box-plots apresentados nas Figuras 34b, 34c e 34d, é exposta uma dificuldade quanto aos *outliers*. Mesmo o modelo com CA ALEAT, que conseguiu representar melhor a série histórica, nota-se que foram gerados *outliers* inferiores para o mês de maio, enquanto na série histórica da Figura 34a, apenas em julho isso está presente. Os modelos com CA DIST.N e DIST.SN além de não geraram nenhum *outlier* inferior para o mês de julho, pelo menos um

deles acabou por gerar valores desse tipo para algum dos demais meses. Além do mais, o modelo de CA DIST.N não produziu nenhum *outlier* superior, sendo que na série histórica, apenas em janeiro não há presença de um desses.

Figura 34 – Box-plots das vazões geradas pelos modelos SVM para usina de Água Vermelha

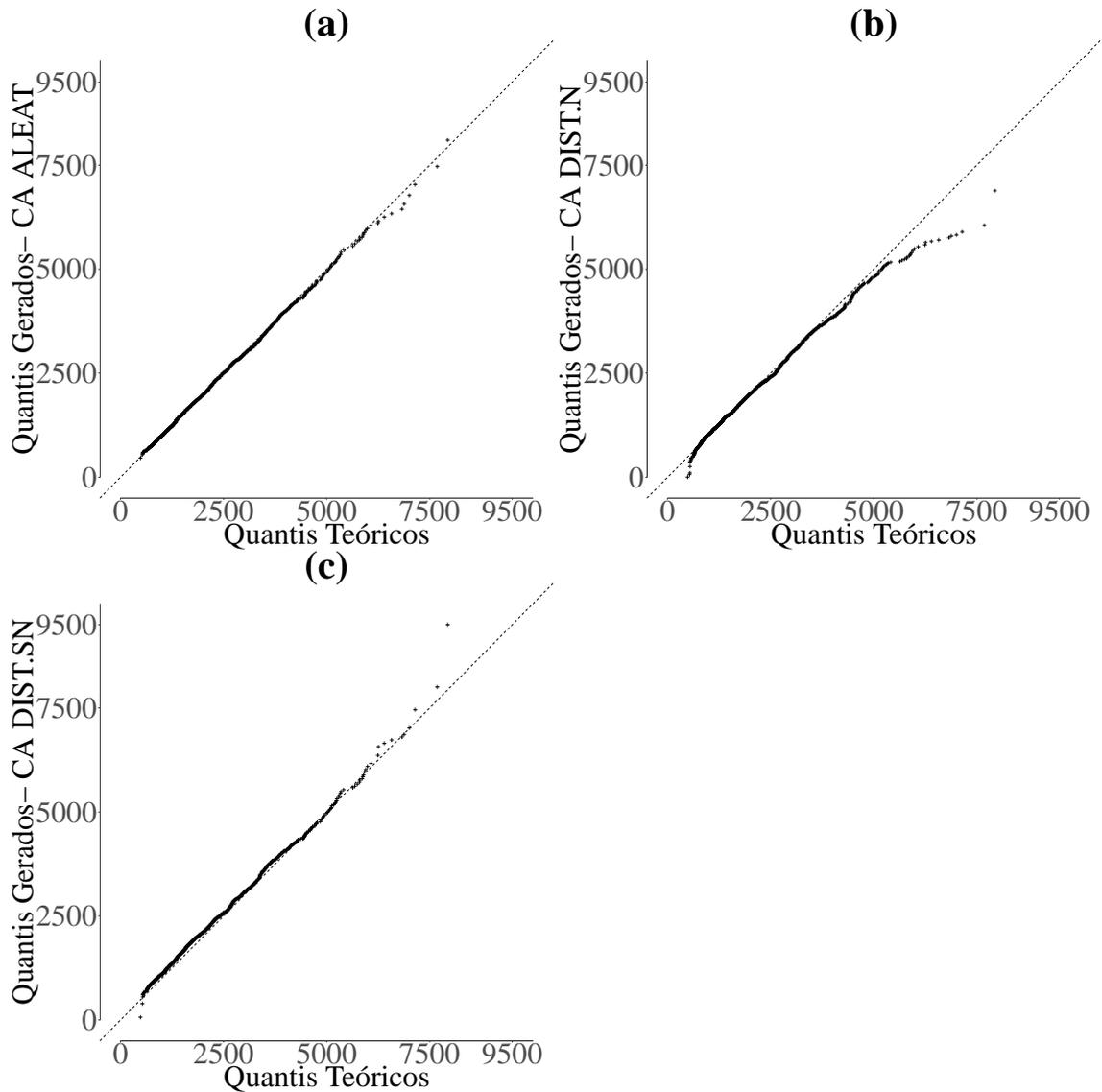


Legenda: a) Série Histórica
 b) CA ALEAT
 c) CA DIST.N
 d) CA DIST.SN

Fonte: Elaborada pelo autor.

Através dos q-qplots apresentados nas Figuras 35b e 35c fica mais claro ainda que ambas as novas propostas de modelagem para o CA não representaram bem aos dados, em especial nos valores extremos da série, onde mostra grande dificuldade em reproduzir esses valores quando comparado ao q-qplot da Figura 35a.

Figura 35 – Q-qplots das vazões geradas pelos modelos SVM para usina de Água Vermelha



Legenda: a) CA ALEAT
 b) CA DIST.N
 c) CA DIST.SN

Fonte: Elaborada pelo autor.

Comparando os erros da FAC e FDP, conforme a Tabela 5, pode-se concluir então que os modelos SVM CA DIST.N e DIST.S não foram uma opção tão viável quanto o modelo SVM CA ALEAT, visto que, como dito anteriormente, além de não conseguir replicar bem os *outliers*, a média e desvio padrão em alguns meses, além da curtose e assimetria em praticamente todos os meses, também apresentou um erro maior tanto na FAC quanto a FDP na maior parte dos casos.

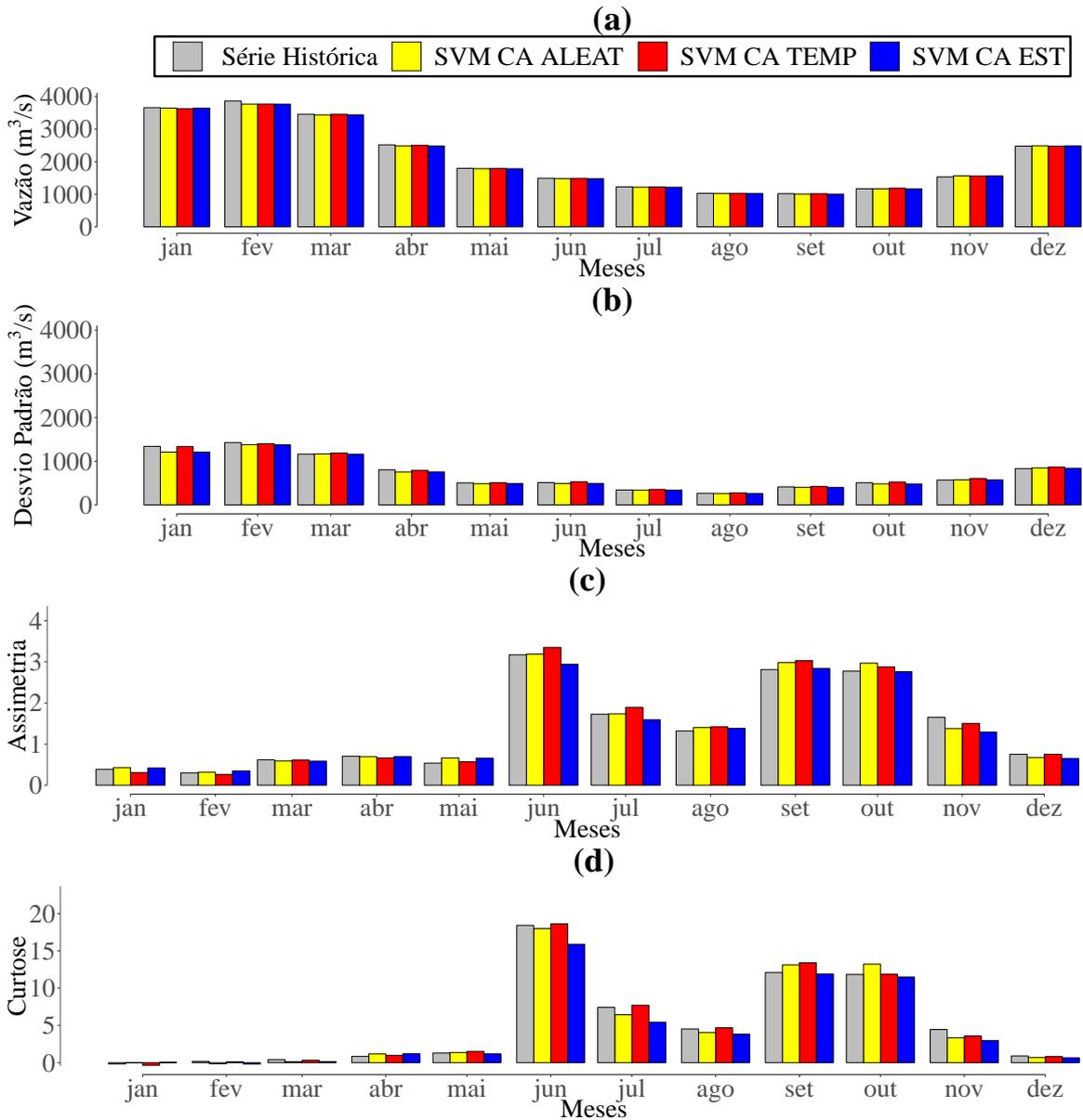
Tabela 5 – RMSE da FAC e FDP mensal dos modelos SVM de CA ALEAT, DIST.N e DIST.SN

RMSE da FAC			
Mês	SVM CA ALEAT	SVM CA DIST.N	SVM CA DIST.SN
janeiro	0,09	0,088	0,085
fevereiro	0,09	0,096	0,097
março	0,11	0,113	0,11
abril	0,081	0,084	0,092
maio	0,084	0,089	0,081
junho	0,075	0,066	0,073
julho	0,093	0,09	0,093
agosto	0,128	0,136	0,126
setembro	0,118	0,108	0,112
outubro	0,104	0,096	0,101
novembro	0,088	0,092	0,083
dezembro	0,119	0,128	0,116
Média dos meses	0,098	0,099	0,097
RMSE da FDP			
Mês	SVM CA ALEAT	SVM CA DIST.N	SVM CA DIST.SN
janeiro	0,026	0,028	0,037
fevereiro	0,026	0,026	0,028
março	0,033	0,037	0,054
abril	0,033	0,038	0,045
maio	0,024	0,032	0,032
junho	0,017	0,047	0,048
julho	0,028	0,035	0,04
agosto	0,024	0,022	0,053
setembro	0,02	0,081	0,048
outubro	0,03	0,053	0,055
novembro	0,023	0,038	0,065
dezembro	0,015	0,027	0,051
Média dos meses	0,025	0,039	0,046

Fonte: Elaborada pelo autor.

Analisando as estatísticas descritivas das outras duas propostas apresentadas nesse trabalho para essa técnica, as Figuras 36a e 36c mostram que os três modelos conseguiram de reproduzir bem tanto a média e o desvio padrão da série histórica do dados em todos os meses. O mesmo efeito pode ser visto na assimetria e curtose, onde todos modelos replicaram essas estatística de maneira razoável nos doze meses (ver Figuras 36c e 36d).

Figura 36 – Comparação entre as estatísticas descritivas das vazões geradas pelos modelos SVM de CA ALEAT, TEMP e EST para usina de Água Vermelha



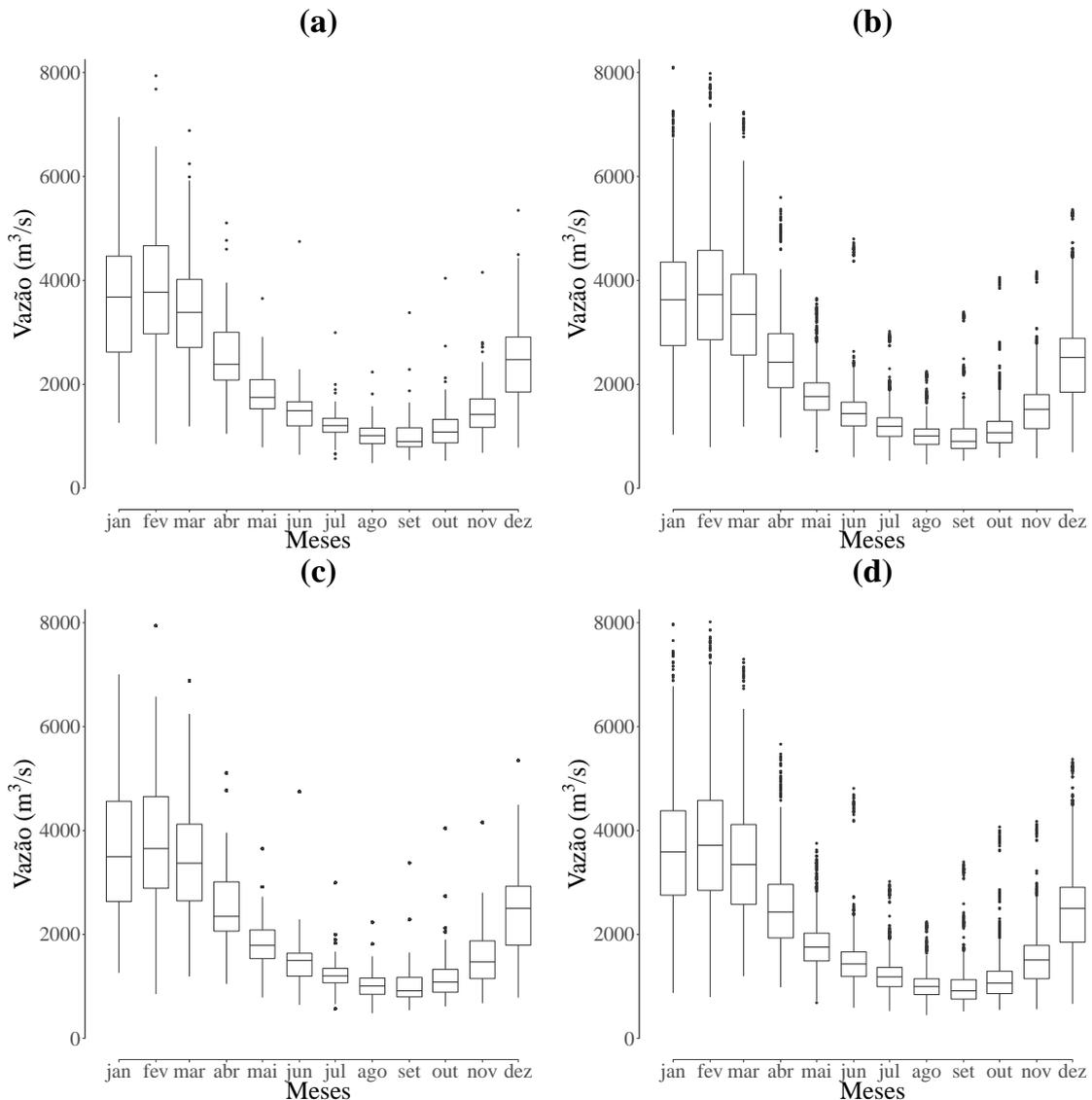
Legenda: a) Média
 b) Desvio Padrão
 c) Assimetria
 d) Curtose

Fonte: Elaborada pelo autor.

Através do box-plot das vazões geradas pelos modelos exibidos nas Figuras 37c e 37d, pode-se notar que o modelo com CA TEMP gera proporcionalmente poucos *outliers*, porém os valores gerados por ele se são bem próximos aos presentes na série histórica, inclusive com um *outlier* inferior no mês de julho, assim como na série histórica. Já o modelo com CA EST gerou *outliers* superiores com valores próximos aos presentes na série histórica, porém assim como o

modelo de CA ALEAT (ver Figura 37b), ele produziu *outliers* inferiores em meses onde isso não ocorre na série histórica, como por exemplo no mês de fevereiro.

Figura 37 – Box-plots das vazões geradas pelos modelos SVM para usina de Água Vermelha

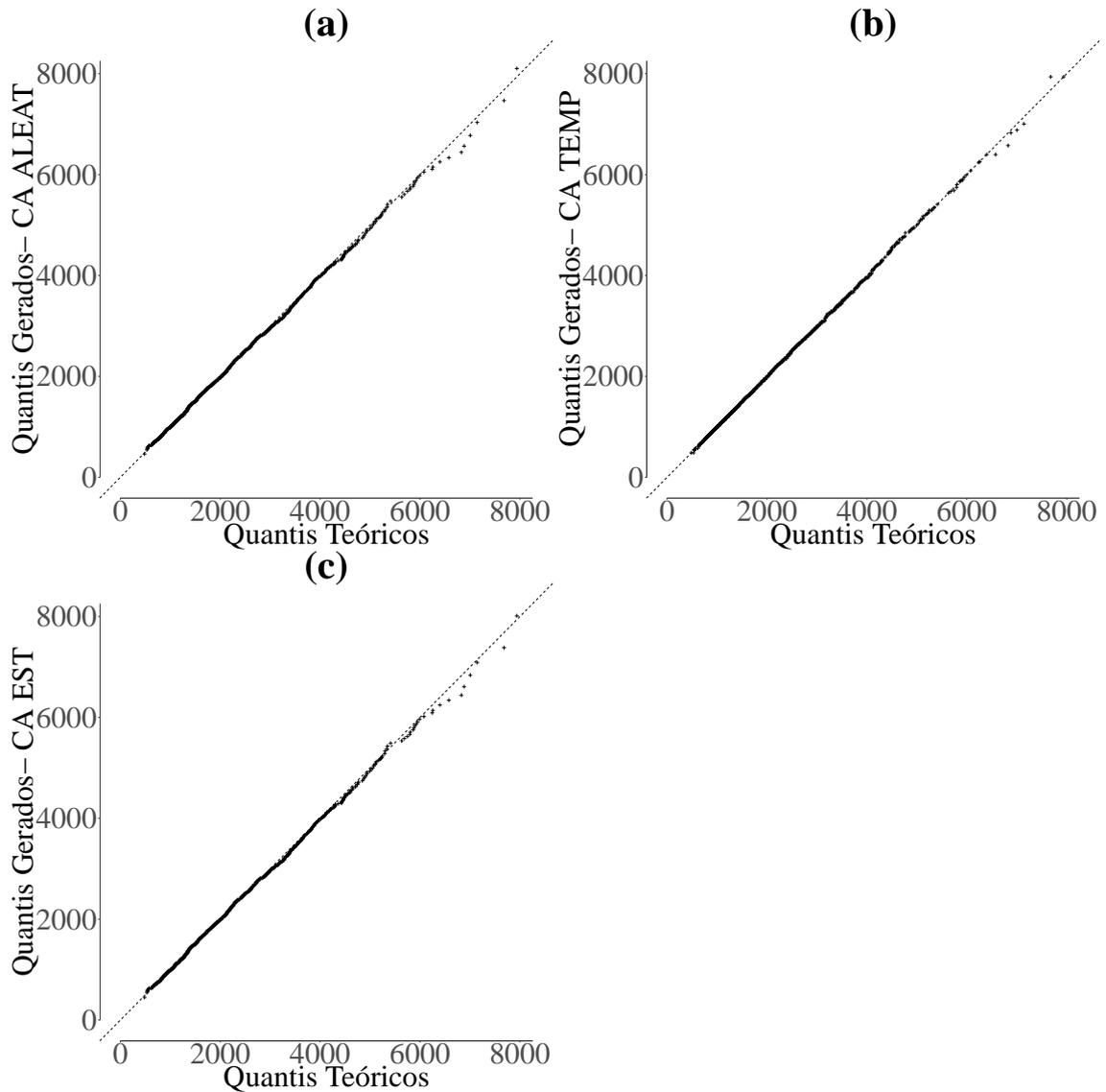


Legenda: a) Série Histórica
 b) CA ALEAT
 c) CA TEMP
 d) CA EST

Fonte: Elaborada pelo autor.

Através dos q-qplots apresentados nas Figuras 38b e 38c fica claro que ambas as propostas de modelagem para os CA se ajustaram tão bem aos dados quanto o modelo de CA ALEAT (ver Figura 38a), representando bem a série histórica durante maior parte, mas apresentando dificuldade nos quantis superiores.

Figura 38 – Q-qplots das vazões geradas pelos modelos SVM para usina de Água Vermelha



Legenda: a) CA ALEAT
 b) CA TEMP
 c) CA EST

Fonte: Elaborada pelo autor.

Por fim, explorando os RMSE da FAC e FDP desses três modelos (ver Tabela 6), constata-se que o modelo de CA TEMP apresentou um erro menor da FAC em todos os meses. O mesmo efeito ocorreu na FDP, onde o modelo de CA TEMP conseguiu diminuir o erro em todos os meses quando comparado aos demais, exceção feita ao mês de dezembro, onde o modelo de CA EST teve uma leve vantagem. No mês de junho destaca-se que o erro do modelo SVM CA TEMP atingiu a terceira casa decimal.

Tabela 6 – RMSE da FAC e FDP mensal dos modelos AR de CA ALEAT, TEMP e EST

RMSE da FAC			
Mês	SVM CA ALEAT	SVM CA TEMP	SVM CA EST
janeiro	0,09	0,056	0,091
fevereiro	0,09	0,074	0,09
março	0,11	0,028	0,111
abril	0,081	0,057	0,081
maio	0,084	0,042	0,083
junho	0,075	0,03	0,073
julho	0,093	0,037	0,092
agosto	0,128	0,043	0,127
setembro	0,118	0,045	0,117
outubro	0,104	0,041	0,103
novembro	0,088	0,029	0,086
dezembro	0,119	0,030	0,116
Média dos meses	0,098	0,043	0,097
RMSE da FDP			
Mês	SVM CA ALEAT	SVM CA TEMP	SVM CA EST
janeiro	0,026	0,016	0,027
fevereiro	0,026	0,010	0,032
março	0,033	0,017	0,035
abril	0,033	0,016	0,041
maio	0,024	0,013	0,024
junho	0,017	0,008	0,024
julho	0,028	0,010	0,033
agosto	0,024	0,011	0,025
setembro	0,020	0,012	0,021
outubro	0,030	0,011	0,027
novembro	0,023	0,016	0,022
dezembro	0,015	0,014	0,012
Média dos meses	0,025	0,013	0,027

Fonte: Elaborada pelo autor.

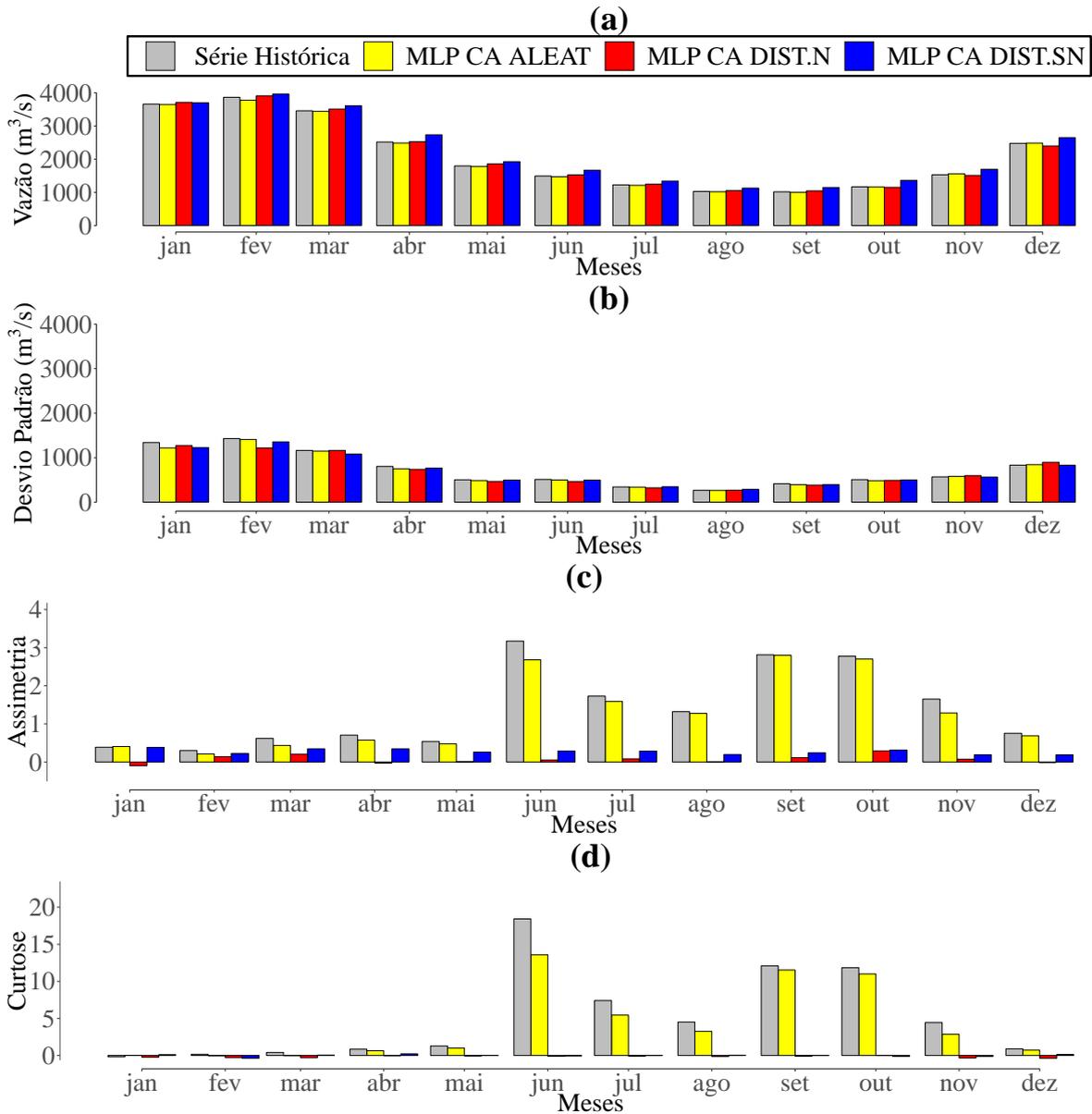
Portanto, concluí-se que para o modelo SVM, o uso da CA TEMP apresenta-se como uma alternativa mais adequada dentre todas as apresentadas nesta seção.

APÊNDICE C - MODELOS MLP PARA USINA DE ÁGUA VERMELHA

Os modelos MLPs gerados para a usina de Água Vermelha utilizaram seis passos anteriores, ou seja, a predição dos dados foi feita utilizando os últimos seis anos da série histórica.

A primeira das comparações executadas foi entre os modelos com CA DIST.N e DIST.SN. Pela Figuras 39a e 39b, percebe-se que os três modelos conseguiram de certa forma replicar de maneira eficaz tanto a média, quanto desvio padrão da série histórica em todos os meses. Porém, analisando as Figuras 39c e 39d, vê-se que os modelos com CA DIST.N e DIST.SN não conseguiram replicar bem tanto a curtose quanto assimetria em praticamente todos os meses, visto que esses modelos só conseguiram replicar de maneira razoável essas estatísticas quando os valores da série histórica eram baixos, como por exemplo nos meses de janeiro e fevereiro.

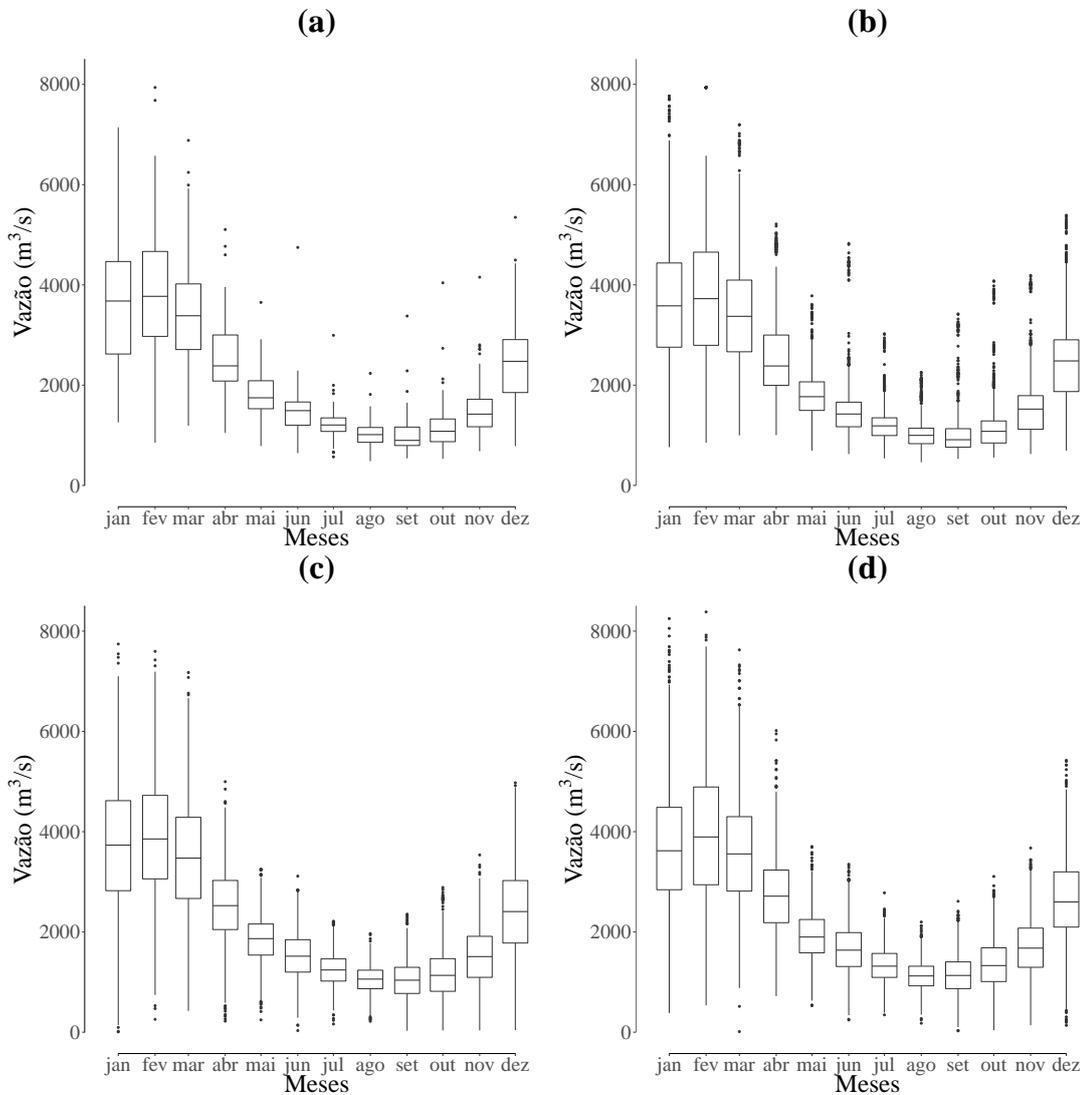
Figura 39 – Comparação entre as estatísticas descritivas das vazões geradas pelos modelos MLP de CA ALEAT, DIST.N e DIST.SN para usina de Água Vermelha



Nos box-plots das vazões geradas apresentados nas Figuras 40b, 40c e 40d, é exposta uma dificuldade quanto aos *outliers*. Mesmo o modelo com CA ALEAT, que conseguiu representar melhor a série histórica, nota-se que os *outliers* superiores conseguiram ser bem replicados, porém os *outliers* inferiores presentes no mês de julho não foram bem representados. Os modelos com CA DIST.N e DIST.SN até geraram esses *outliers* inferiores para o mês de ju-

lho, porém os demais meses também apresentaram *outliers* inferiores e a estrutura dos gráficos em alguns casos não ficaram parecidas com a série histórica conforme a Figura 40a.

Figura 40 – Box-plots das vazões geradas pelos modelos MLP para usina de Água Vermelha

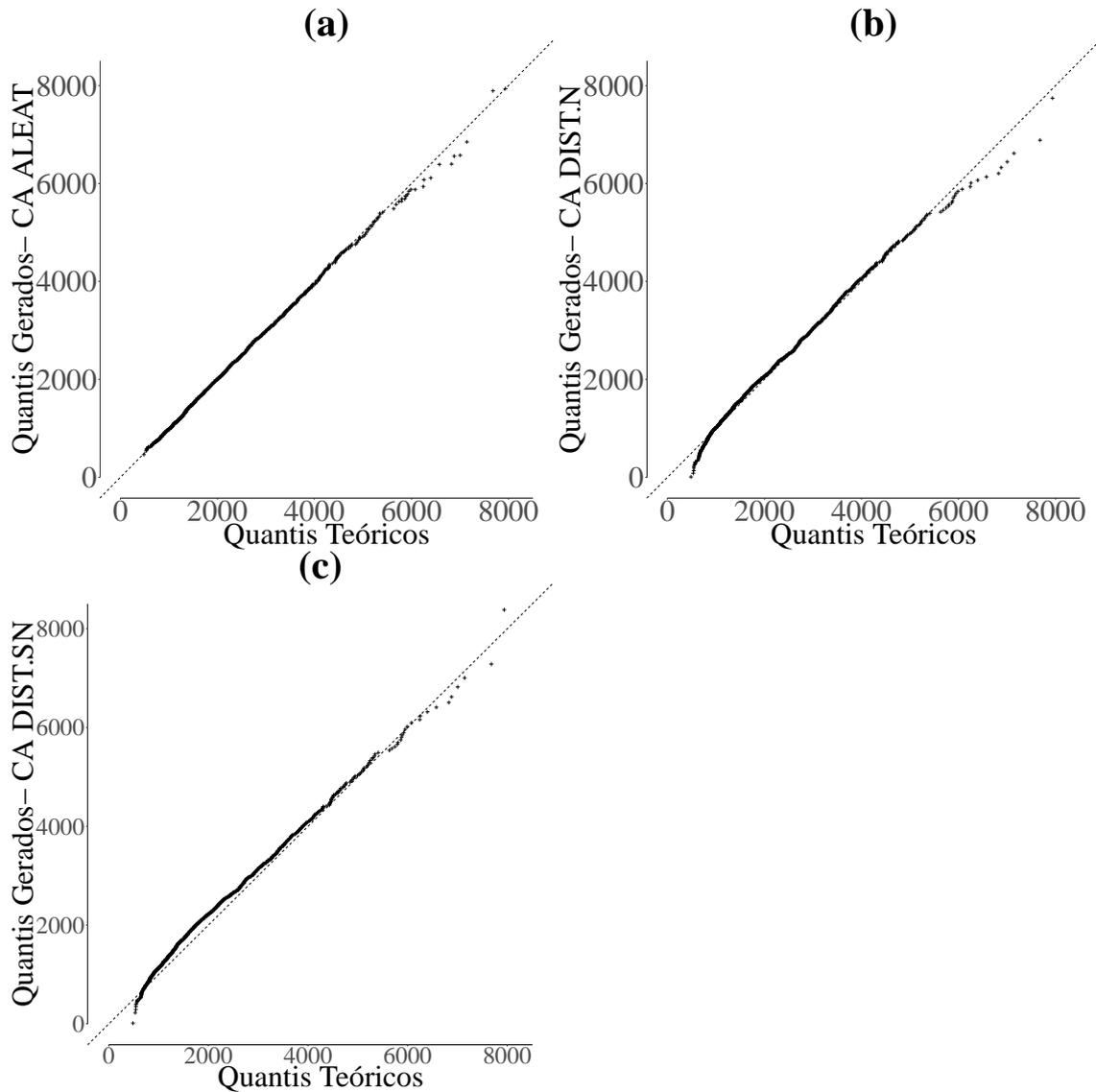


Legenda: a) Série Histórica
 b) CA ALEAT
 c) CA DIST.N
 d) CA DIST.SN

Fonte: Elaborada pelo autor.

Através dos q-qplots das Figuras 41b e 41c fica evidente que ambas as propostas de modelagem para os CA não se ajustaram bem aos dados, em especial nos valores extremos da série, onde mostraram uma dificuldade maior em reproduzir esses valores comparado ao q-qplot da Figura 41a.

Figura 41 – Q-qplots das vazões geradas pelos modelos MLP para usina de Água Vermelha



Legenda: a) CA ALEAT

b) CA DIST.N

c) CA DIST.SN

Fonte: Elaborada pelo autor.

Comparando os erros da FAC e FDP conforme a Tabela 7, pode-se concluir que os modelos MLP CA DIST.N e DIST.SN não são uma opção tão viável quanto o modelo MLP CA ALEAT, visto que como dito anteriormente, além de não conseguirem replicar bem os *outliers*, a curtose e assimetria em praticamente todos os meses, também apresentaram um maior erro na FDP em todos os meses.

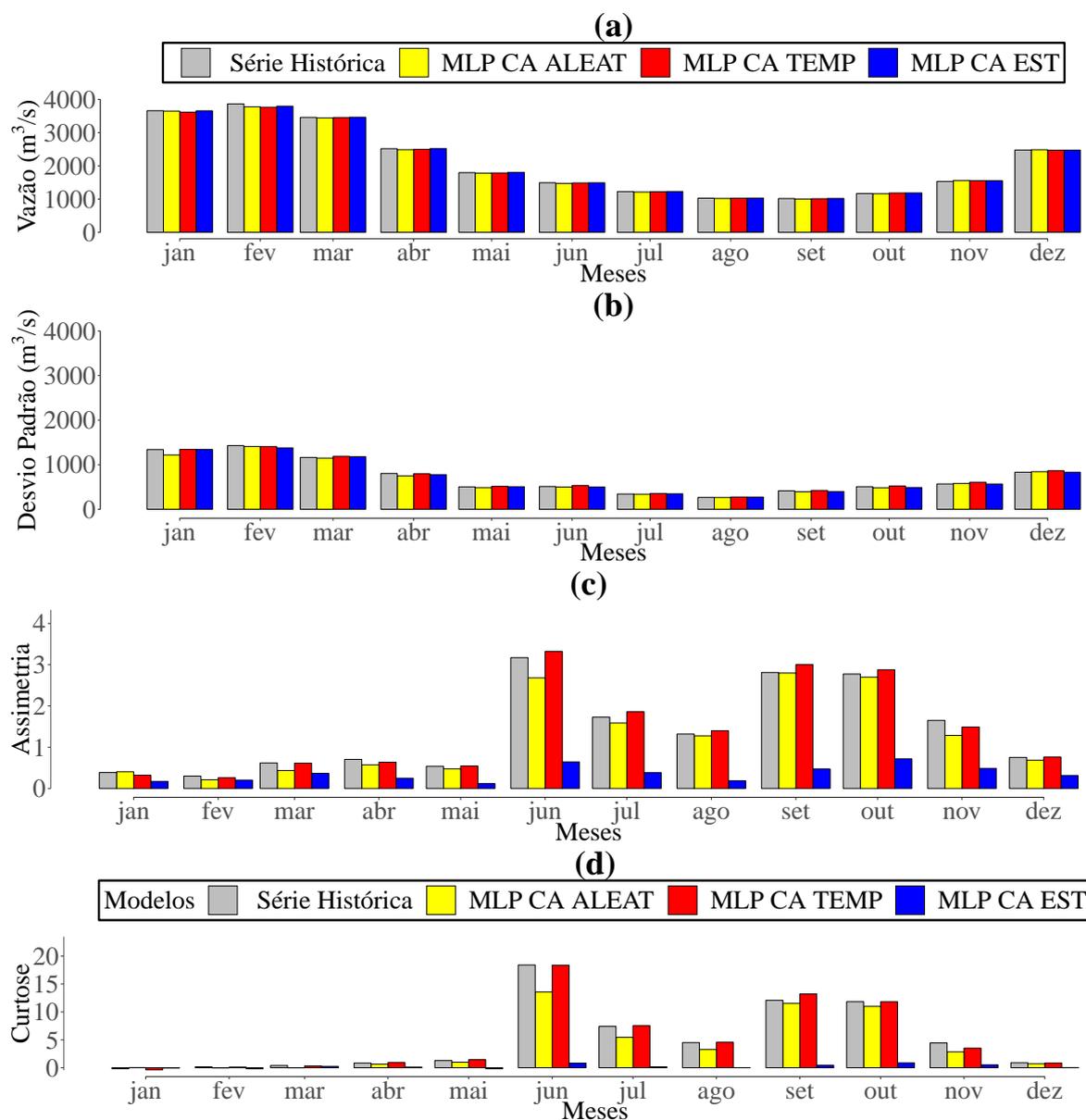
Tabela 7 – RMSE da FAC e FDP mensal dos modelos MLP de CA ALEAT, DIST.N e DIST.SN

RMSE da FAC			
Mês	MLP CA ALEAT	MLP CA DIST.N	MLP CA DIST.SN
janeiro	0,092	0,084	0,079
fevereiro	0,090	0,089	0,090
março	0,115	0,104	0,105
abril	0,079	0,088	0,093
maio	0,079	0,086	0,078
junho	0,070	0,065	0,072
julho	0,090	0,069	0,084
agosto	0,124	0,104	0,116
setembro	0,117	0,110	0,111
outubro	0,103	0,083	0,090
novembro	0,086	0,075	0,083
dezembro	0,117	0,106	0,104
Média dos meses	0,097	0,089	0,092
RMSE da FDP			
Mês	MLP CA ALEAT	MLP CA DIST.N	MLP CA DIST.SN
janeiro	0,026	0,036	0,027
fevereiro	0,018	0,025	0,031
março	0,023	0,034	0,027
abril	0,014	0,04	0,041
maio	0,008	0,032	0,033
junho	0,026	0,047	0,058
julho	0,030	0,047	0,066
agosto	0,026	0,047	0,051
setembro	0,012	0,072	0,082
outubro	0,025	0,065	0,084
novembro	0,031	0,060	0,054
dezembro	0,023	0,036	0,040
Média dos meses	0,022	0,045	0,050

Fonte: Elaborada pelo autor.

Analisando as estatísticas descritivas das outras duas propostas de CA apresentadas nesse trabalho para essa técnica, as Figuras 42a e 42b mostram que os três modelos conseguiram reproduzir bem tanto a média e o desvio padrão da série histórica do dados em todos os meses. Porém, quanto a assimetria e curtose, apenas o modelo de CA TEMP obteve resultados poupáveis para essas estatísticas (ver Figuras 42c e 42d), enquanto modelo de CA EST apresentou muita dificuldades para reproduzir essas características, principalmente nos meses de valores mais altos como junho, setembro e outubro.

Figura 42 – Comparação entre as estatísticas descritivas das vazões mensais geradas pelos modelos MLP de CA ALEAT, TEMP e EST para usina de Água Vermelha



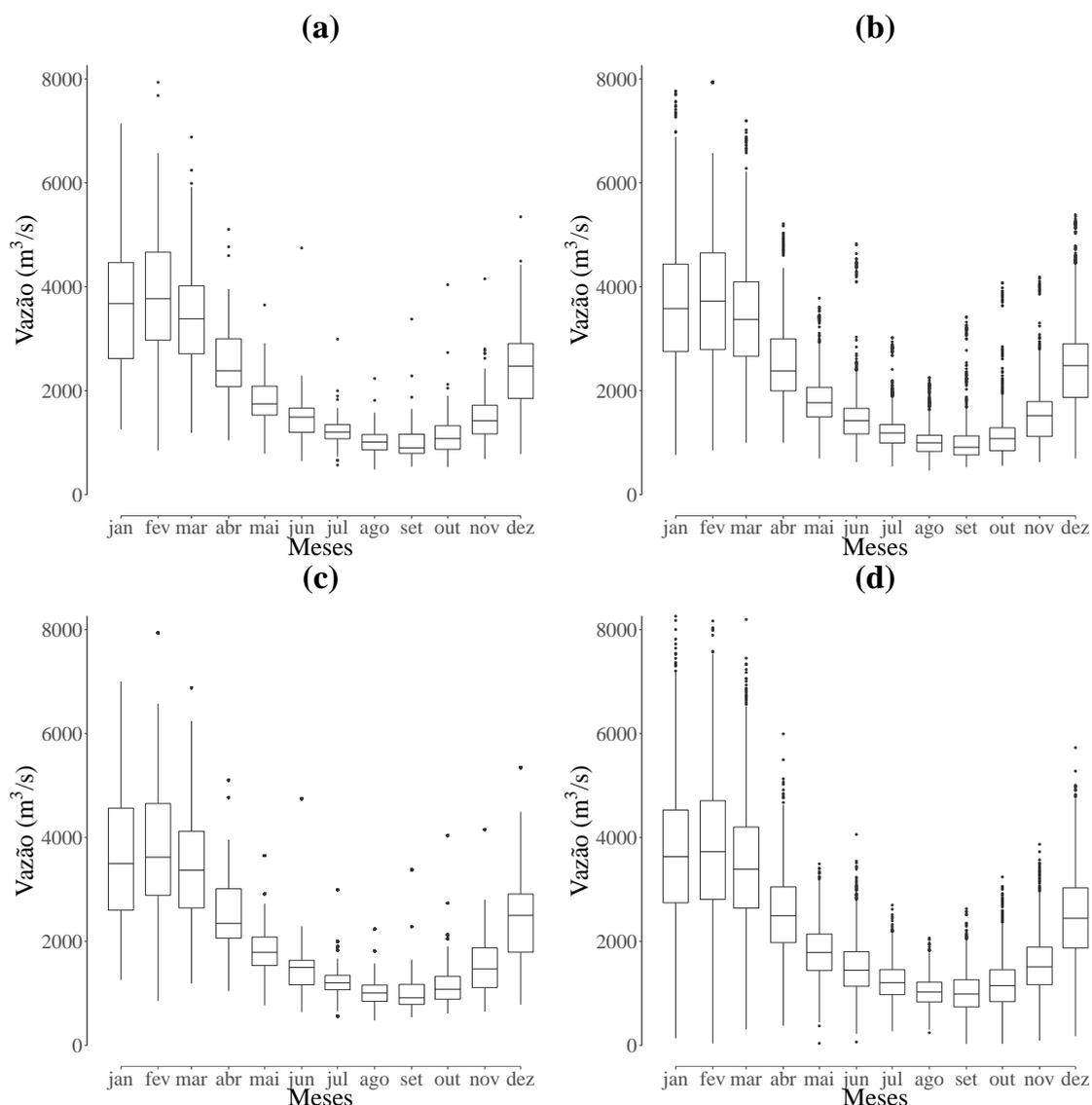
Legenda: a) Média
 b) Desvio Padrão
 c) Assimetria
 d) Curtose

Fonte: Elaborada pelo autor.

Através dos box-plots das vazões geradas pelos modelos exibidos nas Figuras 43c e 43d, nota-se que o modelo com CA TEMP gera proporcionalmente poucos *outliers*, porém os valores gerados por ele são bem próximos aos presentes na série histórica, inclusive com um *outlier* inferior no mês de julho, assim como na série histórica. Já o modelo com CA EST, além de ter apresentado *outliers* superiores com valores que não são semelhantes aos presentes

na série histórica (ver Figura 43a), também gerou *outliers* inferiores em meses onde isso não ocorrem, como em maio e junho, além de ter apresentado dificuldade quanto a amplitude nesses meses.

Figura 43 – Box-plots das vazões geradas pelos modelos MLP para usina de Água Vermelha



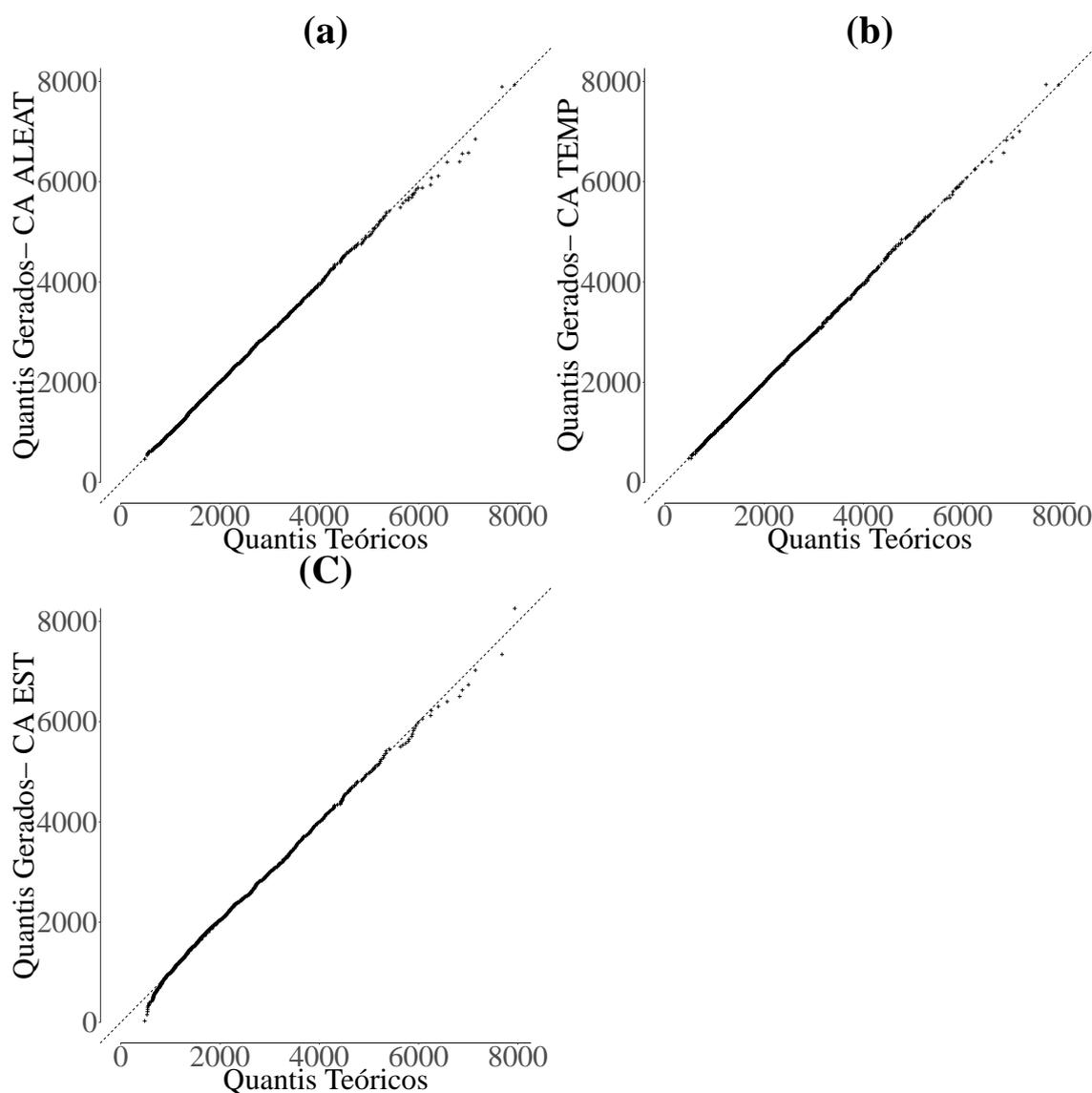
Legenda: a) Série Histórica
 b) CA ALEAT
 c) CA TEMP
 d) CA EST

Fonte: Elaborada pelo autor.

Através dos q-qplots apresentados nas Figuras 44b e 44c fica claro que a proposta de modelagem para os CA TEMP conseguiu representar melhor os dados da série histórica quando comparado ao modelo de CA ALEAT, principalmente nos quantis superiores, já o modelo de CA estimado, apresentou dificuldades em ambos os extremos dos dados, principalmente nos

inferiores (ver Figura 44a).

Figura 44 – Q-qplots das vazões geradas pelos modelos MLP para usina de Água Vermelha



Legenda: a) CA ALEAT
b) CA TEMP
c) CA EST

Fonte: Elaborada pelo autor.

Por fim, explorando os RMSE da FAC e FDP desses três modelos (ver Tabela 8), constata-se que o modelo de CA TEMP apresentou um erro menor da FAC em todos os meses. O mesmo efeito ocorreu na maior parte dos meses para a FDP, onde os erros foram menores ou muito próximos, com exceção do mês de maio cujo erro chegou até a terceira casa decimal. O modelo MLP CA EST não conseguiu resultados satisfatórios neste caso principalmente no erro da FDP, que foi significativamente superior aos outros dois modelos em praticamente todos os meses.

Tabela 8 – RMSE da FAC e FDP mensal dos modelos MLP de CA ALEAT, TEMPO e EST

RMSE da FAC			
Mês	MLP CA ALEAT	MLP CA TEMP	MLP CA EST
janeiro	0,092	0,055	0,085
fevereiro	0,090	0,076	0,094
março	0,115	0,029	0,108
abril	0,079	0,060	0,090
maio	0,079	0,046	0,076
junho	0,070	0,032	0,070
julho	0,090	0,039	0,074
agosto	0,124	0,045	0,113
setembro	0,117	0,050	0,114
outubro	0,103	0,042	0,093
novembro	0,086	0,032	0,081
dezembro	0,117	0,030	0,107
Média dos meses	0,097	0,045	0,092
RMSE da FDP			
Mês	MLP CA ALEAT	MLP CA TEMP	MLP CA EST
janeiro	0,026	0,016	0,025
fevereiro	0,018	0,020	0,026
março	0,023	0,017	0,033
abril	0,014	0,016	0,025
maio	0,008	0,019	0,043
junho	0,026	0,008	0,033
julho	0,030	0,010	0,059
agosto	0,026	0,011	0,03
setembro	0,012	0,012	0,058
outubro	0,025	0,005	0,046
novembro	0,031	0,016	0,036
dezembro	0,023	0,014	0,033
Média dos meses	0,022	0,014	0,037

Fonte: Elaborada pelo autor.

Portanto, concluí-se que o modelo MLP CA TEMP se mostrou uma alternativa melhor dentre todos os modelos apresentados nesta seção.

APÊNDICE D - MODELOS RF PARA USINA DE ÁGUA VERMELHA

Nos modelos *Random Forest* para geração de séries sintéticas de vazões para usina de Água Vermelha, várias estruturas foram testadas para os dados mensais com intuito de se obter a melhor estrutura para o modelo. A estrutura do modelo dos parâmetros na Seção 2.4.4 são exibidos na Tabela 9. O critério utilizado para seleção desses parâmetros foi o RMSE da FDP, onde a estrutura com menor erro foi a selecionada.

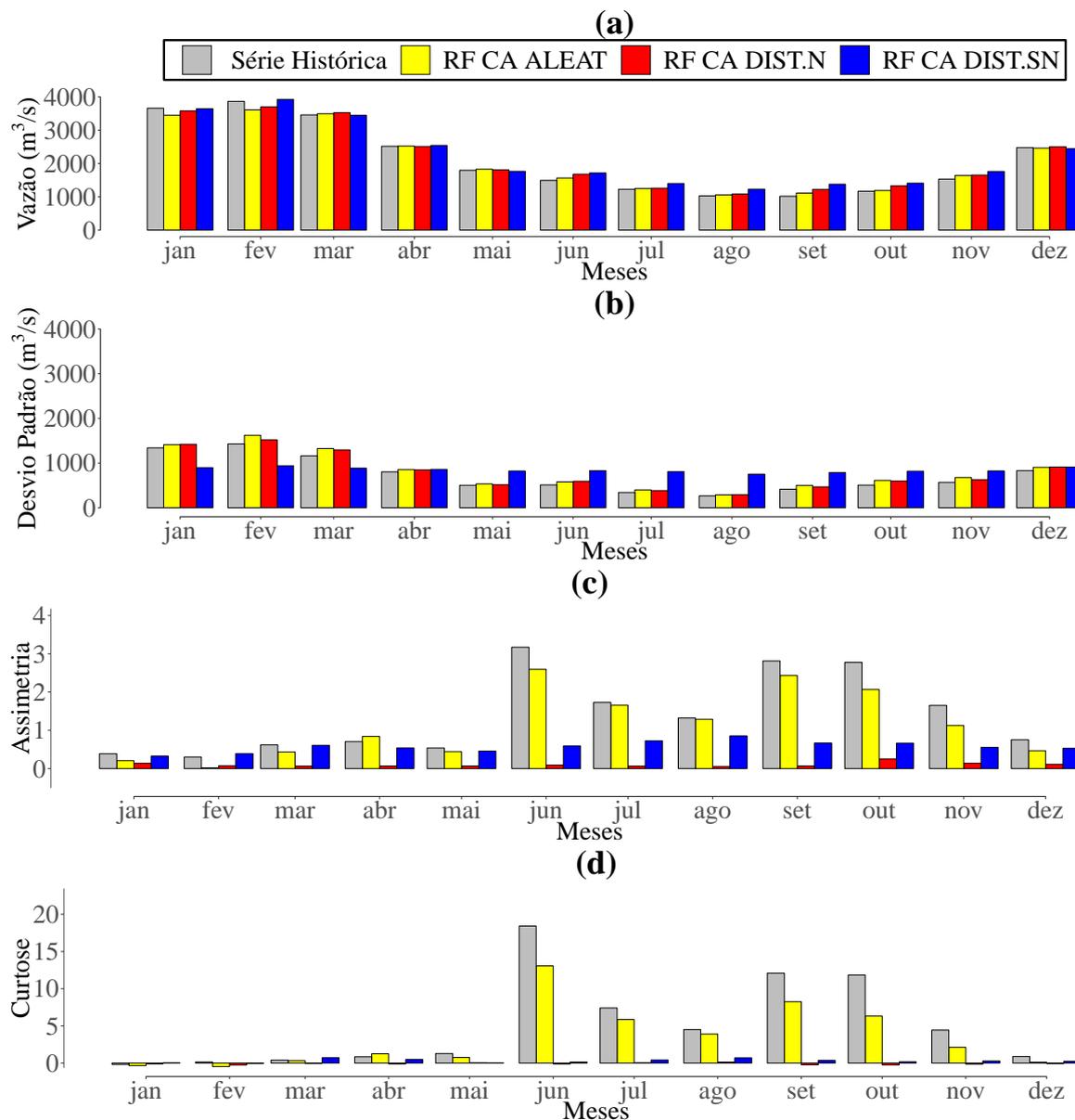
Tabela 9 – Estruturas do modelo *Random Forest* para geração das séries sintéticas mensais na usina de Água Vermelha

Mês	nodesize	ntree	nperm	Corr.Bies	Atrasos
janeiro	3	600	0,1	6	8
fevereiro	9	800	0,1	6	2
março	9	100	0,1	6	9
abril	9	300	0,1	6	9
maio	6	800	0,1	6	7
junho	9	400	0,1	6	11
julho	3	300	0,1	6	8
agosto	12	100	0,1	6	5
setembro	9	700	0,1	6	9
outubro	3	1000	0,1	6	6
novembro	6	500	0,1	6	9
dezembro	3	600	0,1	6	6

Fonte: Elaborada pelo autor.

Após selecionados os parâmetro, a primeira das comparações executadas foi entre os modelos com CA DIST.N e DIS.SN. A Figura 45a mostra que os três modelos até que conseguiram representar bem a série histórica das médias, mas quanto aos desvios padrão, na Figura45b já nota-se uma grande dificuldade do modelo de CA DIST.SN para replicar essa característica, conseguindo um resultado razoável somente no mês de dezembro. Mas nas Figuras 45c e 45d, vê-se que os modelos com CA DIST.N e DIST.SN não conseguiram replicar tanto a curtose quanto assimetria em praticamente todos os meses, em especial nos meses de junho, setembro e outubro, que possuem os maiores valores.

Figura 45 – Comparação entre as estatísticas descritivas das vazões geradas pelos modelos RF de CA ALEAT, DIST.N e DIST.SN para usina de Água Vermelha



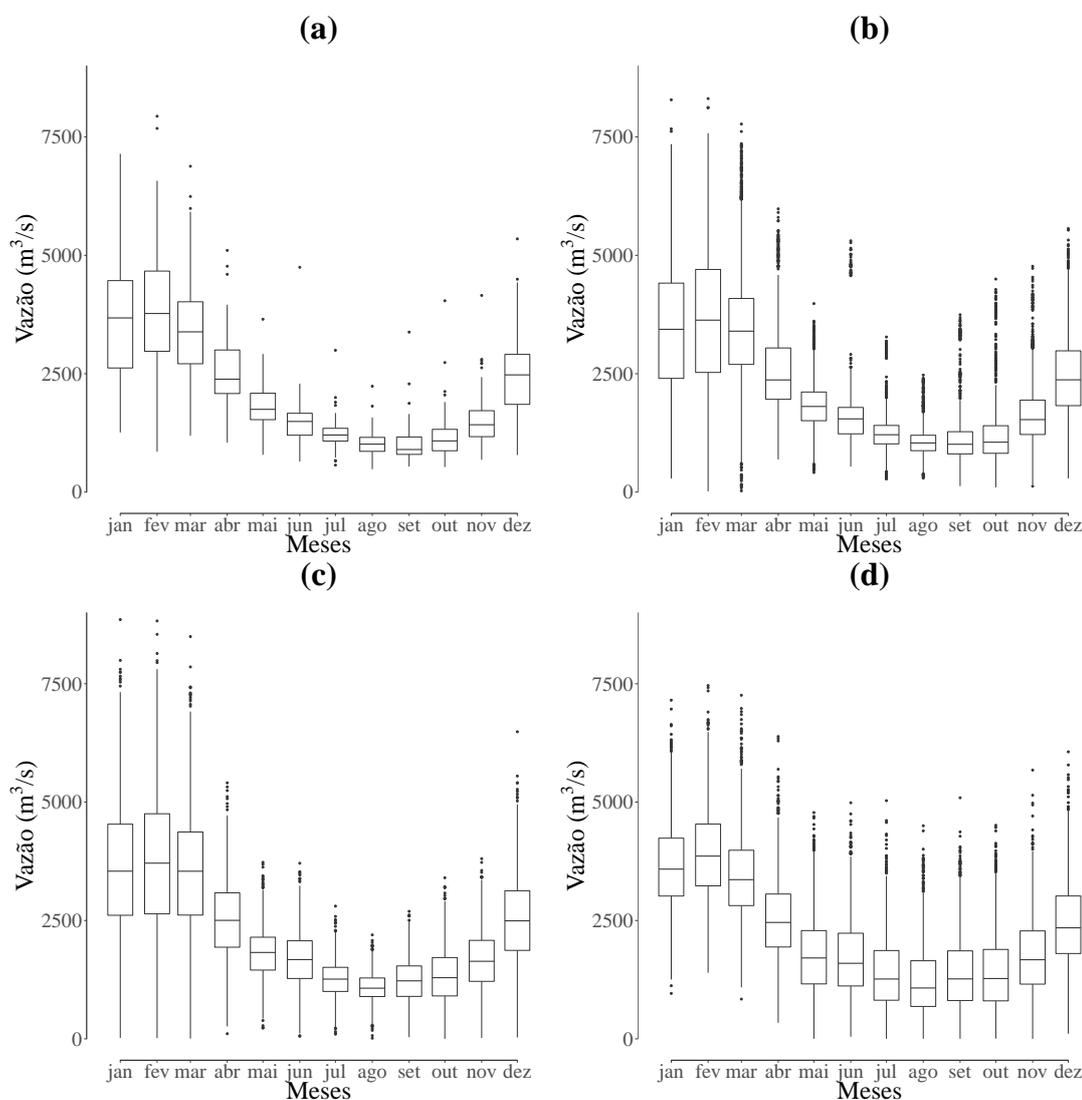
Legenda: a) Média
 b) Desvio Padrão
 c) Assimetria
 d) Curtose

Fonte: Elaborada pelo autor.

Nos box-plots das vazões geradas apresentados nas Figuras 46b, 46c e 46d, é exposta uma dificuldade quanto a representação dos *outliers*. Mesmo o modelo com CA ALEAT, que melhor representou a série histórica, nota-se que *outliers* inferiores foram gerados para os meses de fevereiro, maio e agosto, enquanto na série histórica da Figura 46a, apenas julho isso está presente, onde esses *outliers* inferiores foram bem representados nesse caso. O modelo com

CA DIST.N até gerou esses *outliers* inferiores para julho, mas também gerou em maio, junho, agosto e setembro, enquanto a CA DIST.SN gerou em janeiro e março. A amplitude dos meses também foi mal representada nesses dois modelos, em especial no mês de fevereiro.

Figura 46 – Box-plots das vazões geradas pelos modelos RF para usina de Água Vermelha

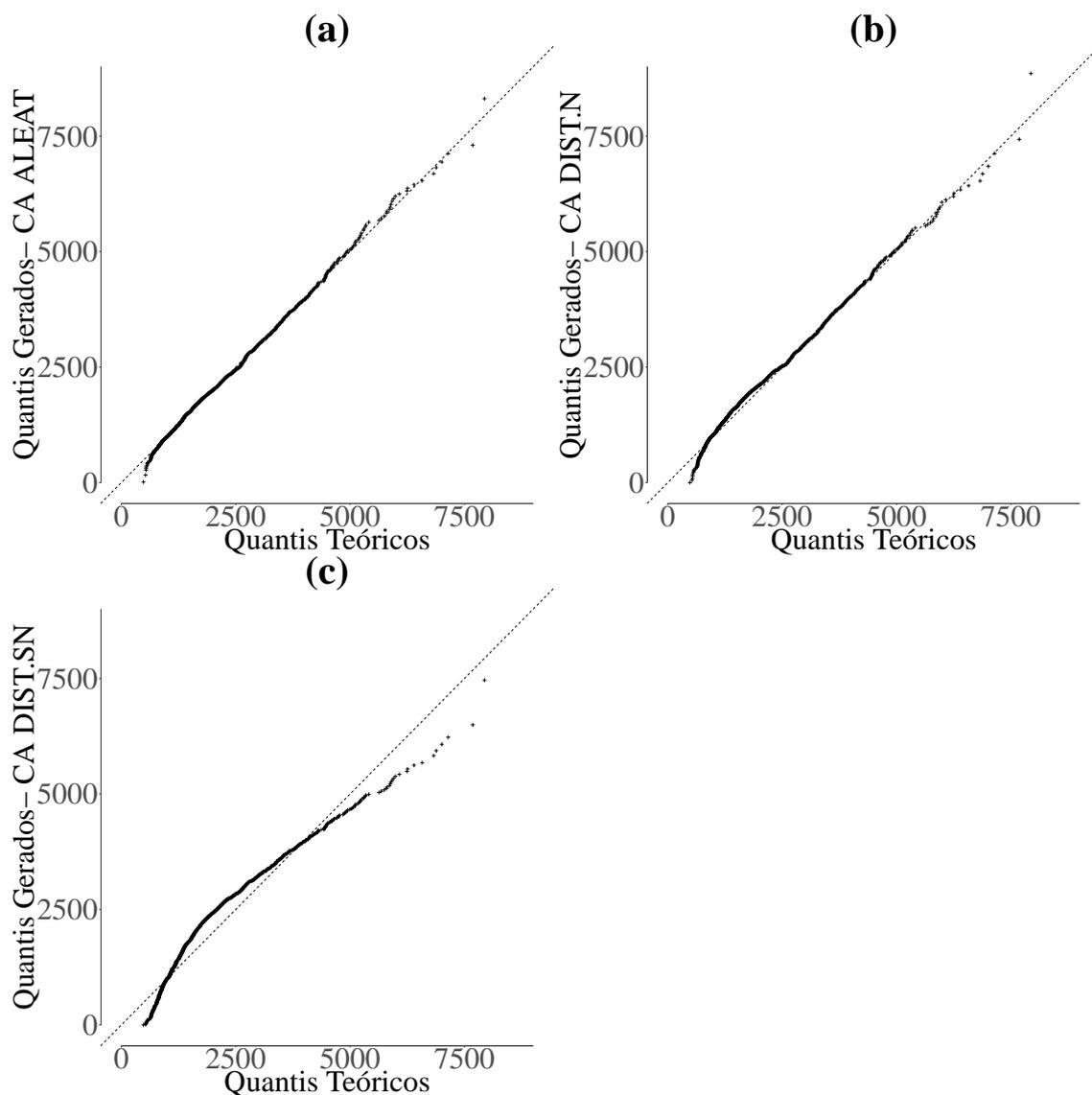


Legenda: a) Série Histórica
 b) CA ALEAT
 c) CA DIST.N
 d) CA DIST.SN

Fonte: Elaborada pelo autor.

Pelos q-qplots apresentados nas Figuras 47b e 47c fica mais claro ainda que ambas as propostas de modelagem para os CA não se ajustaram bem aos dados, em especial nos valores extremos positivos da série, onde mostram uma dificuldade maior em reproduzir esses valores quando comparado ao q-plot da Figura 47a.

Figura 47 – Q-qplots das vazões geradas pelos modelos RF para usina de Água Vermelha



Legenda: a) CA ALEAT
 b) CA DIST.N
 c) CA DIST.SN

Fonte: Elaborada pelo autor.

Comparando os erros da FAC e FDP conforme a Tabela 10, concluí-se que os modelos RF CA DIST.N e DIST.SN não foram uma opção tão viável quanto o modelo com CA ALEAT, visto que, como dito anteriormente, além de não conseguirem replicar bem os *outliers*, desvios padrão, curtoses e assimetrias, vários meses também apresentaram erros da FAC superiores ou muito próximos ao modelo com CA ALEAT, além de erros bem superiores quanto a FDP.

Tabela 10 – RMSE da FAC e FDP mensal dos modelos RF de CA ALEAT, DIST.N e DIST.SN

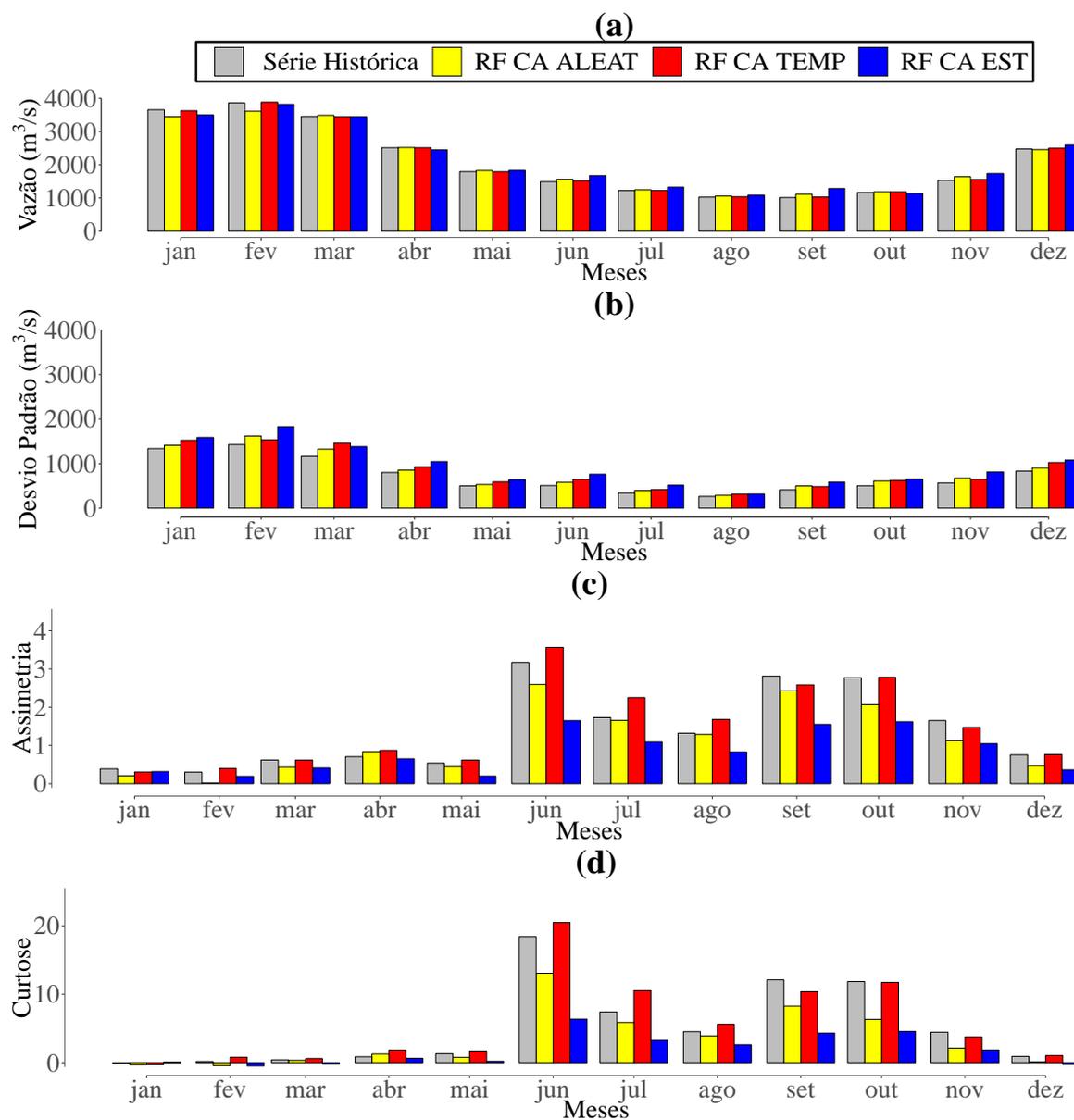
RMSE da FAC			
Mês	RF CA ALEAT	RF CA DIST.N	RF CA DIST.SN
janeiro	0,078	0,064	0,077
fevereiro	0,084	0,092	0,097
março	0,083	0,093	0,093
abril	0,088	0,088	0,086
maio	0,072	0,079	0,082
junho	0,062	0,067	0,062
julho	0,058	0,067	0,079
agosto	0,080	0,086	0,117
setembro	0,080	0,096	0,093
outubro	0,076	0,121	0,069
novembro	0,057	0,049	0,068
dezembro	0,085	0,095	0,078
Média dos meses	0,075	0,083	0,083
RMSE da FDP			
Mês	RF CA ALEAT	RF CA DIST.N	RF CA DIST.SN
janeiro	0,018	0,029	0,052
fevereiro	0,038	0,032	0,052
março	0,021	0,046	0,03
abril	0,012	0,049	0,019
maio	0,013	0,026	0,078
junho	0,027	0,072	0,100
julho	0,011	0,070	0,138
agosto	0,030	0,051	0,129
setembro	0,039	0,101	0,136
outubro	0,041	0,093	0,109
novembro	0,035	0,061	0,083
dezembro	0,026	0,037	0,034
Média dos meses	0,026	0,056	0,080

Fonte: Elaborada pelo autor.

Verificando as outras duas propostas apresentadas nesse trabalho, através da análise das estatísticas descritivas presente nas Figuras 48a, 48b, 48c e 48d mostram que os três modelos conseguiram reproduzir bem a média da série. Em relação ao desvio padrão, os modelos com CA TEMP e EST apresentaram uma maior dificuldade quando comparado ao modelo com CA ALEAT para representação dessa característica. Quanto a assimetria e curtose, o modelo com CA EST não apresentou bons resultados principalmente nos meses de maior valor, já o modelo dos CA TEMP conseguiu replicar essa característica de maneira mais fiel do que os demais modelos testados, com exceção a curtose no mês de março que foi positivo enquanto na série real e nos demais modelos foram negativos. Nos meses de maiores valores, como junho, setembro e

outubro, fica mais nítida essa vantagem no modelo RF CA TEMP quanto a representação dessas estatísticas.

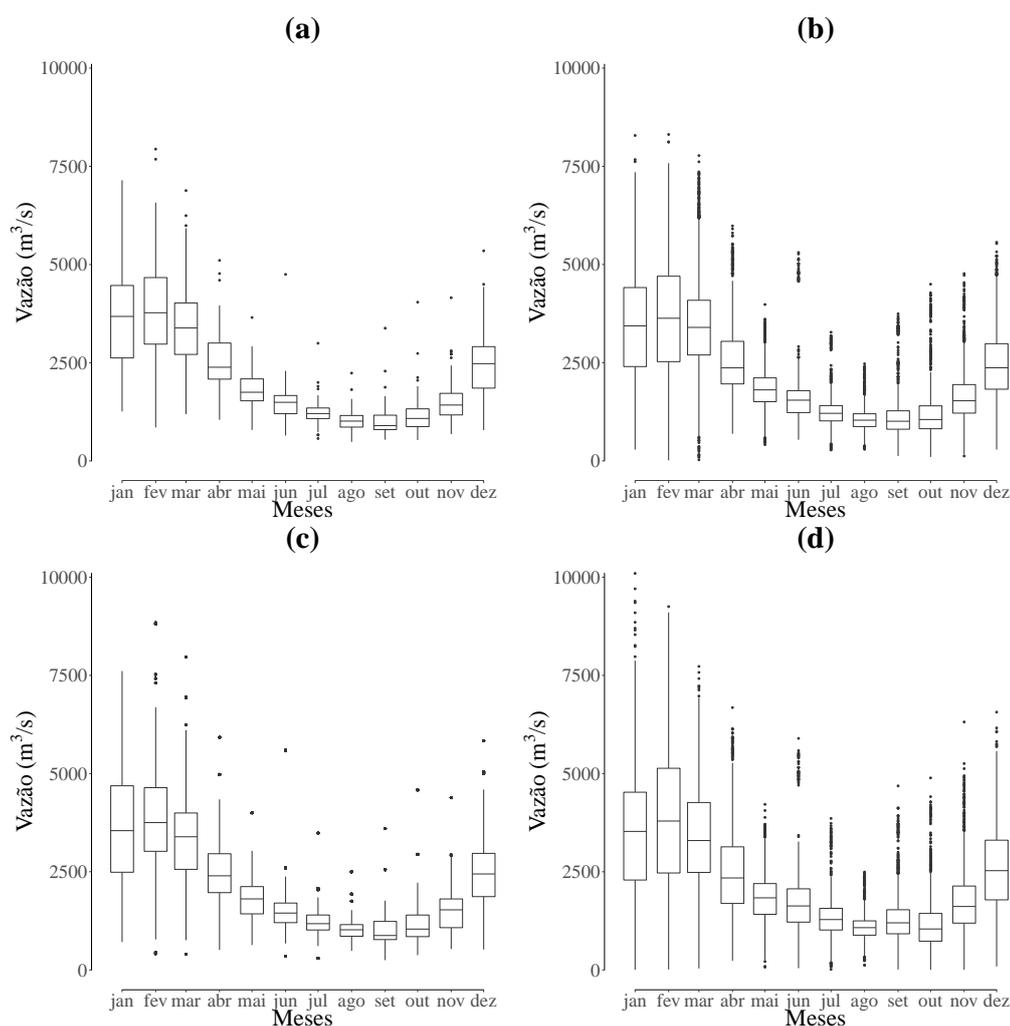
Figura 48 – Comparação entre estatísticas descritivas das vazões geradas pelos modelos RF de CA ALEAT, TEMP e EST para usina de Água Vermelha



Através do box-plot dos modelos exibidos nas Figuras 49c e 49d, pode-se notar que o modelo com CA TEMP gera proporcionalmente poucos *outliers*, porém os valores gerados por ele se são bem próximos presentes na série histórica, inclusive com um *outliers* inferiores

no mês de julho, assim como na série histórica, mas apresentando problemas semelhantes ao do modelo com CA ALEAT, visto os *outliers* inferiores nos meses de fevereiro, março e junho, inexistentes na série histórica. Já o modelo com CA EST produziu *outliers* superiores destoantes dos presentes na série histórica, mostrando grande dificuldade quanto a amplitude, em especial nos meses de janeiro e fevereiro, além de também gerar *outliers* inferiores em meses onde isso não ocorre na série histórica, como maio e agosto.

Figura 49 – Box-plots das vazões geradas pelos modelos RF para usina de Água Vermelha



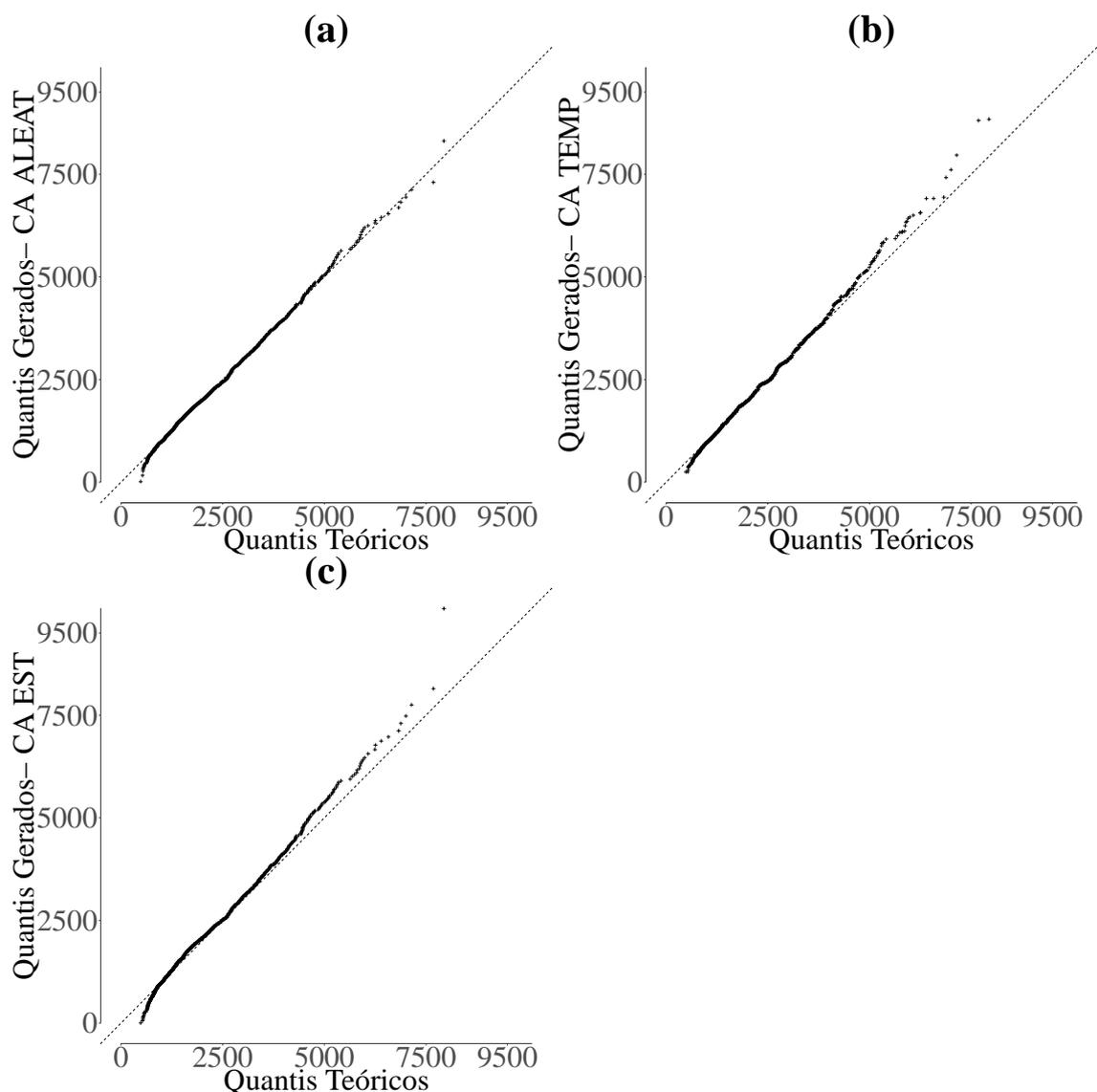
Legenda: a) Série Histórica
 b) CA ALEAT
 c) CA TEMP
 d) CA EST

Fonte: Elaborada pelo autor.

Através dos q-qplots apresentados nas Figuras 50b fica claro a proposta de modelagem utilizando CA TEMP se ajustou tão bem aos dados quanto o modelo de CA ALEAT (ver Figura 50a), apresentando dificuldades quanto aos quantis superiores, porém se ajustando bem a série

histórica nos demais valores. Já o modelo com CA EST (ver Figura 50c) além da maior dificuldade quanto aos quantis mais altos, apresenta também na representação de valores menores da série.

Figura 50 – Q-plots das vazões geradas pelos modelos RF para usina de Água Vermelha



Legenda: a) CA ALEAT
 b) CA TEMP
 c) CA EST

Fonte: Elaborada pelo autor.

Por fim, analisando os RMSE da FAC e FDP desses três modelos (ver Tabela 11), constata-se que o modelo de CA ao longo do tempo apresentou um erro menor da FAC em todos os meses, menos em julho. Na FDP, também houve essa tendência de melhora do erro nesse modelo, visto que conseguiu diminuir significativamente esse erro quanto ao modelo de CA aleatório em meses como setembro, outubro e agosto, obtendo erros próximos nos demais

meses, mas ainda sim menor na maioria dos casos.

Tabela 11 – RMSE da FAC e FDP mensal dos modelos RF de CA ALEAT, TEMP e EST

RMSE da FAC			
Mês	RF CA ALEAT	RF CA TEMP	RF CA EST
janeiro	0,078	0,061	0,072
fevereiro	0,084	0,061	0,095
março	0,083	0,056	0,084
abril	0,088	0,070	0,093
maio	0,072	0,041	0,062
junho	0,062	0,064	0,074
julho	0,058	0,080	0,052
agosto	0,080	0,040	0,084
setembro	0,08	0,036	0,100
outubro	0,076	0,046	0,069
novembro	0,057	0,066	0,055
dezembro	0,085	0,047	0,099
Média dos meses	0,075	0,056	0,078
RMSE da FDP			
Mês	RF CA ALEAT	RF CA TEMP	RF CA EST
janeiro	0,018	0,016	0,045
fevereiro	0,038	0,014	0,051
março	0,021	0,024	0,029
abril	0,012	0,019	0,045
maio	0,013	0,016	0,041
junho	0,027	0,012	0,088
julho	0,011	0,013	0,050
agosto	0,03	0,011	0,044
setembro	0,039	0,009	0,101
outubro	0,041	0,012	0,067
novembro	0,035	0,014	0,059
dezembro	0,026	0,023	0,045
Média dos meses	0,026	0,015	0,055

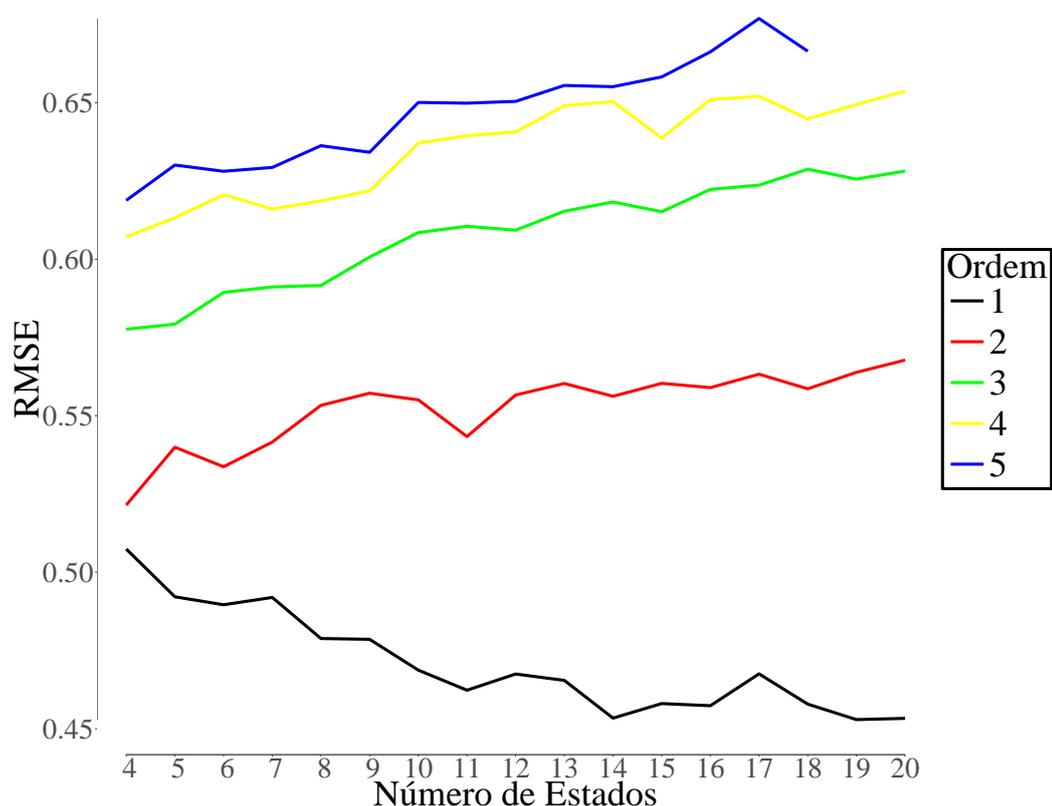
Fonte: Elaborada pelo autor.

Portanto, concluí-se que para a técnica *Random Forest*, o modelo de CA TEMP se mostrou a melhor opção entre os demais modelos, visto que possui menor erro da FDP e FAC na maioria dos casos, obteve valores de assimetria a curtose mais próximos ao da série real e conseguiu estimar de maneira mais eficiente os *outliers* inferiores quando comparado com o modelo de CA ALEAT.

APÊNDICE E - MODELO SYNTISE PARA USINA DE ÁGUA VERMELHA

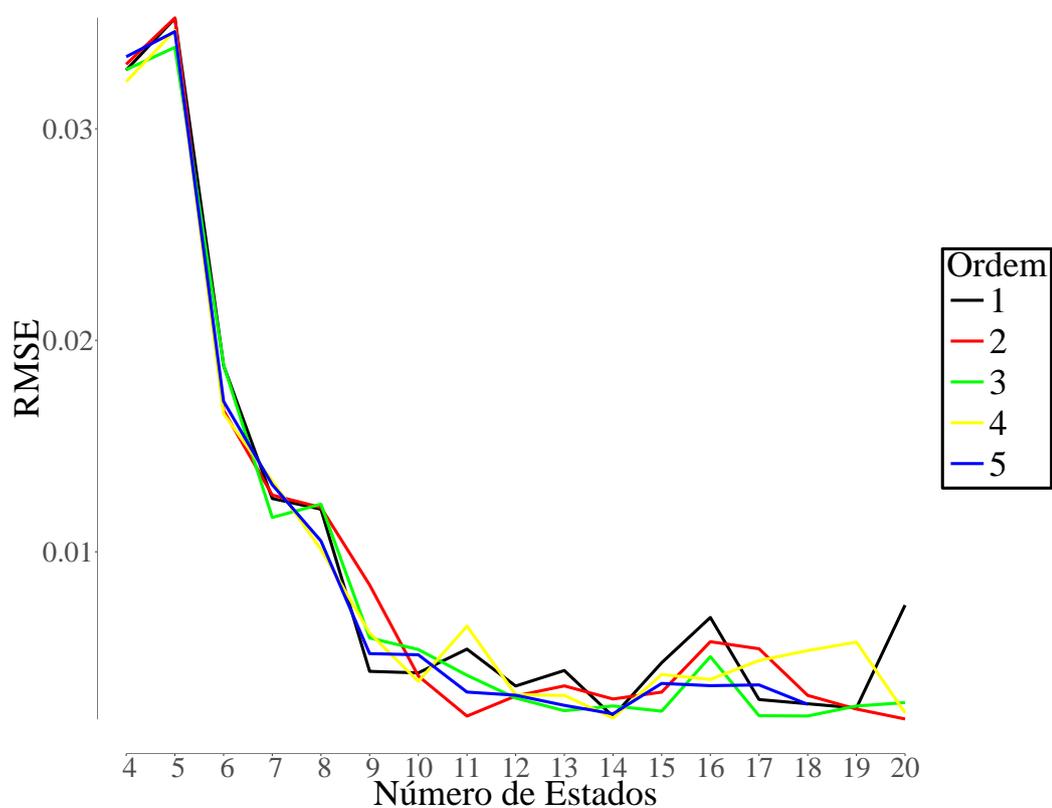
Para realizar a geração da série sintética com o SynTise, antes é necessário definir quais são a ordem e estado mais adequada para o modelo. Para decidir isso, modelos com ordem de 1 até 5 e estados de 4 até 20 foram gerados e comparado quanto ao RMSE das FAC e FDP. Analisando o erro da FAC através da Figura 51, fica evidente que a melhor ordem para o modelo é 1, já que seu erro foi menor que os modelos de demais ordem em todos os estados. Quanto ao estado, através do erro da FDP (ver Figura 52), nota-se que o menor entre os erros mostra ser o de 14 estados. Portanto assim o modelo gerado para essa usina será de ordem 1 e 14 estados.

Figura 51 – RMSE da FAC dos modelos de série possíveis no SynTise para usina de Água Vermelha



Fonte: Elaborada pelo autor.

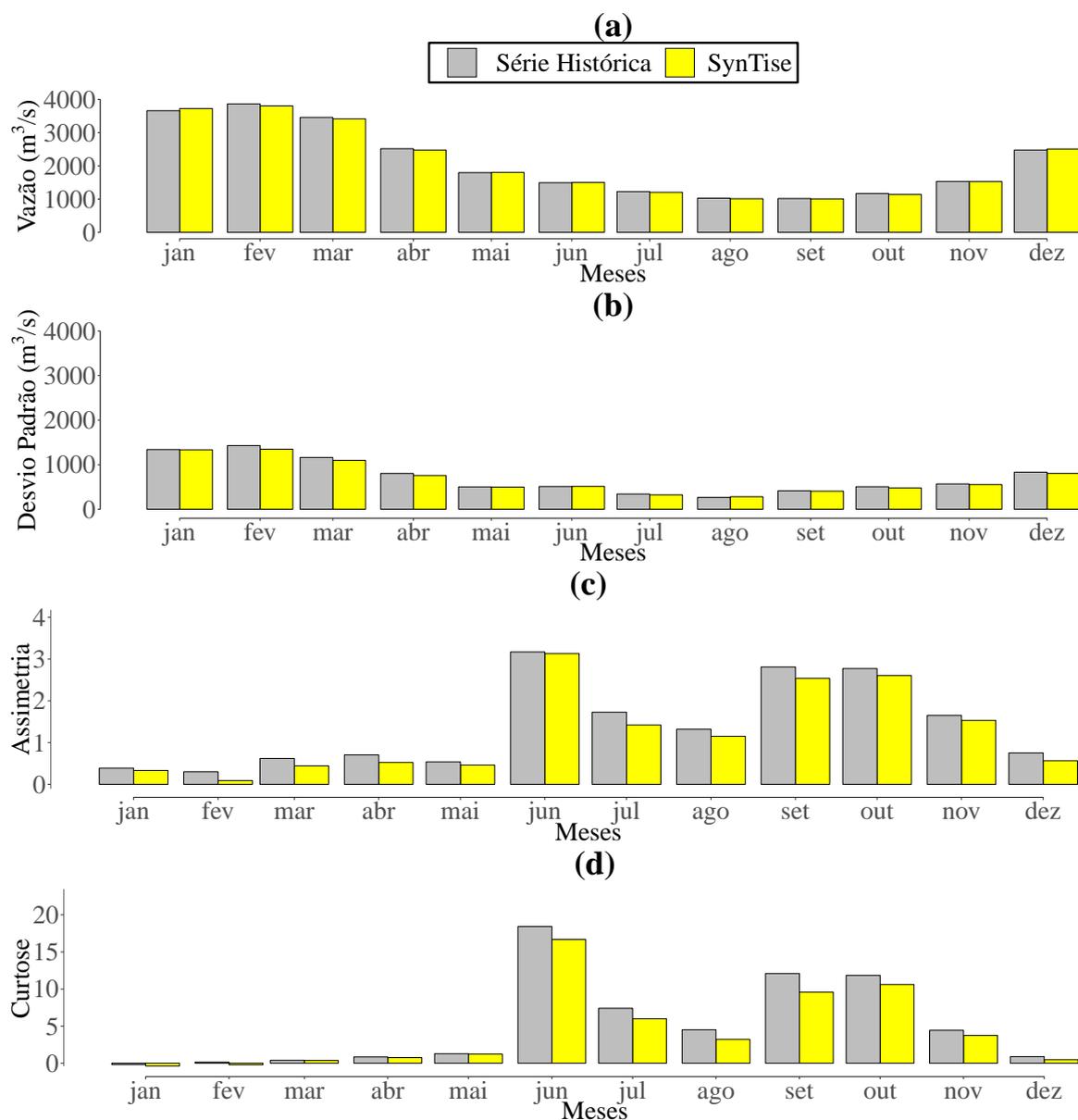
Figura 52 – RMSE da FDP dos modelos de série possíveis no SynTise para usina de Água Vermelha



Fonte: Elaborada pelo autor.

Analisando as estatísticas descritivas do modelo, nota-se que as médias e desvios padrão foram muito bem representadas no SynTise (ver Figuras 53a e 53b). Mesmo não tendo a mesma precisão quanto a estimativa da curtose e assimetria, ainda sim essas estatísticas foram bem replicadas, conforme as Figuras 53c e 53d.

Figura 53 – Comparação entre as estatísticas descritivas das vazões mensais geradas pelo modelo SynTise para usina de Água Vermelha

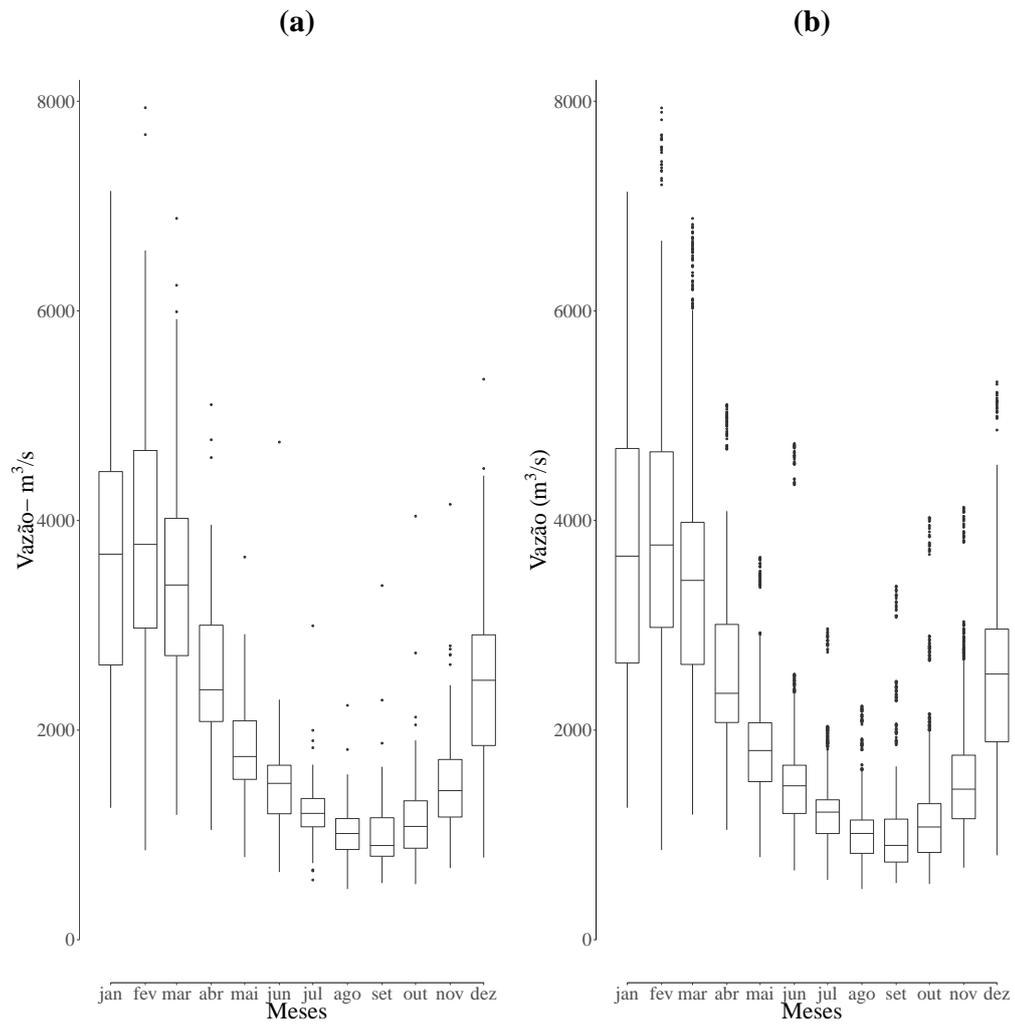


Legenda: a) Média
 b) Desvio Padrão
 c) Assimetria
 d) Curtose

Fonte: Elaborada pelo autor.

O box-plot e q-qplot das Figuras 54b e 55, respectivamente, mostram como que a estrutura da série histórica foi replicada de maneira fidedigna pelo modelo, inclusive nos *outliers*, que em todos os meses foram bem representados. Porém o modelo falhou em replicar os *outliers* inferiores do mês de julho.

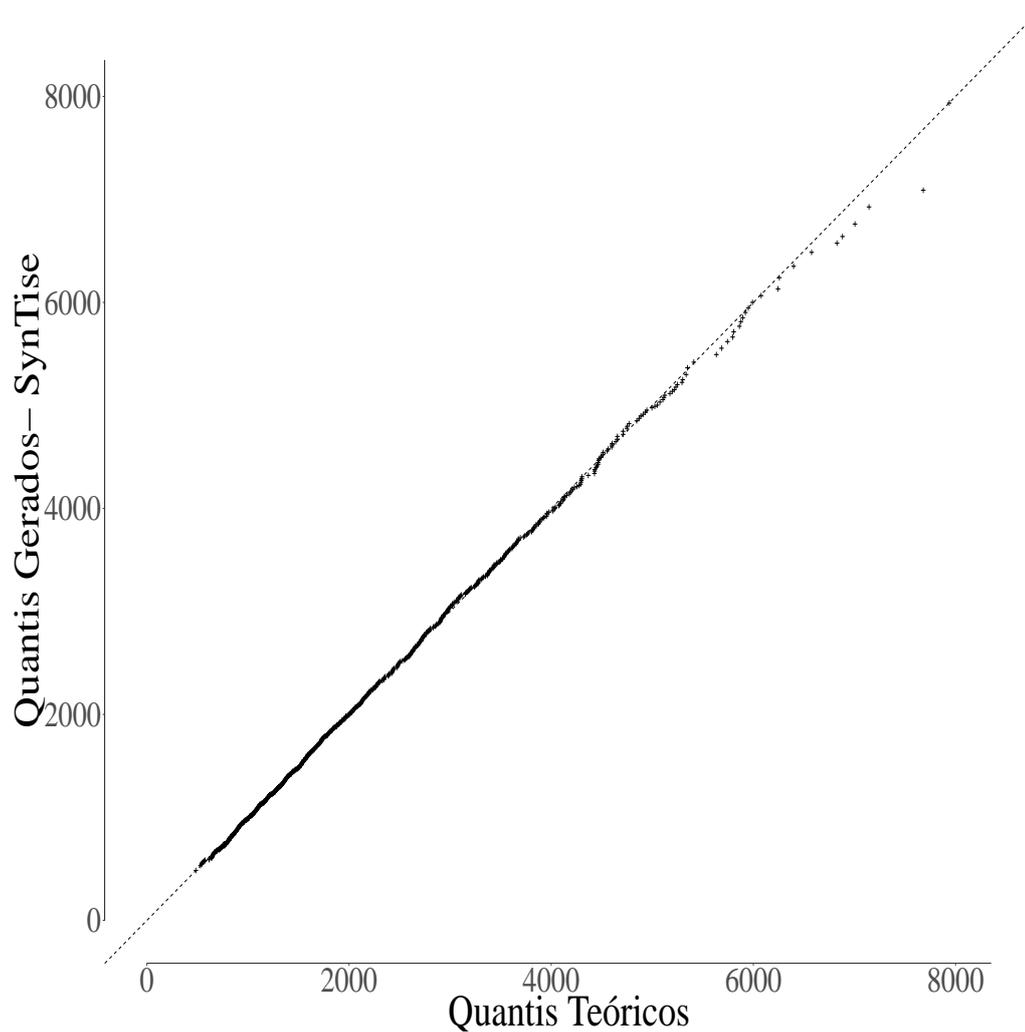
Figura 54 – Box-plots das vazões geradas pelo SynTise para usina de Água Vermelha



Legenda: a) Série Histórica
b) SynTise

Fonte: Elaborada pelo autor.

Figura 55 – Q-qplot das vazões geradas pelo SynTise.



Fonte: Elaborada pelo autor.

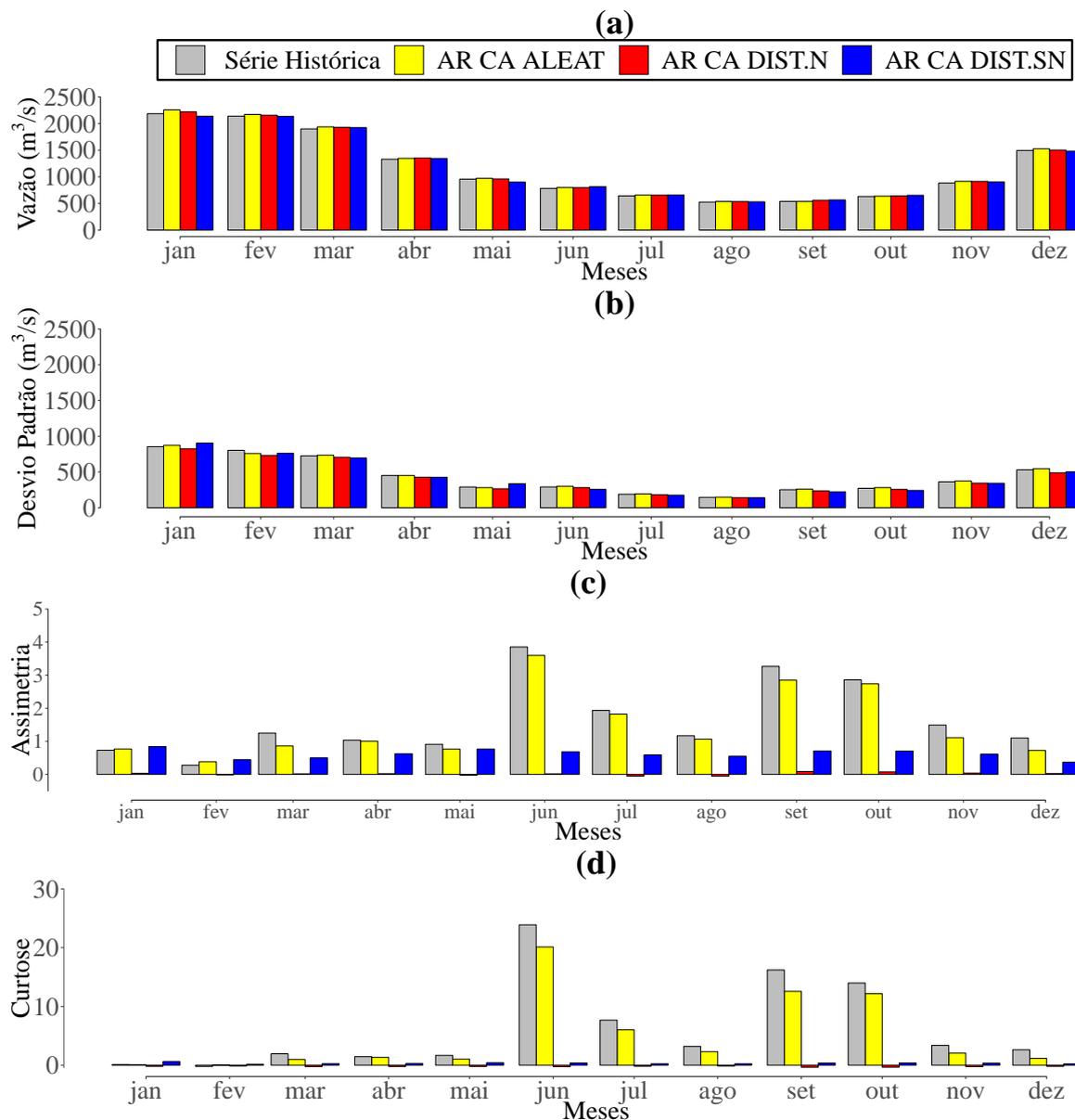
Portanto, através dessas análises, concluí-se que o SynTise foi bem ajustado aos dados de vazões mensais da usina de Água Vermelha, visto que conseguiu replicar bem as estatísticas descritivas e os *outliers* da série em todos os meses, com exceção dos *outliers* inferiores do mês de julho.

APÊNDICE F - MODELOS AR PARA USINA DE VOLTA GRANDE

Para o modelo AR, a ordem que foi considerado para a geração da série sintética para usina de Volta Grande foi o AR(4), ou seja, a predição do dados foi feita utilizando os últimos quatro anos da série histórica.

A princípio, foi executada a comparação entre os modelos com CA DIST.N e DIST.SN. Pelas Figuras 56a e 56b, nota-se que os três modelos conseguiram de certa forma replicar bem tanto a média, quanto o desvio padrão da série histórica do dados em todos os meses. Porém, analisando as Figuras 56c e 56d, assim como na usina Água Vermelha, vê-se que os modelos com CA DIST.N e DIST.SN não conseguiram replicar bem essas características descritivas dos dados, principalmente nos meses que possuíam valores maiores para assimetria e curtose, como junho, setembro e outubro. O modelo de CA ALEAT apresentou resultados mais expressivos do que na usina de Água Vermelha para esses meses, além de ser melhor que os modelos de CA DIST.N e DIST.SN para esta usina.

Figura 56 – Comparação entre as estatísticas descritivas das vazões mensais geradas pelos modelos AR de CA ALEAT, DIST.N e DIST.SN para usina de Volta Grande



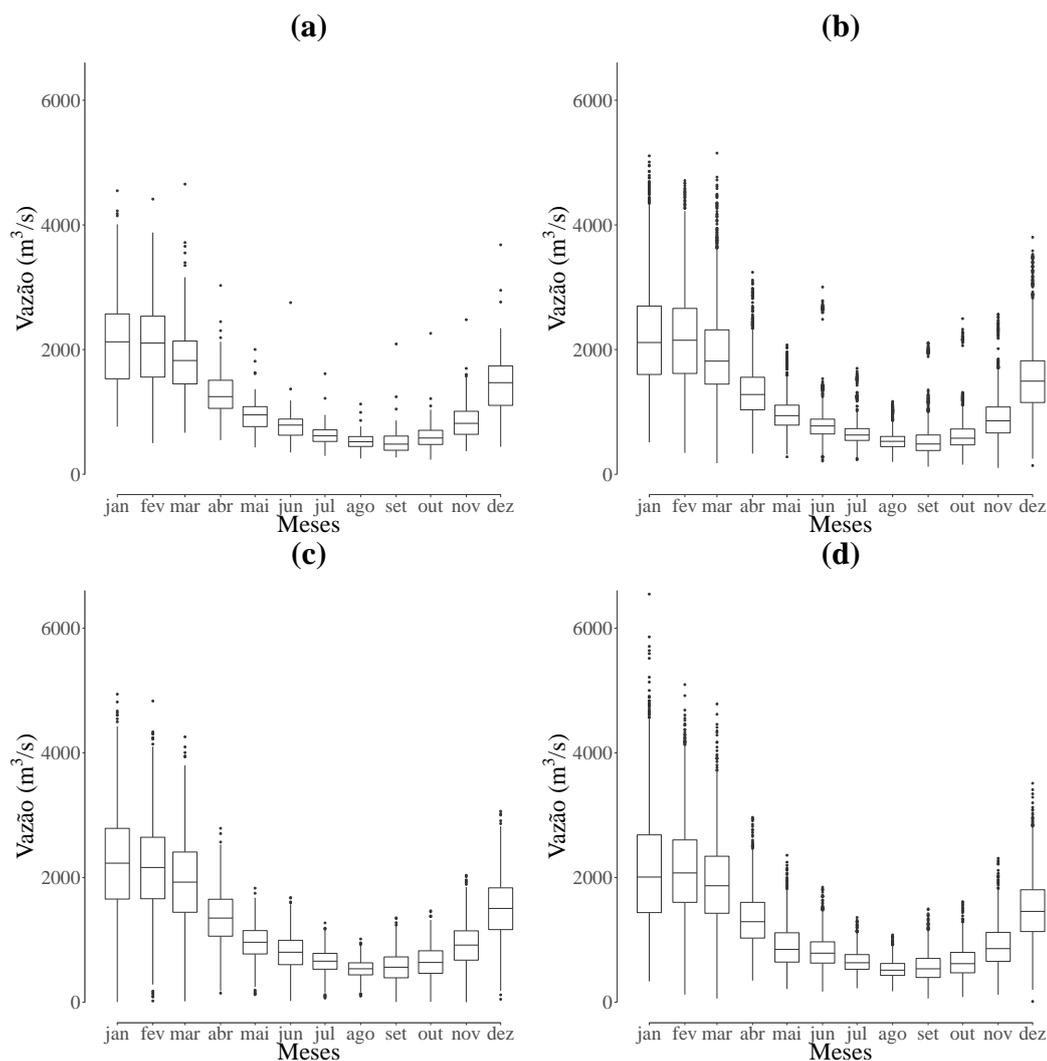
Legenda: a) Média
 b) Desvio Padrão
 c) Assimetria
 d) Curtose

Fonte: Elaborada pelo autor.

Pelos box-plots das vazões geradas pelos modelos mostrados nas Figuras 57b, 57c e 57d, também é exposta a dificuldade desses modelo quanto aos *outliers*, visto que mesmo com o modelo de CA ALEAT, que melhor representou a série histórica, existem *outliers* inferiores gerados para os meses de maio, junho, julho e dezembro, enquanto na série histórica da Figura 57a não apresentam nenhum. Os modelos com CA DIST.N gerou *outliers* inferiores para

fevereiro, maio, julho, agosto e dezembro, além de possuir amplitude bastante distinta da série histórica em alguns meses. Quanto ao modelo com CA DIST.SN, apenas no mês de dezembro gerou-se *outliers* inferiores. Quanto aos *outliers* superiores, todos os modelos apresentaram valores do tipo nos doze meses, assim como na série histórica, porém os modelo de CA ALEAT conseguiu representá-los de maneira mais fidedigna.

Figura 57 – Box-plots das vazões geradas pelos modelos AR para usina de Volta Grande



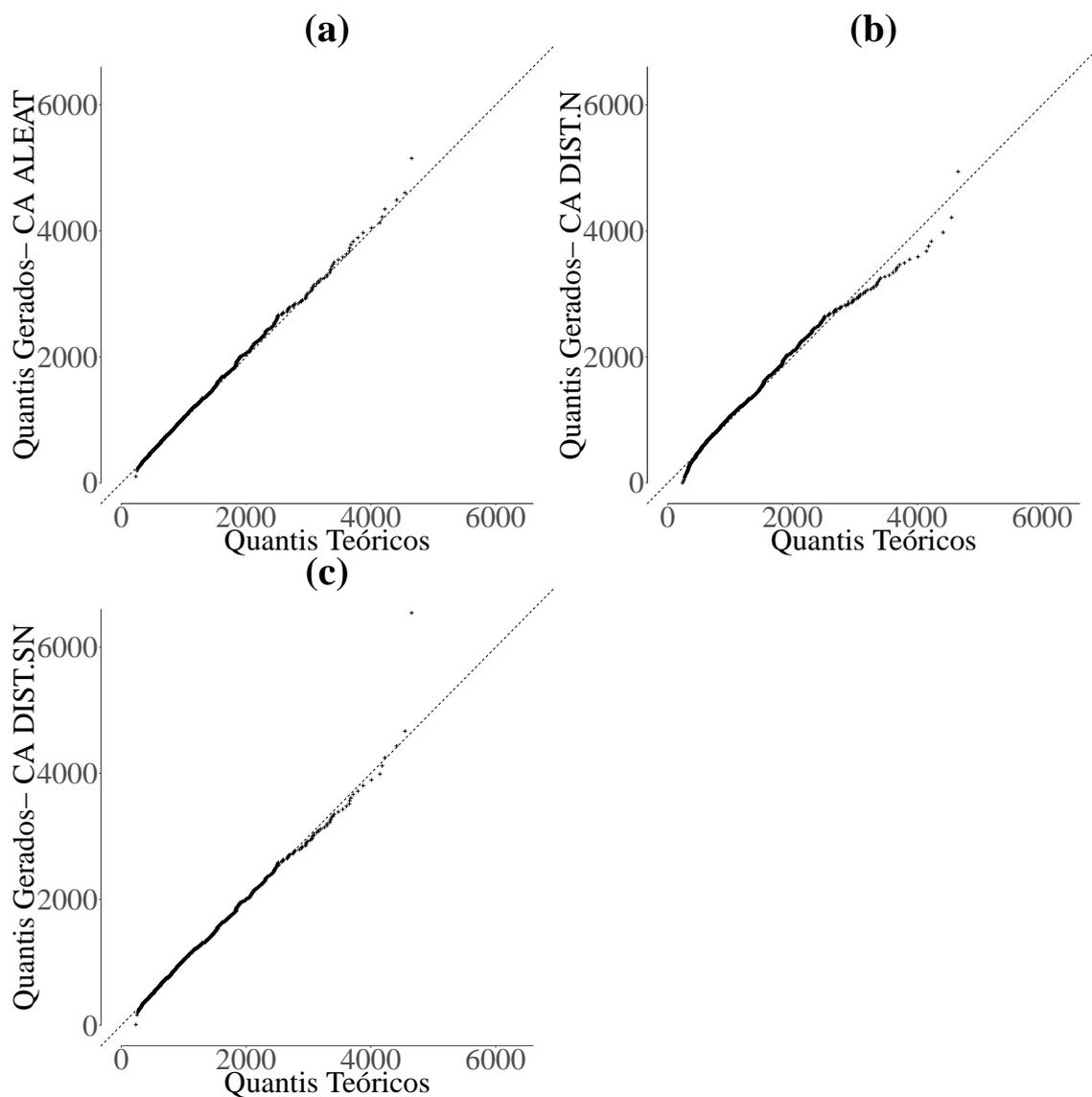
Legenda: a) Série Histórica
 b) CA ALEAT
 c) CA DIST.N
 d) CA DIST.SN

Fonte: Elaborada pelo autor.

Analisando os q-plots das Figuras 58a, 58b e 58c continua evidente a dificuldade que esses modelos apresentam para replicar os quantis mais elevados e os menores da série histórica, mesmo no modelo com CA ALEAT, que apresentou uma capacidade bem superior aos outros

dois quanto as estatísticas descritivas da série histórica.

Figura 58 – Q-plots das vazões geradas pelos modelos AR para usina de Volta Grande



Legenda: a) CA ALEAT
 b) CA DIST.N
 c) CA DIST.SN

Fonte: Elaborada pelo autor.

Comparando os erros da FAC e da FDP conforme a Tabela 12, pode-se concluir então que os modelos AR CA DIST.N e DIST.SN não são uma opção mais viável que o modelo com CA ALEAT, visto que como dito anteriormente, além de não conseguir replicar bem os *outliers* da série histórica, a curtose e assimetria presente, também apresentou um erro maior na FDP, apenas no mês de abril a CA DIST.SN obteve o melhor desempenho. Porém, na FAC, a CA DIST.N apresentou os menores valores.

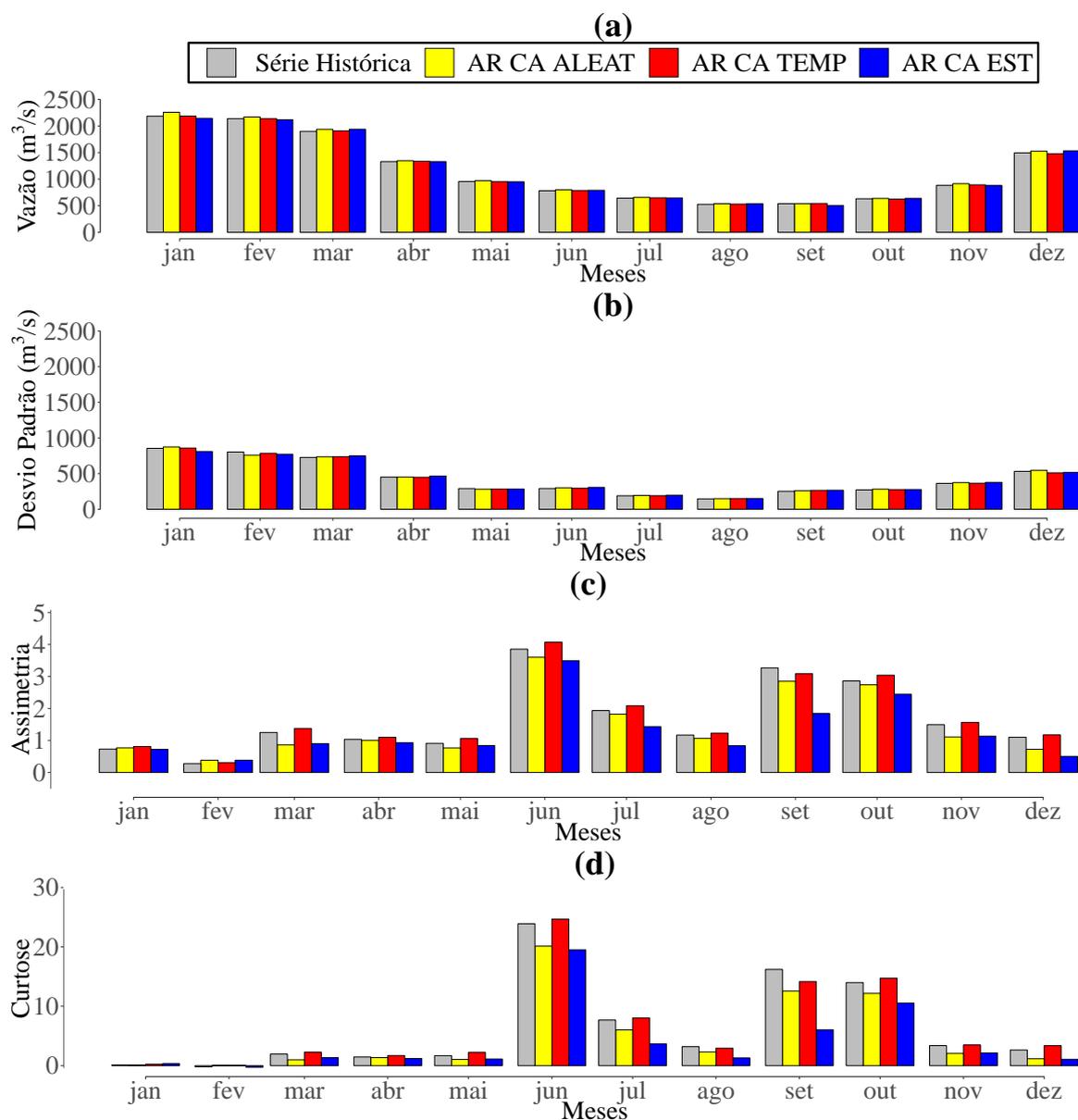
Tabela 12 – RMSE da FDP e FAC mensal dos modelos AR com CA ALEAT, DIST.N e DIST.SN

RMSE da FAC			
Mês	AR CA ALEAT	AR CA DIST.N	AR CA DIST.SN
janeiro	0,090	0,075	0,080
fevereiro	0,095	0,085	0,092
março	0,087	0,067	0,07
abril	0,091	0,070	0,082
maio	0,063	0,067	0,081
junho	0,066	0,053	0,072
julho	0,066	0,050	0,068
agosto	0,083	0,060	0,079
setembro	0,097	0,076	0,069
outubro	0,059	0,047	0,057
novembro	0,049	0,047	0,038
dezembro	0,080	0,068	0,080
Média dos meses	0,077	0,064	0,072
RMSE da FDP			
Mês	AR CA ALEAT	AR CA DIST.N	AR CA DIST.SN
janeiro	0,016	0,039	0,031
fevereiro	0,022	0,030	0,026
março	0,023	0,061	0,045
abril	0,024	0,034	0,016
maio	0,019	0,030	0,050
junho	0,029	0,059	0,039
julho	0,032	0,043	0,032
agosto	0,023	0,042	0,027
setembro	0,005	0,079	0,048
outubro	0,009	0,073	0,057
novembro	0,019	0,059	0,035
dezembro	0,023	0,033	0,033
Média dos meses	0,020	0,048	0,037

Fonte: Elaborada pelo autor.

Numa análise inicial das outras duas propostas do trabalho, através das estatísticas descritivas as Figuras 59a e 59b mostram que os três modelos conseguiram de reproduzir bem tanto a média quanto desvio padrão da série histórica dos dados em todos os meses. Já na assimetria, os modelos de CA TEMP e ALEAT reproduziram a série de forma mais efetiva, enquanto na curtose, o de CA TEMP se mostrou mais efetivo (ver Figuras 59c e 59d).

Figura 59 – Comparação entre as estatísticas descritivas das vazões mensais geradas pelos modelos AR de CA ALEAT, TEMP e EST para usina de Volta Grande



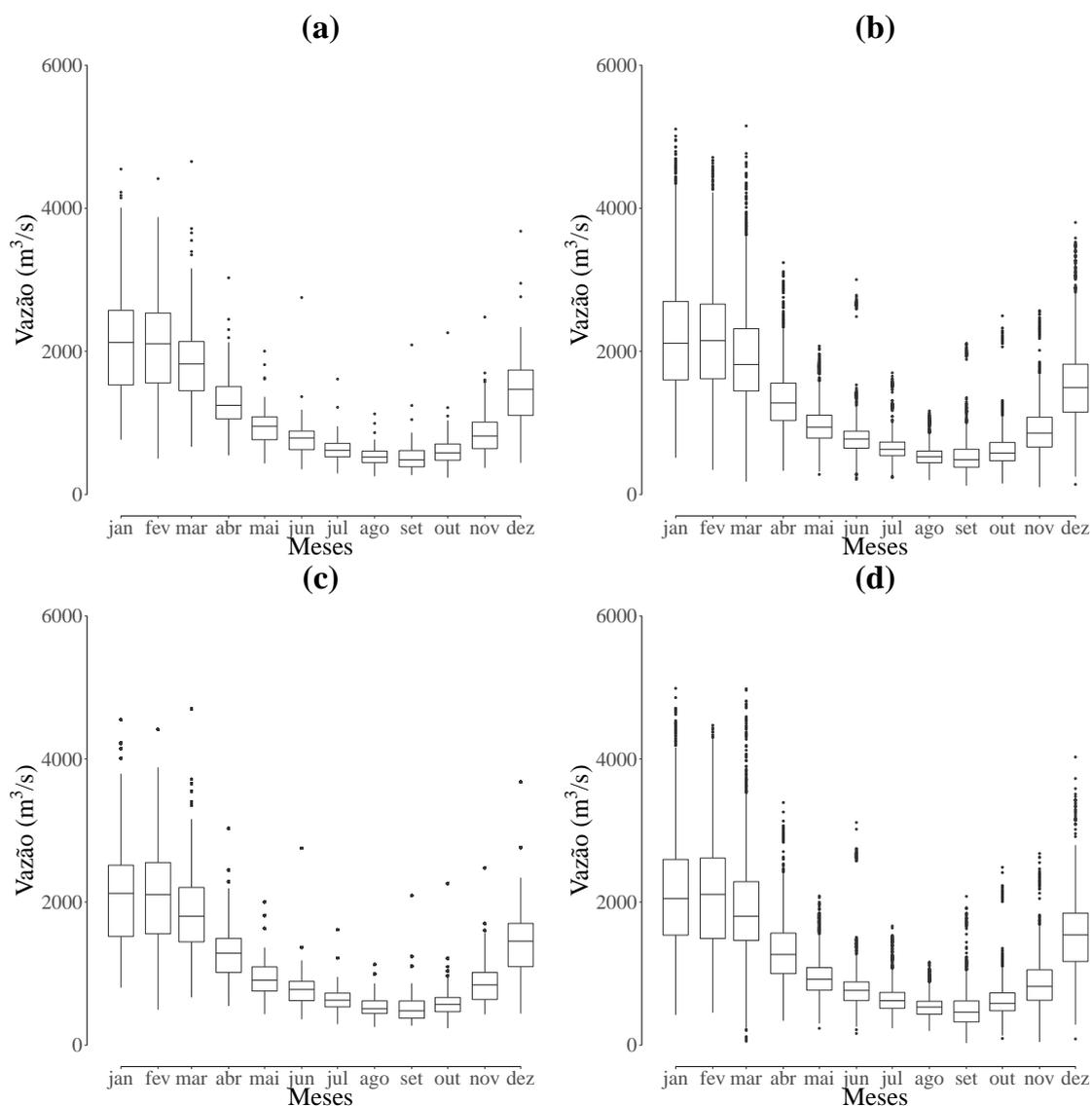
Legenda: a) Média
 b) Desvio Padrão
 c) Assimetria
 d) Curtose

Fonte: Elaborada pelo autor.

Através dos box-plots das vazões geradas pelos modelos exibidos nas Figuras 60c e 60d, pode-se notar que o modelo com CA TEMP não gera nenhum *outlier* inferior, assim como na série histórica. Já o modelo com CA EST gerou *outliers* superiores com valores próximos aos presentes na série histórica, porém (ver Figura 60b), ele gerou *outliers* inferiores em meses onde isso não ocorre na série histórica como março, maio, junho, julho, outubro e dezembro. Já o

modelo com CA TEMP cometeu esse equívoco apenas no mês de fevereiro.

Figura 60 – Box-plots das vazões geradas pelos modelos AR para usina de Volta Grande

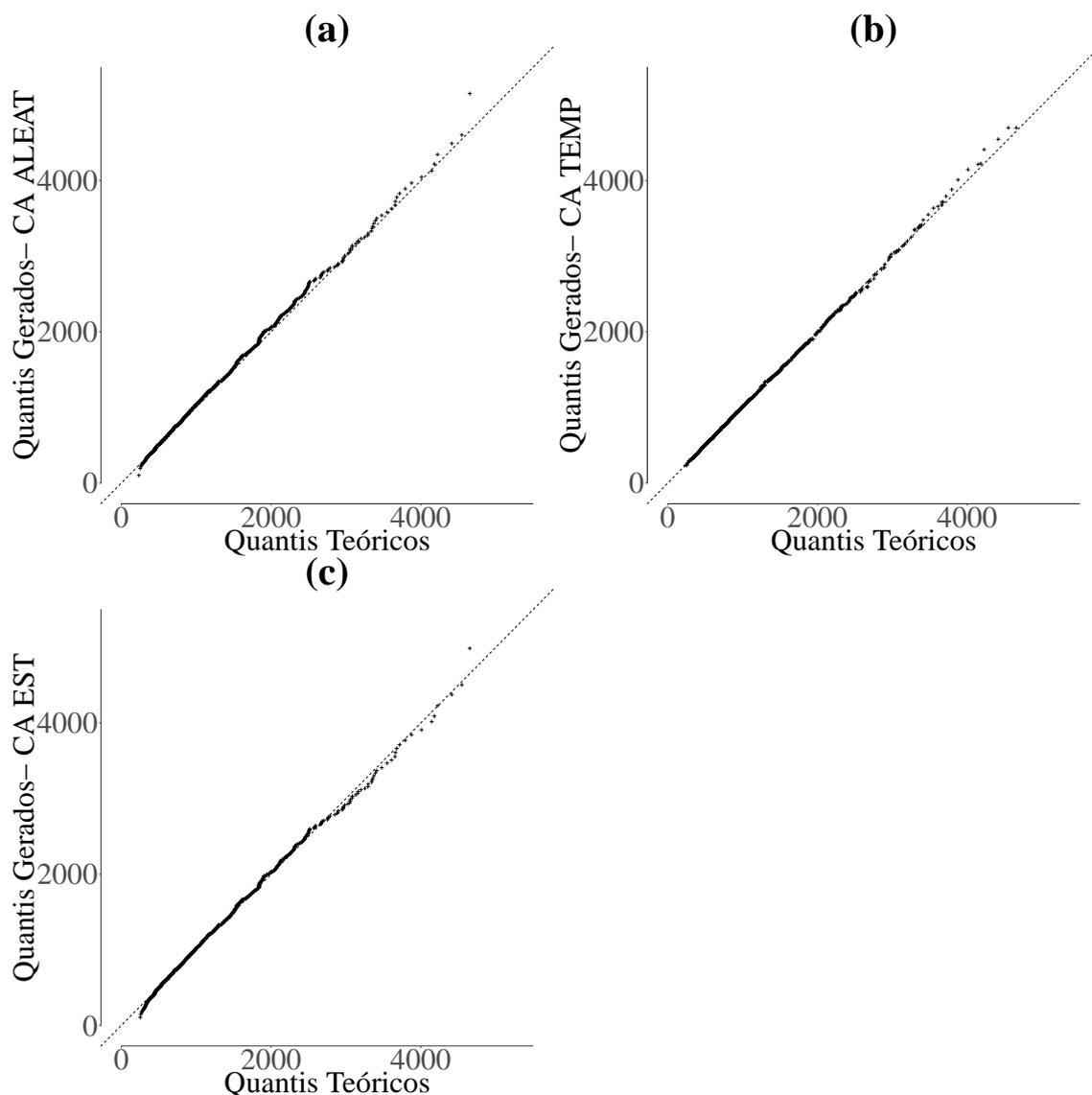


Legenda: a) Série Histórica
 b) CA ALEAT
 c) CA TEMP
 d) CA EST

Fonte: Elaborada pelo autor.

Os q-qplots apresentados nas Figuras 61b e 61c ressaltam que ambas as propostas de modelagem para o CA se ajustaram bem aos dados, mostrando-se superior ao modelo de CA ALEAT, principalmente quando se compara com o q-qplot da Figura 61a.

Figura 61 – Q-plots das vazões geradas pelos modelos AR para usina de Volta Grande



Legenda: a) CA ALEAT

b) CA TEMP

c) CA EST

Fonte: Elaborada pelo autor.

Por fim, explorando os RMSE da FAC e FDP desses três modelos (ver Tabela 13), constata-se que o modelo de CA EST apresentou um erro maior da FAC dentre os três modelos nos meses de maio, agosto, novembro e dezembro, enquanto o modelo de CA TEMP obteve um erro menor em todos os meses. Quanto a FDP, o modelo com CA TEMP também obteve os menores erros em praticamente todos os meses, com exceção do mês de outubro e novembro no qual o de CA ALEAT foi melhor ou igual. Destaca-se o fato de nos meses de janeiro, junho e setembro, o erro da FDP do modelo de CA ao longo do tempo foi tão baixo que atingiu a terceira casa decimal.

Tabela 13 – RMSE da FAC e FDP mensal dos modelos AR de CA ALEAT, TEMP e EST

RMSE da FAC			
Mês	AR CA ALEAT	AR CA TEMP	AR CA EST
janeiro	0,090	0,045	0,086
fevereiro	0,095	0,070	0,083
março	0,087	0,048	0,074
abril	0,091	0,052	0,072
maio	0,063	0,048	0,070
junho	0,066	0,022	0,063
julho	0,066	0,029	0,047
agosto	0,083	0,039	0,063
setembro	0,097	0,032	0,068
outubro	0,059	0,025	0,049
novembro	0,049	0,027	0,050
dezembro	0,080	0,044	0,085
Média dos meses	0,077	0,040	0,067
RMSE da FDP			
Mês	AR CA ALEAT	AR CA TEMP	AR CA EST
janeiro	0,016	0,007	0,030
fevereiro	0,022	0,022	0,030
março	0,023	0,012	0,024
abril	0,024	0,018	0,031
maio	0,019	0,017	0,028
junho	0,029	0,007	0,035
julho	0,032	0,010	0,016
agosto	0,023	0,015	0,028
setembro	0,005	0,004	0,074
outubro	0,009	0,012	0,012
novembro	0,019	0,019	0,023
dezembro	0,023	0,014	0,028
Média dos meses	0,020	0,013	0,030

Fonte: Elaborada pelo autor.

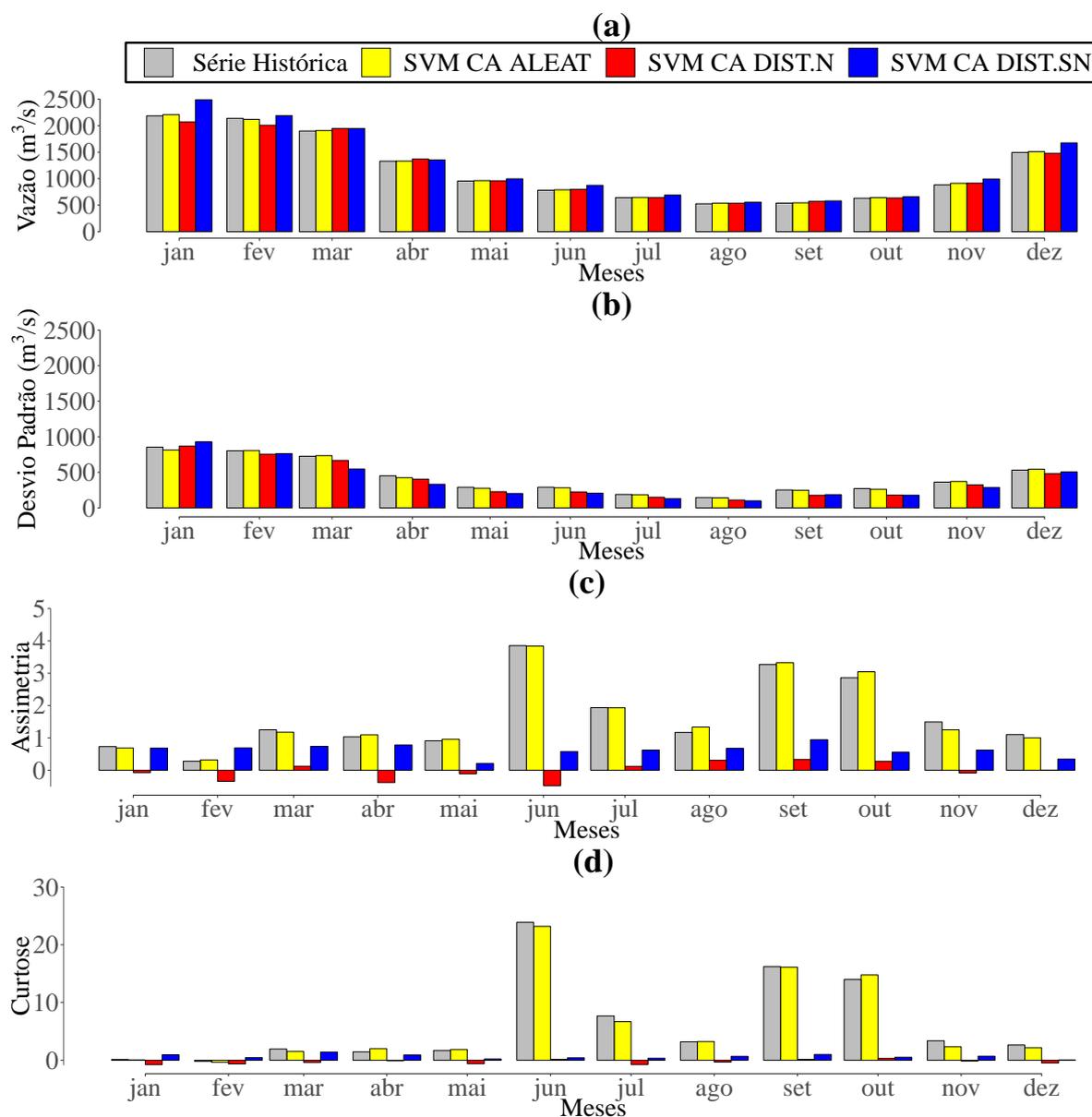
Dada todas essas análises feitas sobre as propostas de modelos apresentadas, concluí-se que o modelo AR CA TEMP aparenta ser a alternativa mais adequada dentre todos os modelos apresentados nesta seção.

APÊNDICE G - MODELOS SVM PARA USINA DE VOLTA GRANDE

Os modelos SVM gerados utilizaram seis passos anteriores, ou seja, a predição foi feita utilizando dados dos últimos seis anos da série histórica para cada mês.

A primeira das comparações executadas foi entre os modelos com CA DIST.N e DIST.SN. Pela Figuras 62a e 62b, nota-se que os três modelos conseguiram de certa forma replicar de forma razoável tanto a média, quanto o desvio padrão da série histórica do dados em todos os meses, com exceção de janeiro, fevereiro e dezembro, onde os modelos que utilizam CA DIST.N e DIST.SN acabaram obtendo resultados consideravelmente menores que o modelo de CA ALEAT. Porém, analisando as Figuras 62c e 62d, vê-se que os modelos com CA DIST.N e DIST.SN não conseguiram replicar bem tanto a curtose quanto assimetria da série em praticamente todos os meses, sendo que em alguns casos, até assimetrias negativas em meses com alta assimetria positiva acabaram por serem gerados pelo modelo de CA DIST.N.

Figura 62 – Comparação entre as estatísticas descritivas das vazões mensais geradas pelos modelos SVM de CA ALEAT, DIST.N e DIST.SN para usina de Volta Grande



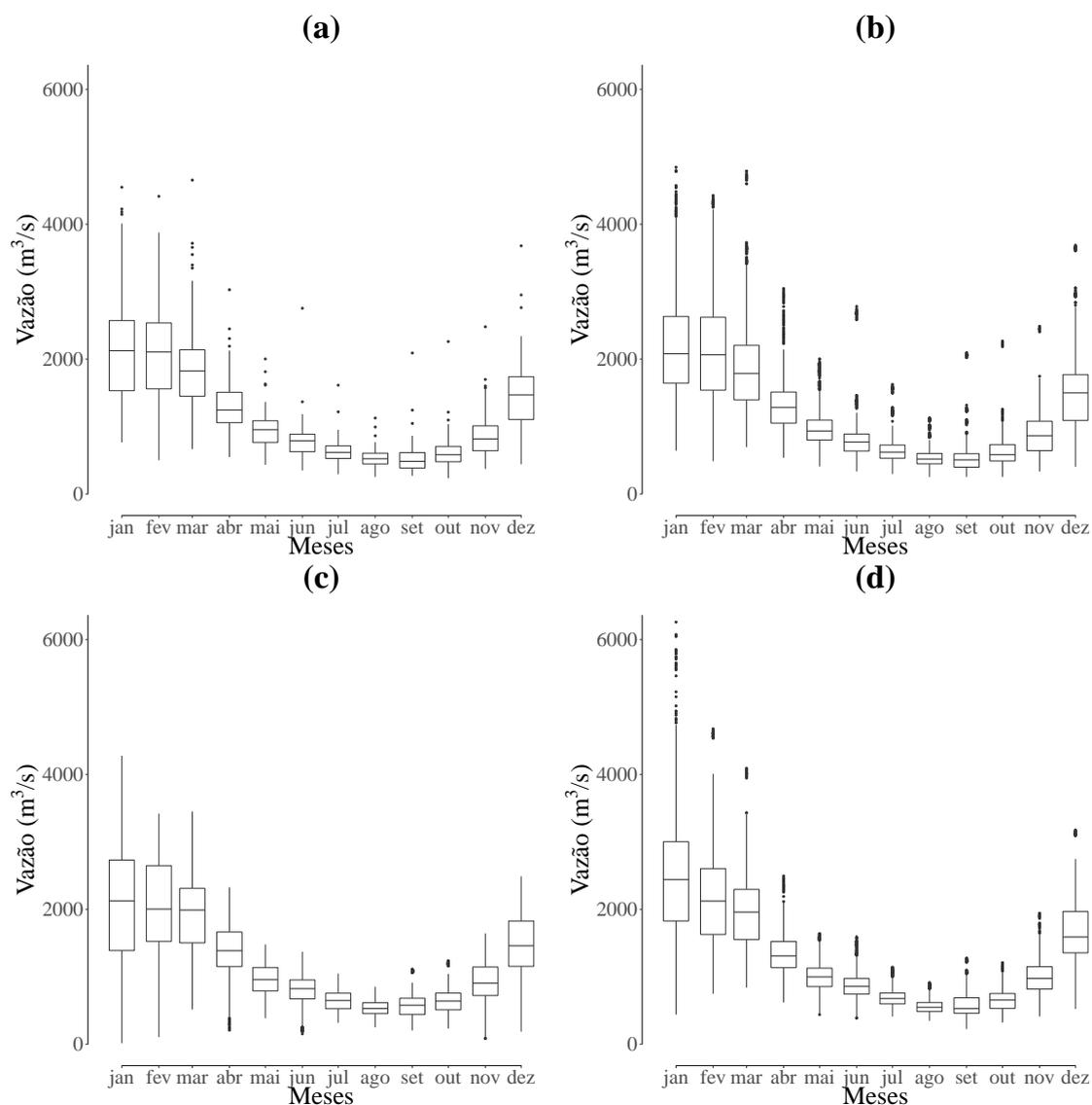
Legenda: a) Média
 b) Desvio Padrão
 c) Assimetria
 d) Curtose

Fonte: Elaborada pelo autor.

Nos box-plots das vazões geradas apresentados nas Figuras 63b, 63c e 63d, é exposta uma dificuldade quanto aos *outliers*, porém, menos do que na usina anterior. O modelo com CA ALEAT, que conseguiu representar melhor a série histórica, também reproduziu os *outliers* inferiores, visto que não foram observados nenhum, tanto na série quanto no modelo. O modelo de CA DIST.N apresentou *outliers* inferiores no mês de abril, junho e novembro. Enquanto que

o modelo CA DIST.SN, assim como o de CA ALEAT, não apresentou nenhum.

Figura 63 – Box-plots das vazões geradas pelos modelos SVM para usina de Volta Grande

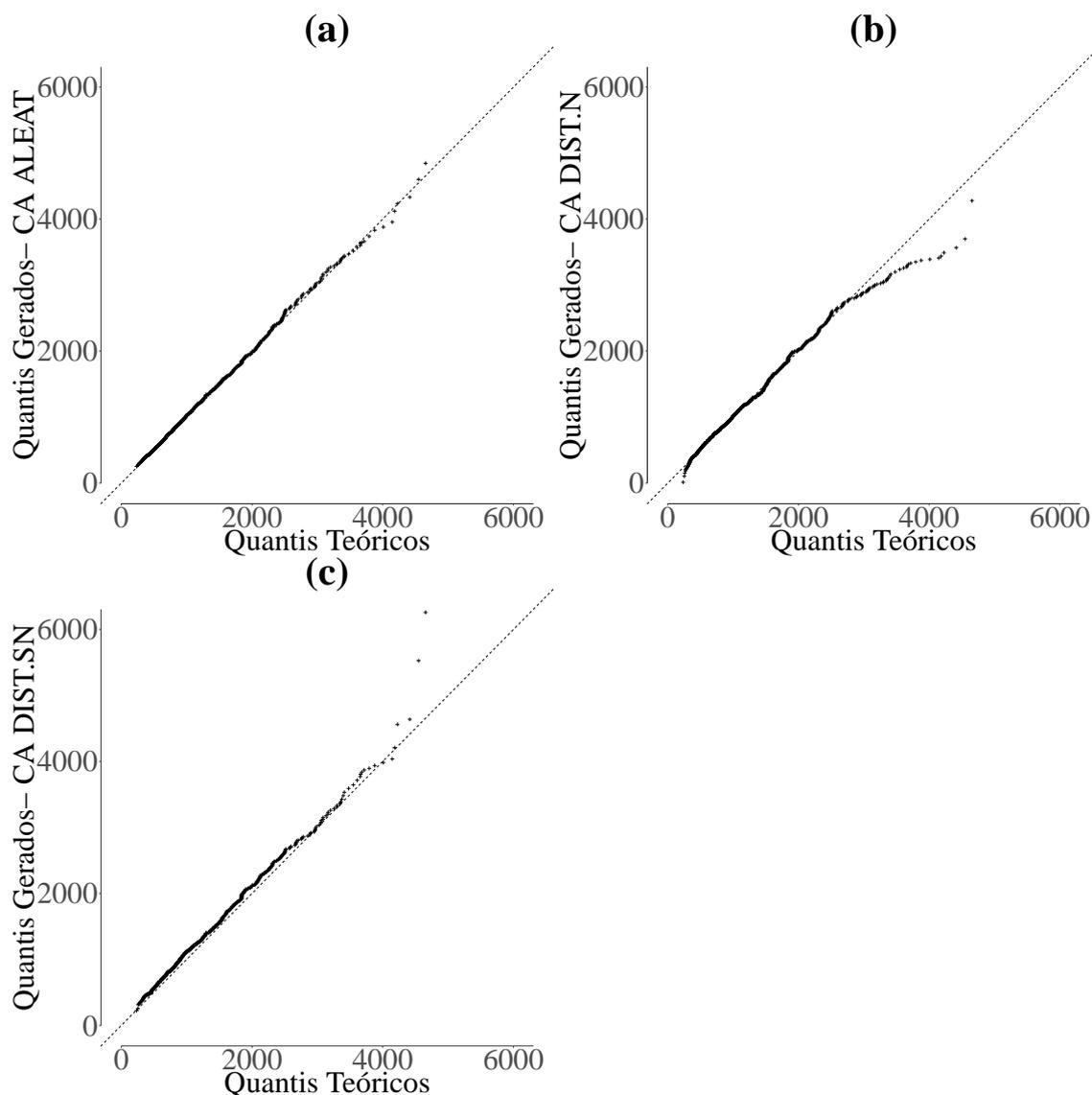


Legenda: a) Série Histórica
 b) CA ALEAT
 c) CA DIST.N
 d) CA DIST.SN

Fonte: Elaborada pelo autor.

Através dos q-qplots apresentados nas Figuras 64b e 64c fica evidente ainda que ambas as propostas de modelagem para os CA não se ajustaram bem aos dados, em especial nos valores extremos da série, onde mostra grande dificuldade em reproduzir esses valores quando comparado ao q-qplot da Figura 64a.

Figura 64 – Q-plots das vazões geradas pelos modelos SVM para usina de Volta Grande



Legenda: a) CA ALEAT

b) CA DIST.N

c) CA DIST.SN

Fonte: Elaborada pelo autor.

Comparando os erros da FAC e FDP conforme a Tabela 14, entende-se que os modelos SVM CA ALEAT, DIST.N e DIST.SN apresentaram resultados próximos entre si em relação a FAC, porém, no que se refere a FDP, o modelo com CA ALEAT obteve erros bem menores em relação aos outros.

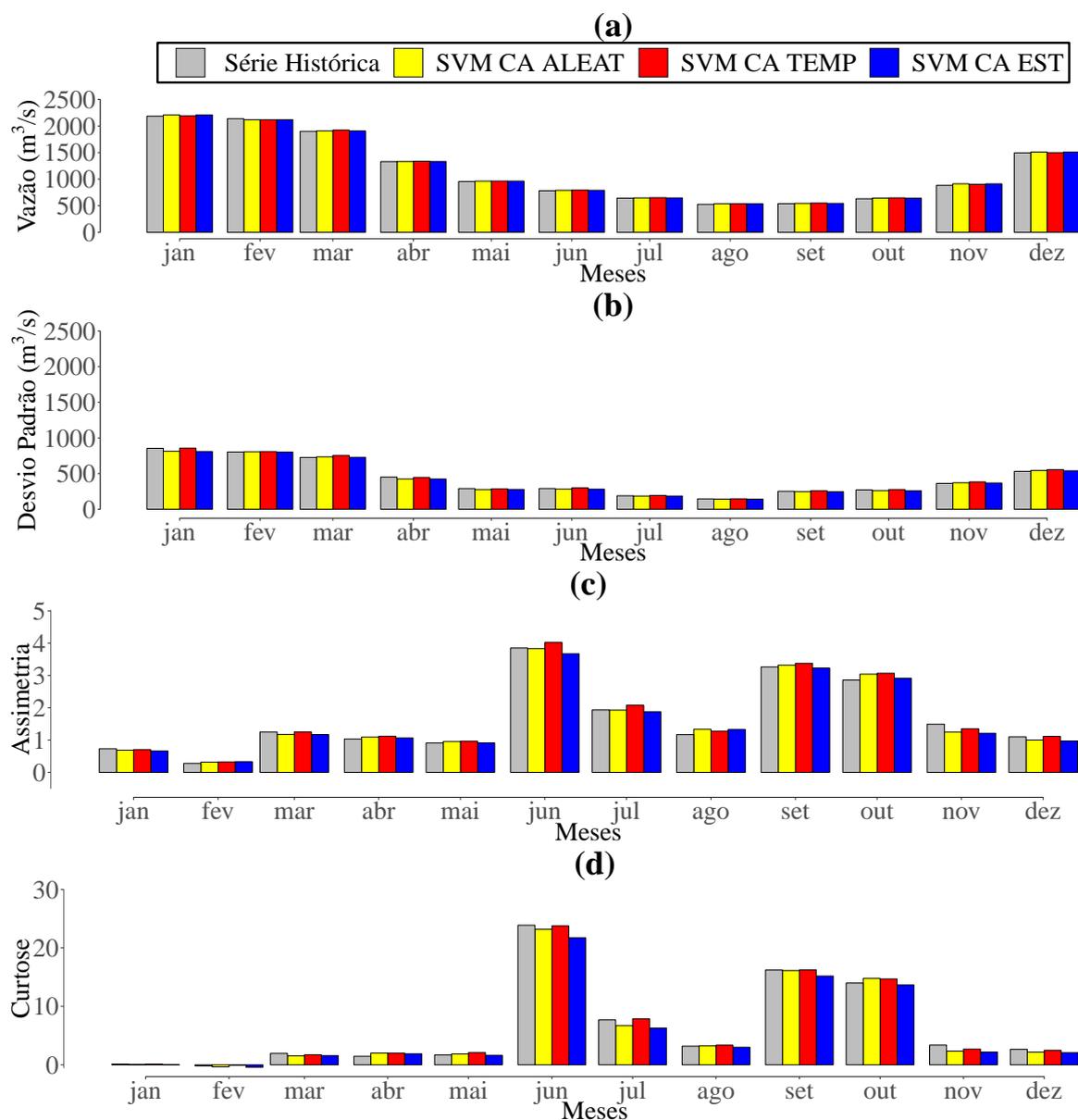
Tabela 14 – RMSE da FAC e FDP mensal dos modelos SVM de CA ALEAT, DIST.N e DIST.SN

RMSE da FAC			
Mês	SVM CA ALEAT	SVM CA DIST.N	SVM CA DIST.SN
janeiro	0,103	0,101	0,103
fevereiro	0,103	0,116	0,114
março	0,113	0,117	0,112
abril	0,081	0,083	0,092
maio	0,092	0,103	0,095
junho	0,078	0,074	0,078
julho	0,078	0,079	0,083
agosto	0,107	0,117	0,104
setembro	0,106	0,099	0,096
outubro	0,069	0,070	0,070
novembro	0,077	0,085	0,074
dezembro	0,116	0,121	0,112
Média dos meses	0,094	0,097	0,094
RMSE da FDP			
Mês	SVM CA ALEAT	SVM CA DIST.N	SVM CA DIST.SN
janeiro	0,016	0,041	0,070
fevereiro	0,018	0,043	0,030
março	0,024	0,060	0,046
abril	0,022	0,068	0,042
maio	0,019	0,034	0,047
junho	0,009	0,047	0,066
julho	0,024	0,042	0,052
agosto	0,017	0,024	0,051
setembro	0,024	0,069	0,056
outubro	0,016	0,042	0,073
novembro	0,012	0,061	0,074
dezembro	0,017	0,056	0,058
Média dos meses	0,018	0,049	0,055

Fonte: Elaborada pelo autor.

Analisando as estatísticas descritivas das outras duas propostas apresentadas nesse trabalho para essa técnica, as Figuras 65a e 65b mostram que os três modelos conseguiram reproduzir bem tanto a média quanto desvio padrão da série histórica dos dados em todos os meses. O mesmo efeito pode ser visto na assimetria e curtose, onde os três modelos replicaram essas estatísticas de maneira razoável em todos os meses, principalmente os de CA ALEAT e TEMP (ver Figuras 65c e 65d).

Figura 65 – Comparação entre as estatísticas descritivas das vazões mensais geradas pelos modelos SVM de CA ALEAT, TEMP e EST para usina de Volta Grande



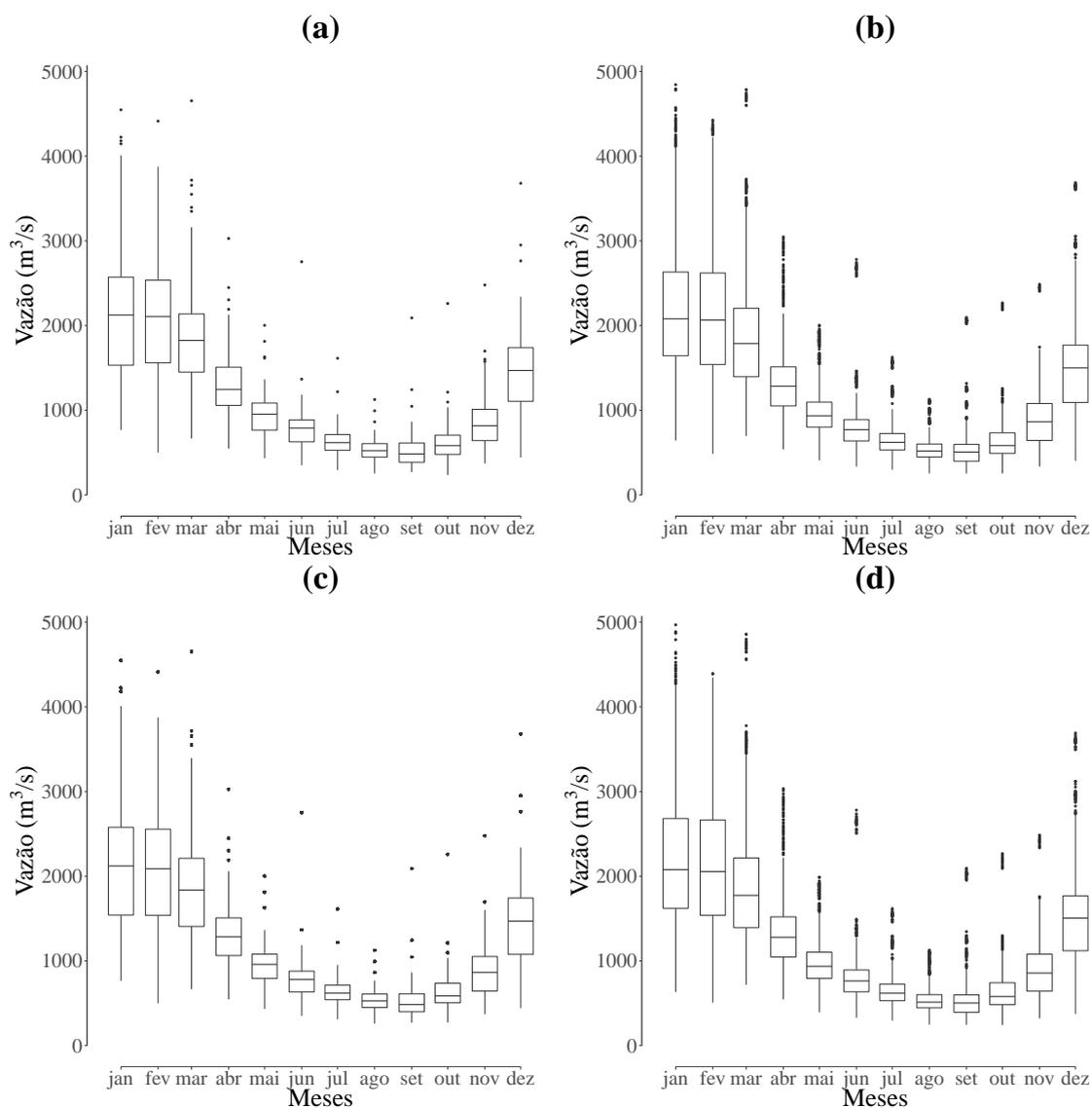
Legenda: a) Média
 b) Desvio Padrão
 c) Assimetria
 d) Curtose

Fonte: Elaborada pelo autor.

Através do box-plot dos modelos exibidos nas Figuras 66c e 66d, pode-se notar que o modelo com CA TEMP gera proporcionalmente poucos *outliers*, porém os valores gerados por ele são bem próximos aos presentes na série histórica, inclusive com nenhum *outlier* inferior, assim como na série histórica. Já o modelo com CA EST gerou *outliers* superiores com valores próximos aos presentes na série histórica (ver Figura 66b) e também produziu *outliers* inferiores

em meses onde isso não ocorre na série histórica.

Figura 66 – Box-plots das vazões geradas pelos modelos SVM para usina de Volta Grande

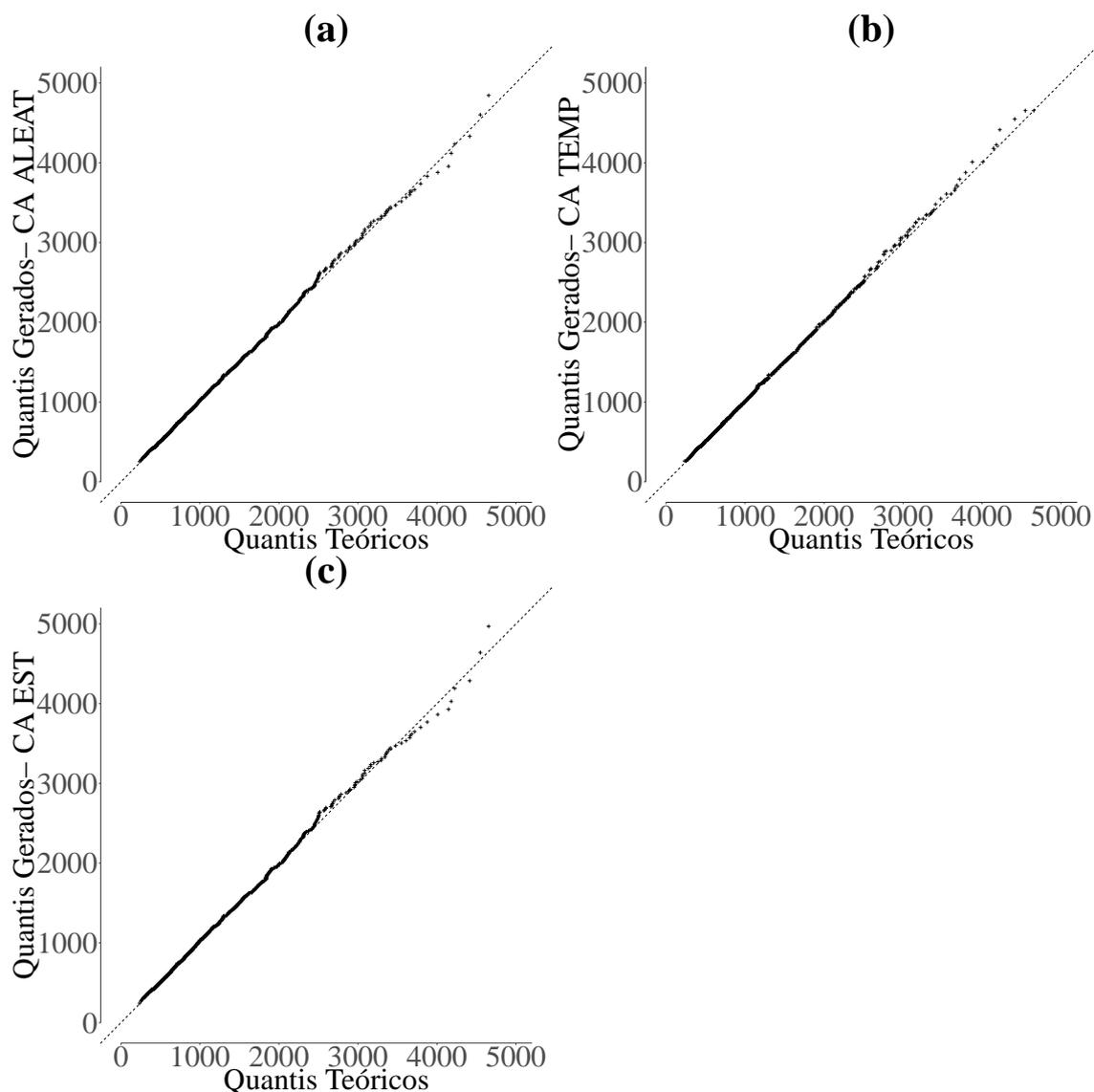


Legenda: a) Série Histórica
 b) CA ALEAT
 c) CA TEMP
 d) CA EST

Fonte: Elaborada pelo autor.

Através dos q-qplots apresentados nas Figuras 67b e 67c fica evidente que ambas as propostas de modelagem para os CA se ajustaram tão bem aos dados quanto o modelo de CA ALEAT (ver Figura 67a).

Figura 67 – Q-plots das vazões geradas pelos modelos SVM para usina de Volta Grande



Legenda: a) CA ALEAT

b) CA TEMP

c) CA EST

Fonte: Elaborada pelo autor.

Por fim, explorando os RMSE da FAC e FDP desses três modelos (ver Tabela 15), constata-se que o modelo de CA TEMP apresentou um erro menor da FAC em todos os meses. Quanto a FDP, o modelo de CA TEMP conseguiu diminuir o erro na maioria dos meses quando comparado aos demais, exceção feita aos meses de junho, outubro e novembro, onde o modelo de CA EST teve um leve vantagem.

Tabela 15 – RMSE da FAC e FDP mensal dos modelos SVM de CA ALEAT, TEMP e EST

RMSE da FAC			
Mês	SVM CA ALEAT	SVM CA TEMP	SVM CA EST
janeiro	0,103	0,063	0,103
fevereiro	0,103	0,062	0,103
março	0,113	0,026	0,112
abril	0,081	0,042	0,081
maio	0,092	0,032	0,091
junho	0,078	0,028	0,077
julho	0,078	0,029	0,077
agosto	0,107	0,042	0,106
setembro	0,106	0,031	0,106
outubro	0,069	0,039	0,067
novembro	0,077	0,030	0,075
dezembro	0,116	0,031	0,113
Média dos meses	0,094	0,038	0,093
RMSE da FDP			
Mês	SVM CA ALEAT	SVM CA TEMP	SVM CA EST
janeiro	0,016	0,014	0,016
fevereiro	0,018	0,011	0,018
março	0,024	0,009	0,026
abril	0,022	0,013	0,024
maio	0,019	0,012	0,021
junho	0,009	0,012	0,017
julho	0,024	0,010	0,022
agosto	0,017	0,012	0,023
setembro	0,024	0,013	0,022
outubro	0,016	0,022	0,021
novembro	0,012	0,019	0,016
dezembro	0,017	0,012	0,012
Média dos meses	0,018	0,013	0,020

Fonte: Elaborada pelo autor.

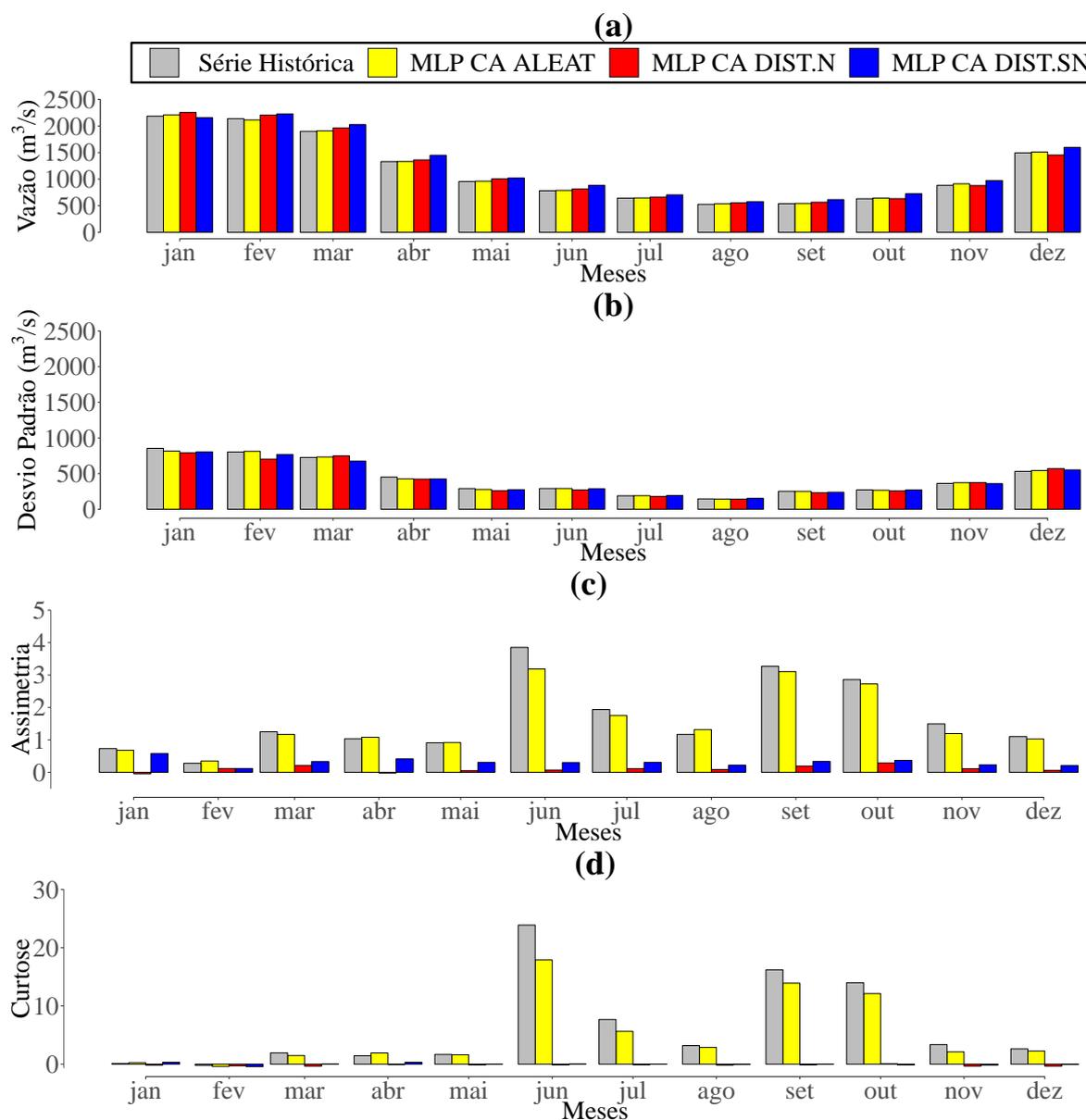
Portanto, concluí-se que o modelo SVM CA TEMP se apresenta como uma alternativa mais adequada dentre todos os modelos apresentados nesta seção.

APÊNDICE H - MODELOS MLP PARA USINA DE VOLTA GRANDE

Os modelos MLPs gerados para a usina de Volta Grande utilizaram seis passos anteriores, ou seja, a predição foi feita utilizando dados dos seis anos últimos da série histórica nos meses em que são avaliados.

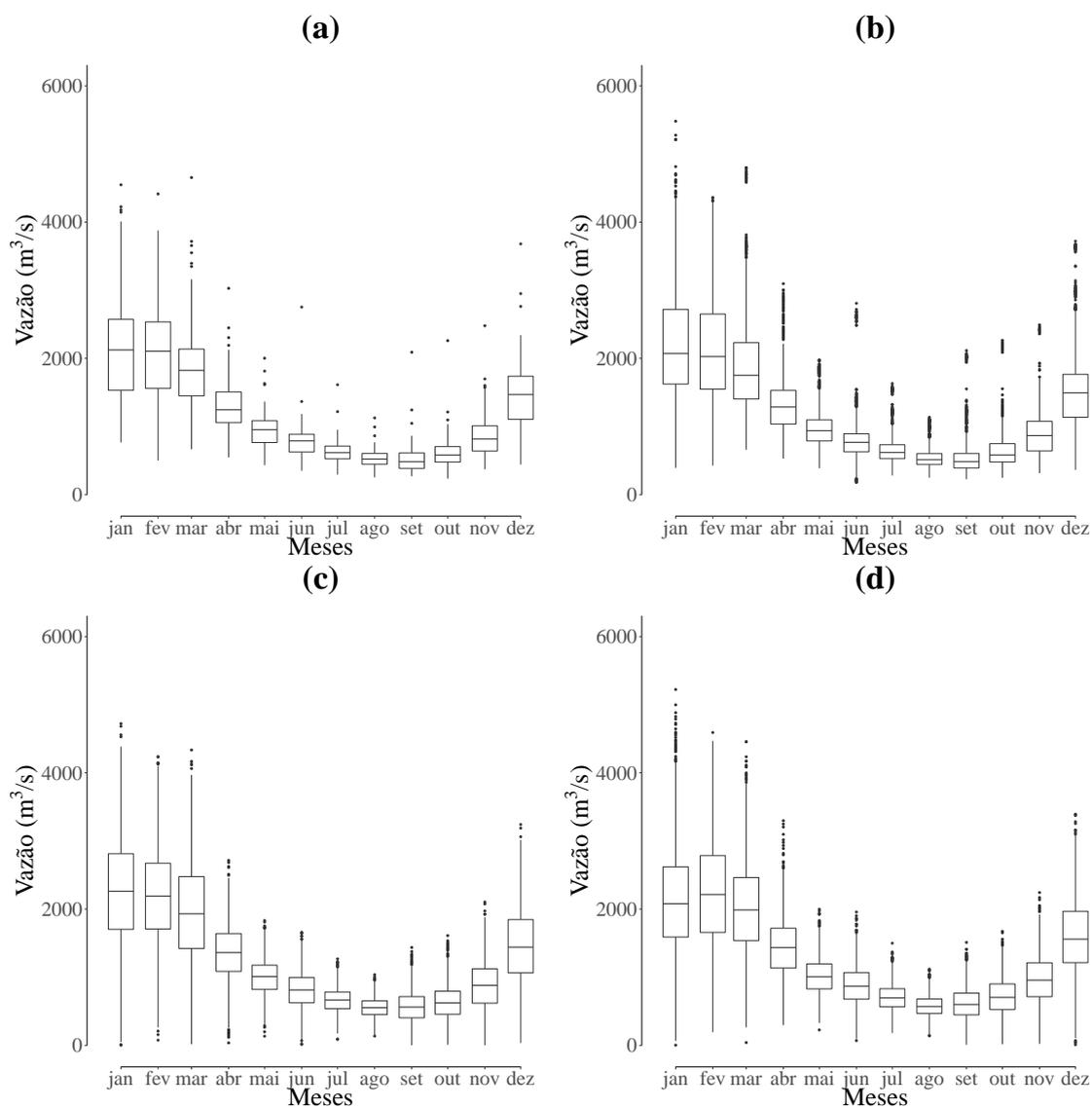
A primeira das comparações executadas foi entre os modelos com CA DIST.N e DIST.SN. Pelas Figuras 68a e 68b, percebe-se que os três modelos conseguiram de certa forma replicar de forma eficaz tanto a média, quanto desvio padrão da série histórica em todos os meses, com exceção do modelo de CA DIST.N que ficou aquém dos demais no mês de fevereiro. Porém, analisando as Figuras 68c e 68d, vê-se que os modelos com CA DIST.N e DIST.SN não conseguiram replicar bem tanto a curtose quanto assimetria em praticamente todos os meses, visto que esses modelos só conseguiam replicar de maneira razoável essas estatísticas quando o valores da série histórica eram baixos, como por exemplo nos meses de janeiro e fevereiro.

Figura 68 – Comparação entre as estatísticas descritivas das vazões mensais geradas pelos modelos MLP de CA ALEAT, DIST.N e DIST.SN para usina de Volta Grande



Nos box-plots das vazões geradas pelos modelos apresentados nas Figuras 69b, 69c e 69d, é exposta uma dificuldade quanto aos *outliers*. Todos os modelos geraram *outliers* inferiores, algo inexistente na série histórica. Além disso, os modelos de CA DIST.N e DIST.SN representaram mal a amplitude dos dados e, conseqüentemente, os *outliers* superiores também.

Figura 69 – Box-plots das vazões geradas pelos modelos MLP para usina de Volta Grande



Legenda: a) Série Histórica

b) CA ALEAT

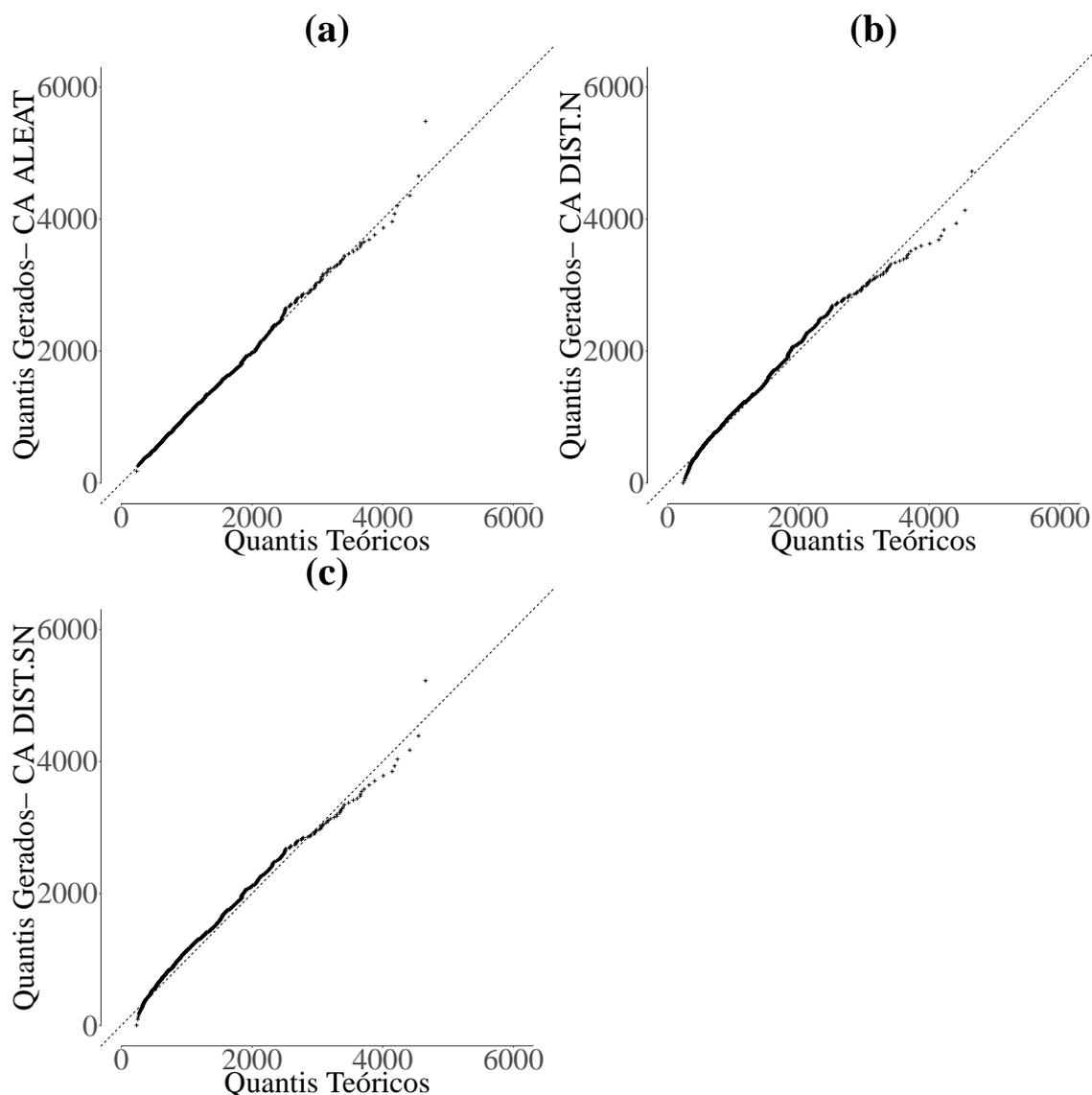
c) CA DIST.N

d) CA DIST.SN

Fonte: Elaborada pelo autor.

Através dos q-qplots das Figuras 70b e 70c fica evidente que ambas as propostas de modelagem para os CA não se ajustaram bem aos dados, em especial nos valores extremos positivos da série, onde mostraram uma dificuldade maior em reproduzir esses valores quando comparados ao q-qplot da Figura 70a.

Figura 70 – Q-qplots das vazões geradas pelos modelos MLP para usina de Volta Grande



Legenda: a) CA ALEAT

b) CA DIST.N

c) CA DIST.SN

Fonte: Elaborada pelo autor.

Comparando os erros da FAC e FDP conforme a Tabela 16, pode-se concluir que os modelos MLP de CA DIST.N e DIST.SN não são uma opção tão viável quanto o modelo com CA ALEAT, visto que, como dito anteriormente, além de não conseguirem replicar bem os *outliers*, a curtose e a assimetria em praticamente todos os meses, também apresentaram um maior erro na FDP.

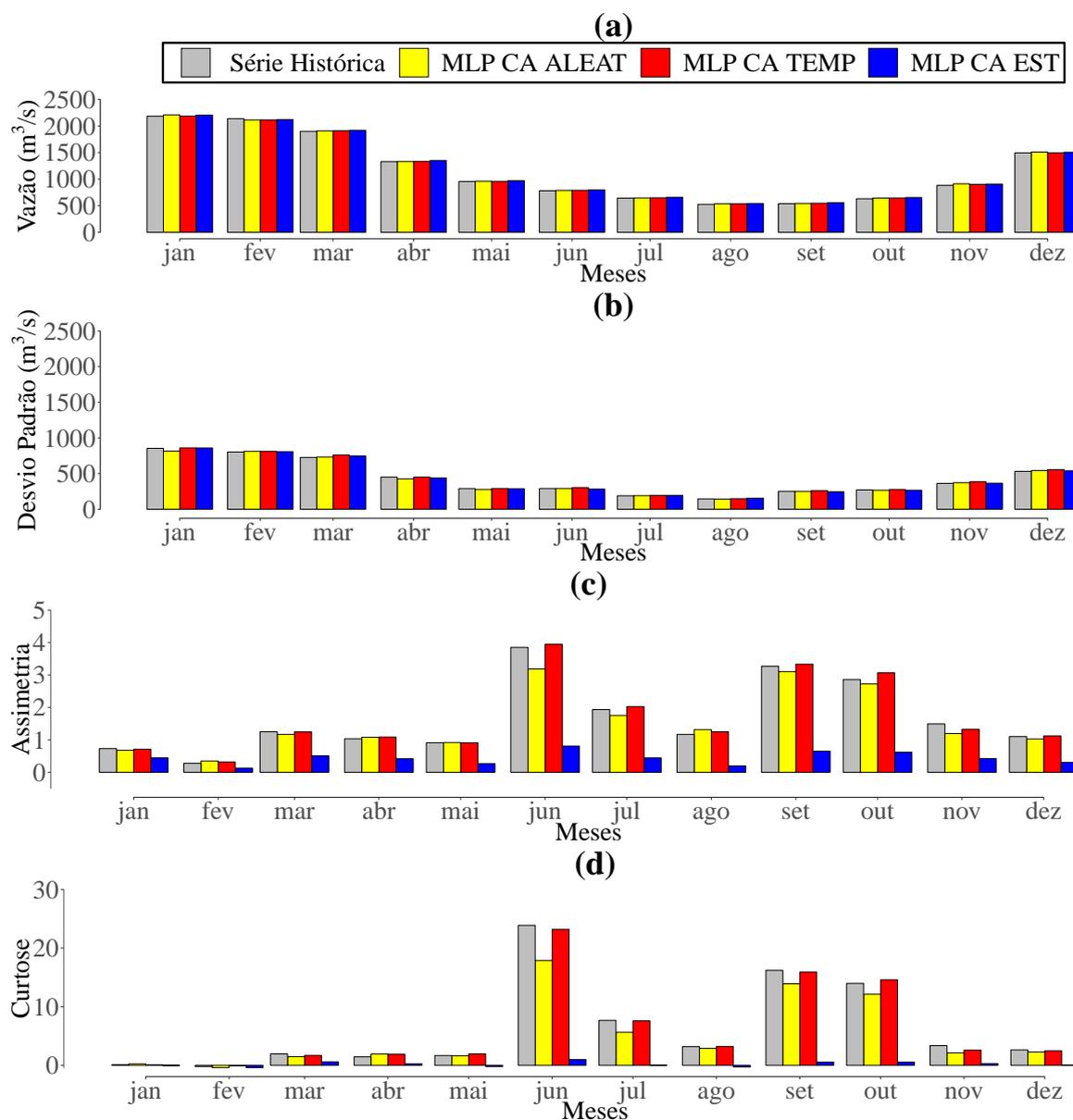
Tabela 16 – RMSE da FAC e FDP mensal dos modelos MLP de CA ALEAT, DIST.N e DIST.SN

RMSE da FAC			
RMSE da FAC			
Mês	MLP CA ALEAT	MLP CA DIST.N	MLP CA DIST.SN
janeiro	0,107	0,061	0,100
fevereiro	0,102	0,063	0,109
março	0,113	0,029	0,110
abril	0,081	0,048	0,090
maio	0,085	0,038	0,084
junho	0,072	0,026	0,081
julho	0,074	0,027	0,075
agosto	0,105	0,035	0,097
setembro	0,108	0,038	0,106
outubro	0,066	0,036	0,065
novembro	0,073	0,034	0,071
dezembro	0,110	0,03	0,097
Média dos meses	0,091	0,093	0,088
RMSE da FDP			
Mês	MLP CA ALEAT	MLP CA DIST.N	MLP CA DIST.SN
janeiro	0,026	0,036	0,027
fevereiro	0,018	0,025	0,031
março	0,023	0,034	0,027
abril	0,014	0,040	0,041
maio	0,008	0,032	0,033
junho	0,026	0,047	0,058
julho	0,030	0,047	0,066
agosto	0,026	0,047	0,051
setembro	0,012	0,072	0,082
outubro	0,025	0,065	0,084
novembro	0,031	0,060	0,054
dezembro	0,023	0,036	0,040
Média dos meses	0,019	0,049	0,056

Fonte: Elaborada pelo autor.

Analisando as estatísticas descritivas das outras duas propostas apresentadas nesse trabalho para essa técnica, as Figuras 71a e 71b é mostrado que os três modelos conseguiram reproduzir bem tanto a média e quanto desvio padrão da série histórica do dados em todos os meses. Porém, quanto a assimetria e curtose, apenas os modelos de CA TEMP e ALEAT obtiveram resultados poupáveis para essas estatística (ver Figuras 71c e 71d), enquanto modelo de CA EST apresentou muita dificuldades para reproduzir essas características, principalmente nos meses de valores mais altos como junho, setembro e outubro. Nesses meses inclusive, o modelo com CA ALEAT não foi tão bem quanto o de CA TEMP, mesmo tendo resultados bem próximos do da série histórica.

Figura 71 – Comparação entre as estatísticas descritivas das vazões mensais geradas pelos modelos MLP de CA ALEAT, TEMPO e EST para usina de Volta Grande

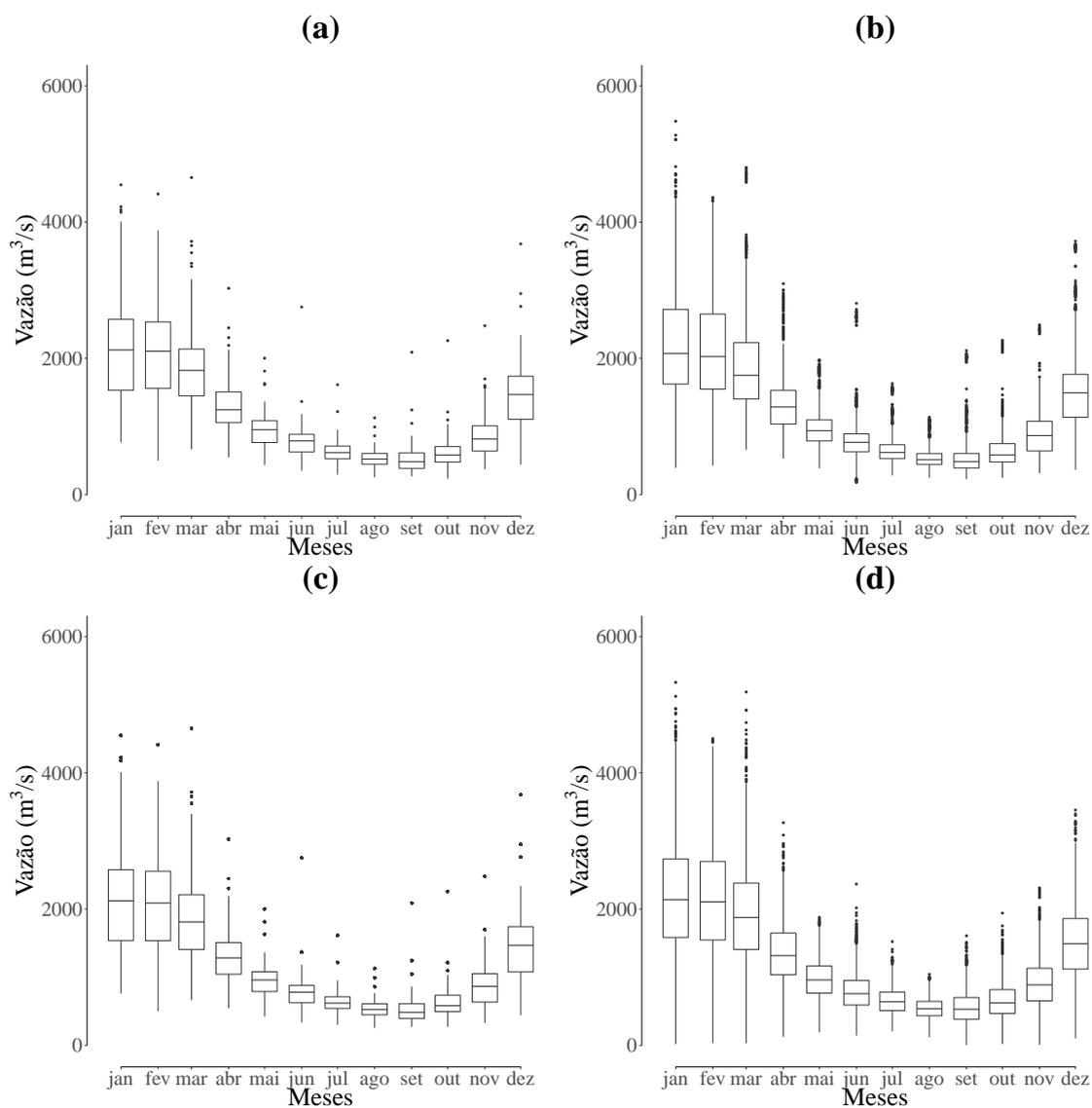


Legenda: a) Média
 b) Desvio Padrão
 c) Assimetria
 d) Curtose

Fonte: Elaborada pelo autor.

Através do box-plot das vazões geradas pelos modelos exibidos nas Figuras 72c e 72d, pode-se notar que o modelo com CA TEMP gera poucos *outliers*, porém os valores gerados por ele são bem próximos aos presentes na série histórica. Já o modelo com CA EST representou as amplitudes muito acima da real, cometendo equívocos quanto aos *outliers*, enquanto o modelo com CA ALEAT produziu *outliers* inferiores em junho, algo que não ocorre na série histórica.

Figura 72 – Box-plots das vazões geradas pelos modelos MLP para usina de Volta Grande



Legenda: a) Série Histórica

b) CA ALEAT

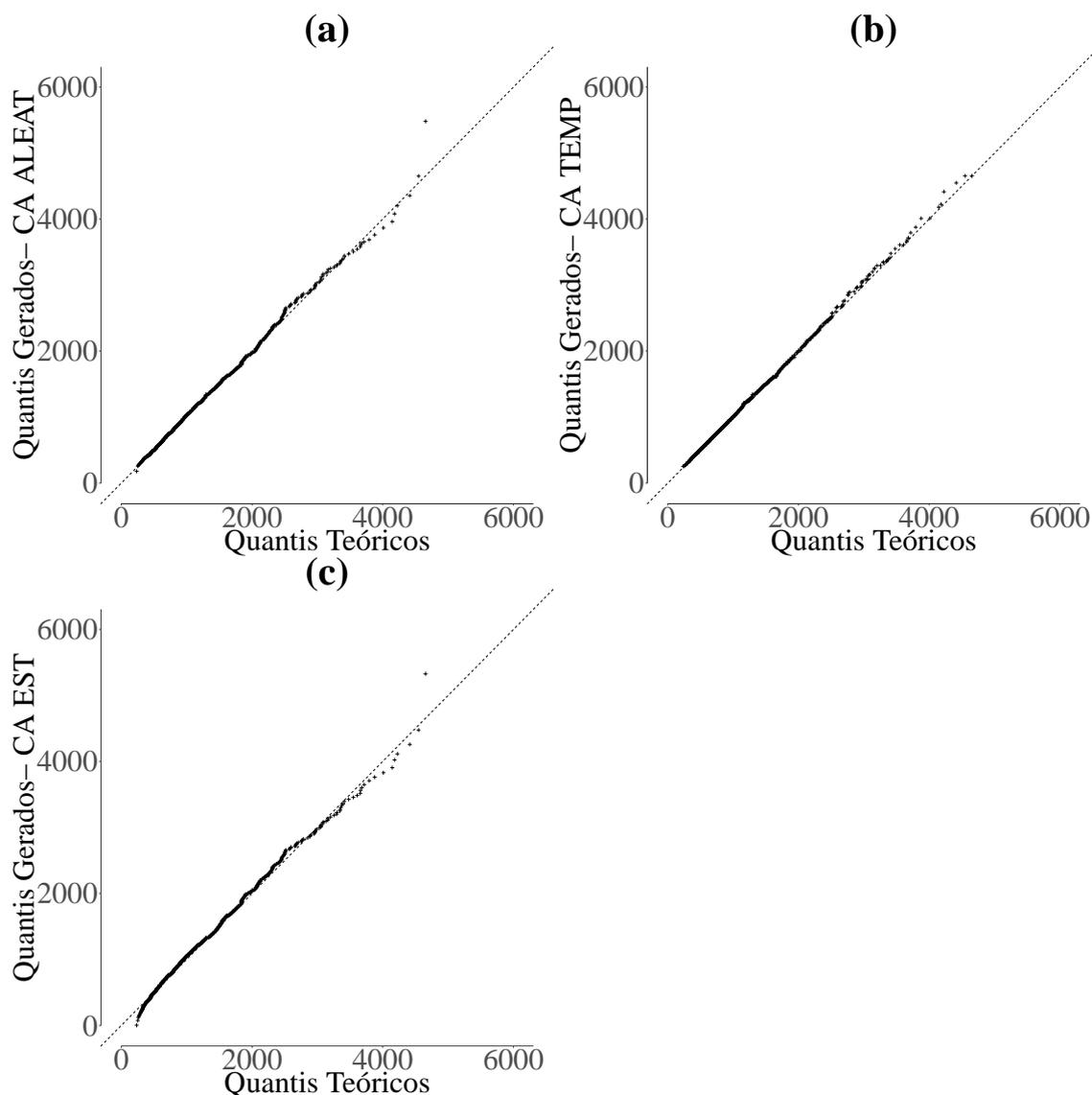
c) CA TEMP

d) CA EST

Fonte: Elaborada pelo autor.

Analisando os q-qplots apresentados nas Figuras 73b e 73c fica evidente que o modelo de CA TEMP se ajustou tão bem aos dados quanto o modelo de CA ALEAT, enquanto o com CA EST já apresentou mais dificuldades quanto a essa representação (ver Figura 73a).

Figura 73 – Q-plots das vazões geradas pelos modelos MLP para usina de Volta Grande



Legenda: a) CA ALEAT

b) CA TEMP

c) CA EST

Fonte: Elaborada pelo autor.

Por fim, explorando os RMSE da FAC e FDP desses três modelos (ver Tabela 17), constata-se que o modelo de CA TEMP apresentou erros menores para a FAC em todos os meses, com exceção de março. O mesmo efeito ocorreu na maior parte dos meses FDP, onde os erros foram menores ou muito próximos, com exceção do mês de setembro, onde houve empate, e março. O modelo com CA EST para MLP não conseguiu resultados satisfatórios neste caso, principalmente no erro da FDP, que foi significativamente superior aos dos outros dois modelos em praticamente todos os meses.

Tabela 17 – RMSE da FAC e FDP mensal dos modelos MLP de CA ALEAT, TEMP e EST

RMSE da FAC			
Mês	MLP CA ALEAT	MLP CA TEMP	MLP CA EST
janeiro	0,021	0,012	0,020
fevereiro	0,018	0,019	0,018
março	0,020	0,010	0,043
abril	0,032	0,016	0,030
maio	0,019	0,012	0,036
junho	0,014	0,013	0,061
julho	0,022	0,005	0,034
agosto	0,019	0,013	0,036
setembro	0,013	0,007	0,072
outubro	0,024	0,022	0,053
novembro	0,016	0,021	0,054
dezembro	0,009	0,012	0,033
Média dos meses	0,091	0,039	0,091
RMSE da FDP			
Mês	MLP CA ALEAT	MLP CA TEMP	MLP CA EST
janeiro	0,021	0,012	0,020
fevereiro	0,018	0,019	0,018
março	0,020	0,01	0,043
abril	0,032	0,016	0,03
maio	0,019	0,012	0,036
junho	0,014	0,013	0,061
julho	0,022	0,005	0,034
agosto	0,019	0,013	0,036
setembro	0,013	0,007	0,072
outubro	0,024	0,022	0,053
novembro	0,016	0,021	0,054
dezembro	0,009	0,012	0,033
Média dos meses	0,019	0,014	0,041

Fonte: Elaborada pelo autor.

Portanto, concluí-se que o modelo MLP CA TEMP se apresenta como uma alternativa melhor dentre todos os modelos apresentados nesta seção.

APÊNDICE I - MODELOS RF PARA USINA DE VOLTA GRANDE

No modelos *Random Forest* para geração de séries sintéticas das vazões para usina de Volta Grande, várias estruturas foram testadas para os dados mensais com intuito de se obter a melhor estrutura para o modelo. A estrutura do modelo dos parâmetros na Seção 2.4.4 são exibidos na Tabela 18. O critério utilizado para seleção desses parâmetros foi o RMSE da FDP, onde a estrutura com menor erro foi a selecionada.

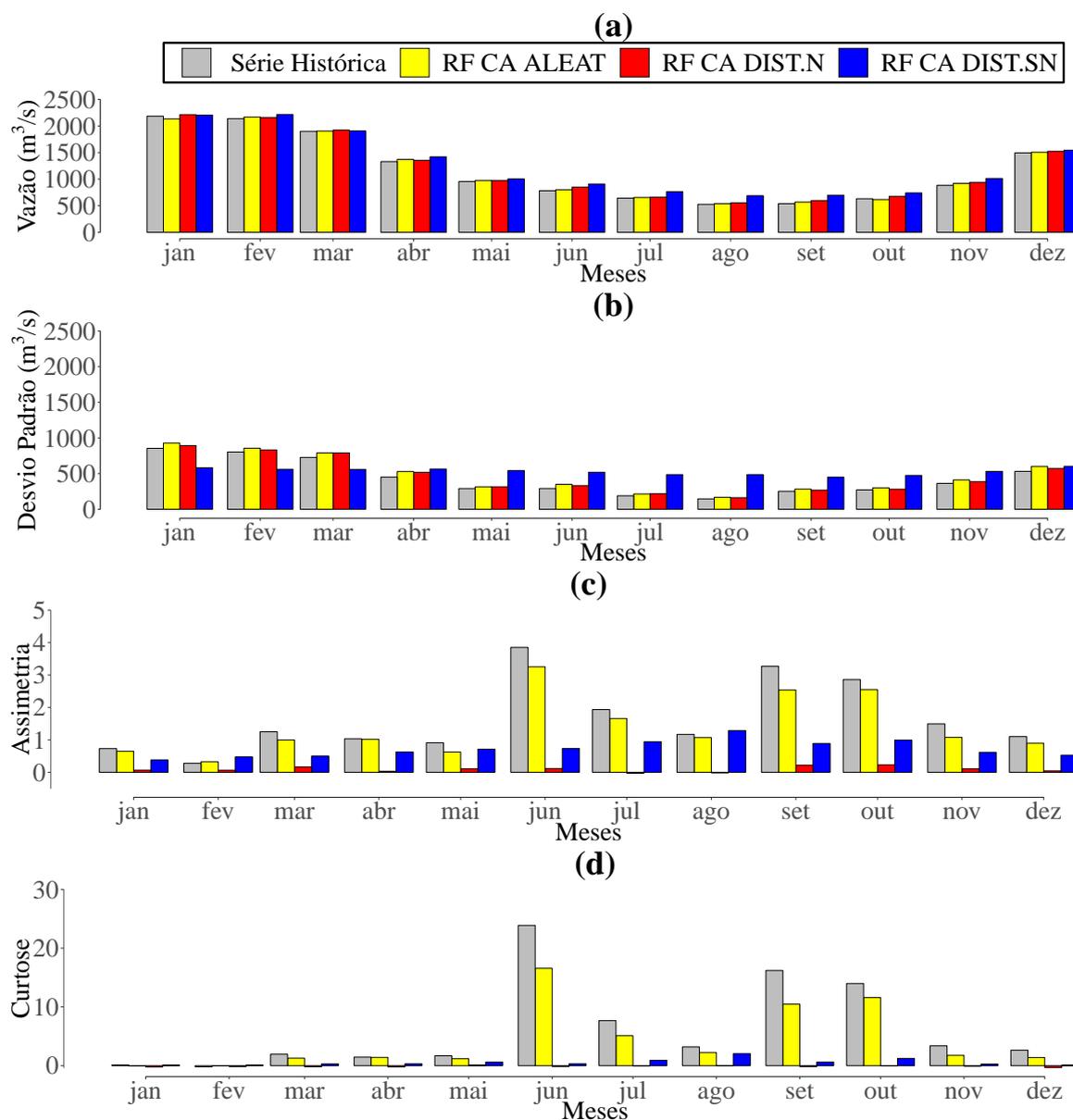
Tabela 18 – Estruturas do modelo *Random Forest* para geração de séries sintéticas mensais na usina de Água Vermelha

Mês	nodesize	ntree	nperm	Corr.Bies	Atrasos
janeiro	3	600	0.1	6	8
fevereiro	9	800	0.1	6	2
março	9	100	0.1	6	9
abril	9	300	0.1	6	9
maio	6	800	0.1	6	7
junho	9	400	0.1	6	11
julho	3	300	0.1	6	8
agosto	12	100	0.1	6	5
setembro	9	700	0.1	6	9
outubro	3	1000	0.1	6	6
novembro	6	500	0.1	6	9
dezembro	3	600	0.1	6	6

Fonte: Elaborada pelo autor.

Após selecionados os parâmetros, a primeira das comparações executadas foi entre os modelos com CA DIST.N e DIST.SN. A Figura 74a mostra que os três modelos até que conseguiram representar bem a série histórica das médias, mas quanto aos desvios padrão, na Figura 74b já nota-se uma grande dificuldade do modelo com CA DIST.N para replicar essa característica, conseguindo um resultado razoável somente no mês de dezembro. Mas nas Figuras 74c e 74d, vê-se que os modelos com CA DIST.N e DIST.SN apresentaram uma enorme dificuldade em replicar tanto a curtose quanto assimetria em praticamente todos os meses.

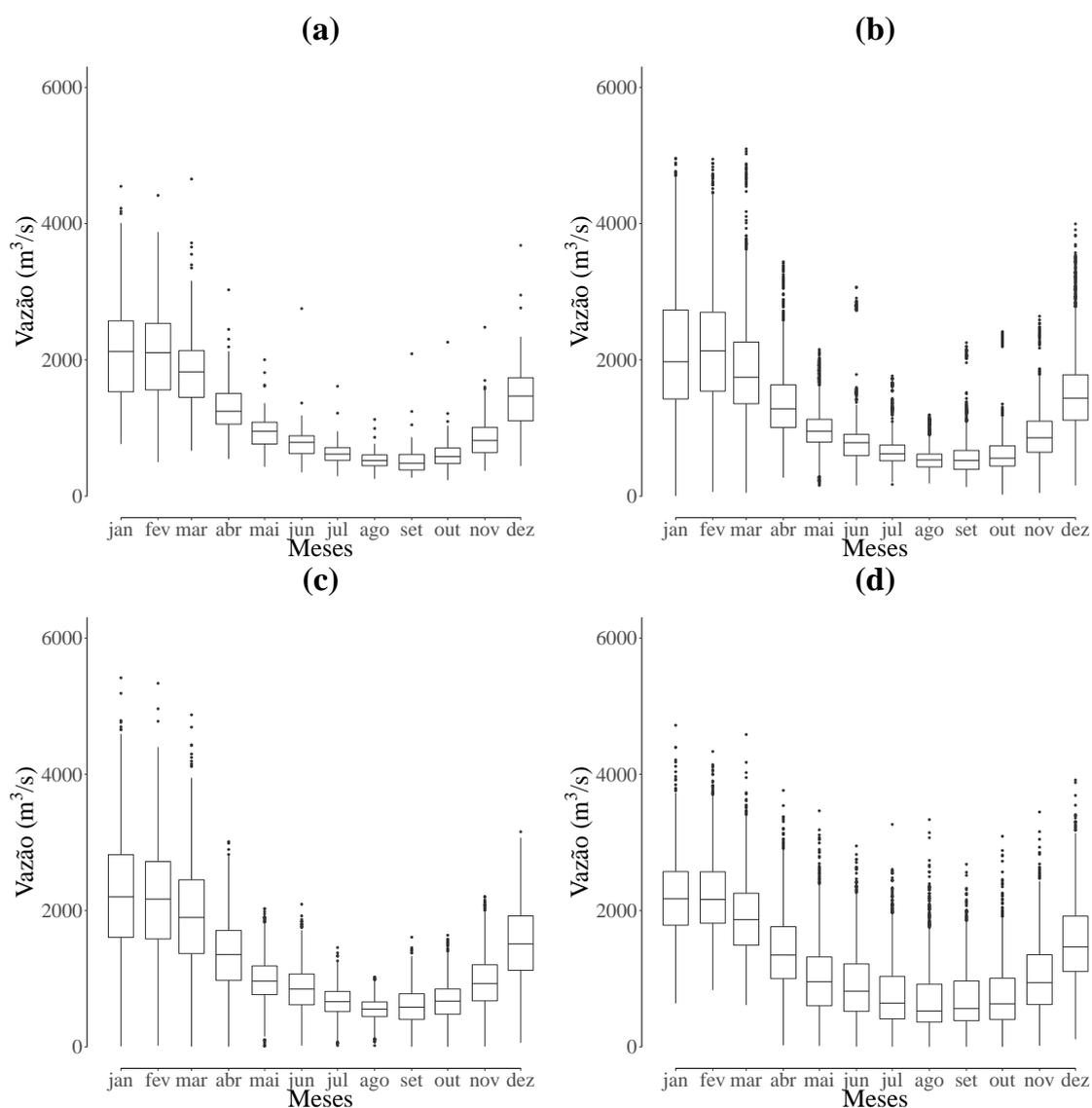
Figura 74 – Comparação entre as estatísticas descritivas das vazões geradas pelos modelos RF de CA ALEAT, DIST.N e DIST.SN para usina de Volta Grande



Nos box-plots das vazões geradas pelos modelos apresentados nas Figuras 75b, 75c e 75d, é exposta uma dificuldade quanto aos *outliers*. Mesmo o modelo com CA ALEAT, que conseguiu representar melhor a série histórica, nota-se que *outliers* inferiores foram gerados para os meses de fevereiro, maio, agosto e novembro, enquanto na série histórica da Figura 75a não há nenhum. O modelo com CA DIST.N gerou *outliers* em abril, maio, junho, agosto e

setembro, enquanto o com CA DIST.N produziu em abril, maio, junho e agosto. A amplitude dos meses também foi mal representada nos modelos de CA DIST.N e DIST.SN, em especial no mês de fevereiro.

Figura 75 – Box-plots das vazões geradas pelos modelos RF para usina de Volta Grande



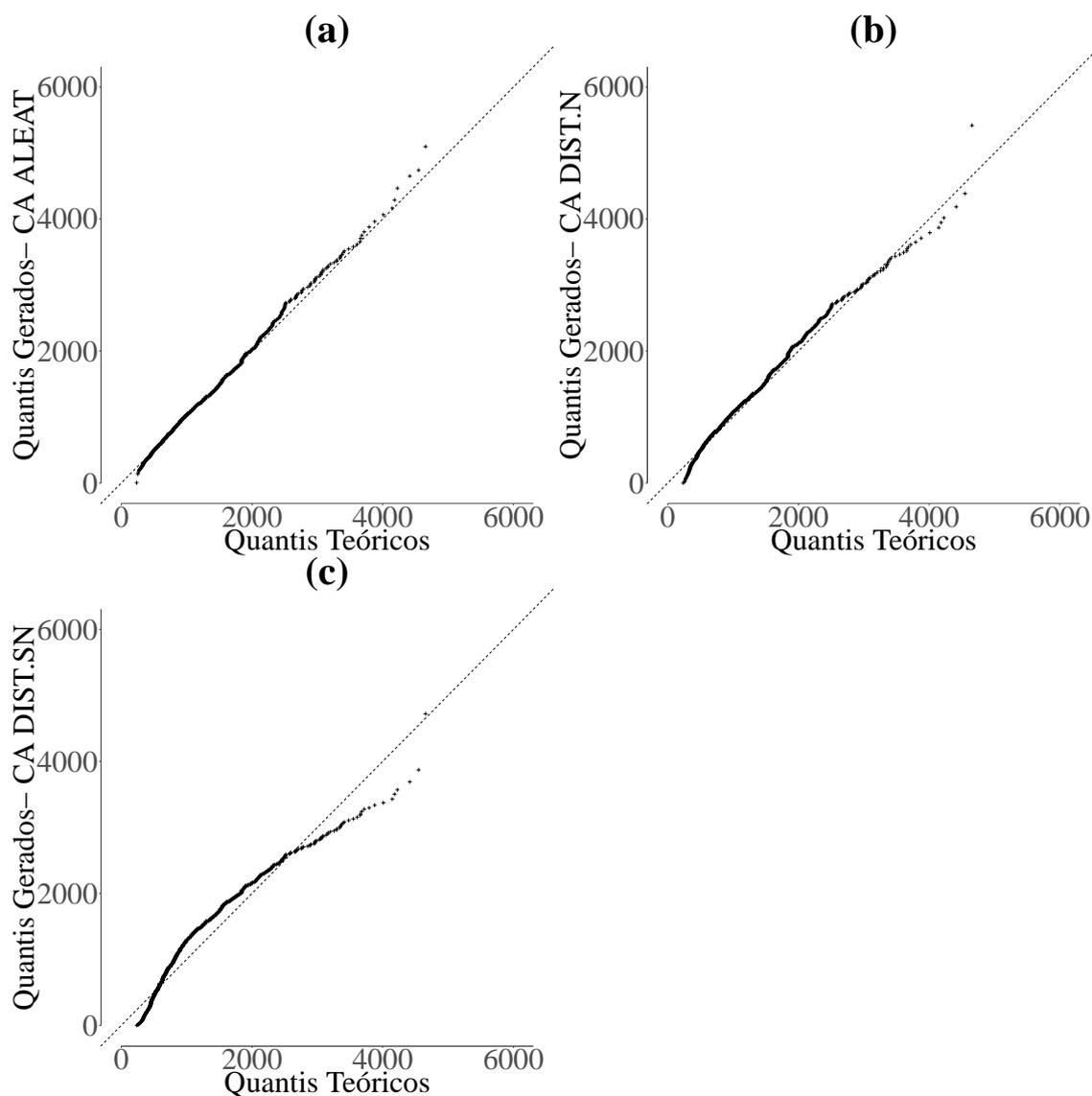
Legenda: a) Série Histórica
 b) CA ALEAT
 c) CA DIST.N
 d) CA DIST.SN

Fonte: Elaborada pelo autor.

Pelos q-qplots apresentados nas Figuras 76b e 76c fica mais claro ainda que ambas as propostas de modelagem para os CA não se ajustaram bem aos dados, em especial nos valores extremos da série, onde mostram uma dificuldade maior em reproduzir esses valores quando comparado ao q-qplot da Figura 76a. Todos os três modelos também mostraram dificuldade

na representação dos extremos inferiores da série, porém essa dificuldade foi menor no de CA ALEAT.

Figura 76 – Q-qplots das vazões geradas pelos modelos RF para usina de Volta Grande



Legenda: a) CA ALEAT
b) CA DIST.N
c) CA DIST.SN

Fonte: Elaborada pelo autor.

Na comparação dos erros da FAC e FDP conforme a Tabela 19, concluí-se que os modelos RF de CA DIST.N e DIST.SN não foram uma opção tão viável quanto o modelo com CA ALEAT, visto que como dito anteriormente, além de não conseguirem replicar bem os *outliers*, desvios padrão, curtoses e assimetrias em vários meses, além de apresentarem um erro da FAC maior ou muito próximo ao do modelo com CA ALEAT e erros bem superiores quanto a FDP.

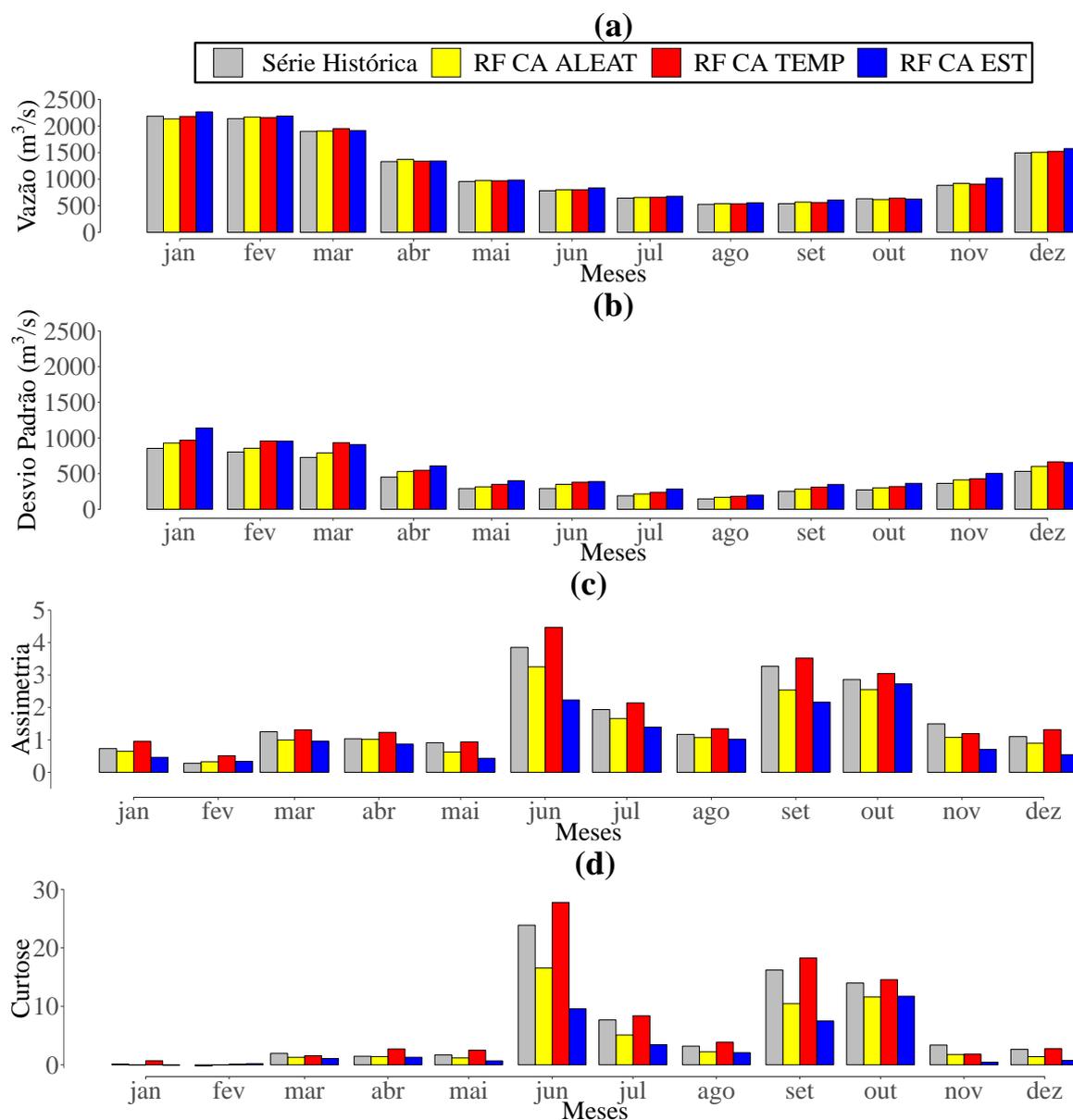
Tabela 19 – RMSE da FAC e FDP mensal dos modelos RF de CA ALEAT, DIST.N e DIST.SN

RMSE da FAC			
Mês	RF CA ALEAT	RF CA DIST.N	RF CA DIST.SN
janeiro	0,096	0,087	0,092
fevereiro	0,088	0,102	0,094
março	0,104	0,101	0,093
abril	0,075	0,091	0,081
maio	0,072	0,081	0,085
junho	0,062	0,070	0,073
julho	0,072	0,061	0,070
agosto	0,092	0,080	0,098
setembro	0,09	0,093	0,087
outubro	0,065	0,058	0,070
novembro	0,059	0,069	0,051
dezembro	0,093	0,093	0,082
Média dos meses	0,081	0,082	0,081
RMSE da FDP			
Mês	RF CA ALEAT	RF CA DIST.N	RF CA DIST.SN
janeiro	0,038	0,032	0,047
fevereiro	0,017	0,025	0,053
março	0,031	0,069	0,031
abril	0,026	0,044	0,047
maio	0,025	0,043	0,079
junho	0,021	0,080	0,118
julho	0,019	0,065	0,126
agosto	0,031	0,042	0,116
setembro	0,027	0,093	0,090
outubro	0,026	0,085	0,108
novembro	0,034	0,065	0,066
dezembro	0,032	0,037	0,036
Média dos meses	0,027	0,057	0,077

Fonte: Elaborada pelo autor.

Pelas outras duas propostas apresentadas nesse trabalho, através da análise das estatísticas descritivas presente nas Figuras 77a, 77b, 77c e 77d mostram que os três modelos conseguiram de reproduzir a média da série, quanto ao desvio padrão, os modelos de CA TEMP e EST apresentaram uma dificuldade maior em relação ao modelo do CA ALEAT. Já na assimetria e curtose, o modelo de CA EST não apresentou bons resultados principalmente nos meses de maior valor, enquanto o modelo com CA TEMP conseguiu replicar essa característica de maneira mais fiel do que os demais modelos testados, mesmo para os meses de maior valor, que apresentou essa dificuldade em todos os demais modelos.

Figura 77 – Comparação entre as estatísticas descritivas das vazões mensais geradas pelos modelos RF de CA ALEAT, TEMP e EST para usina de Volta Grande

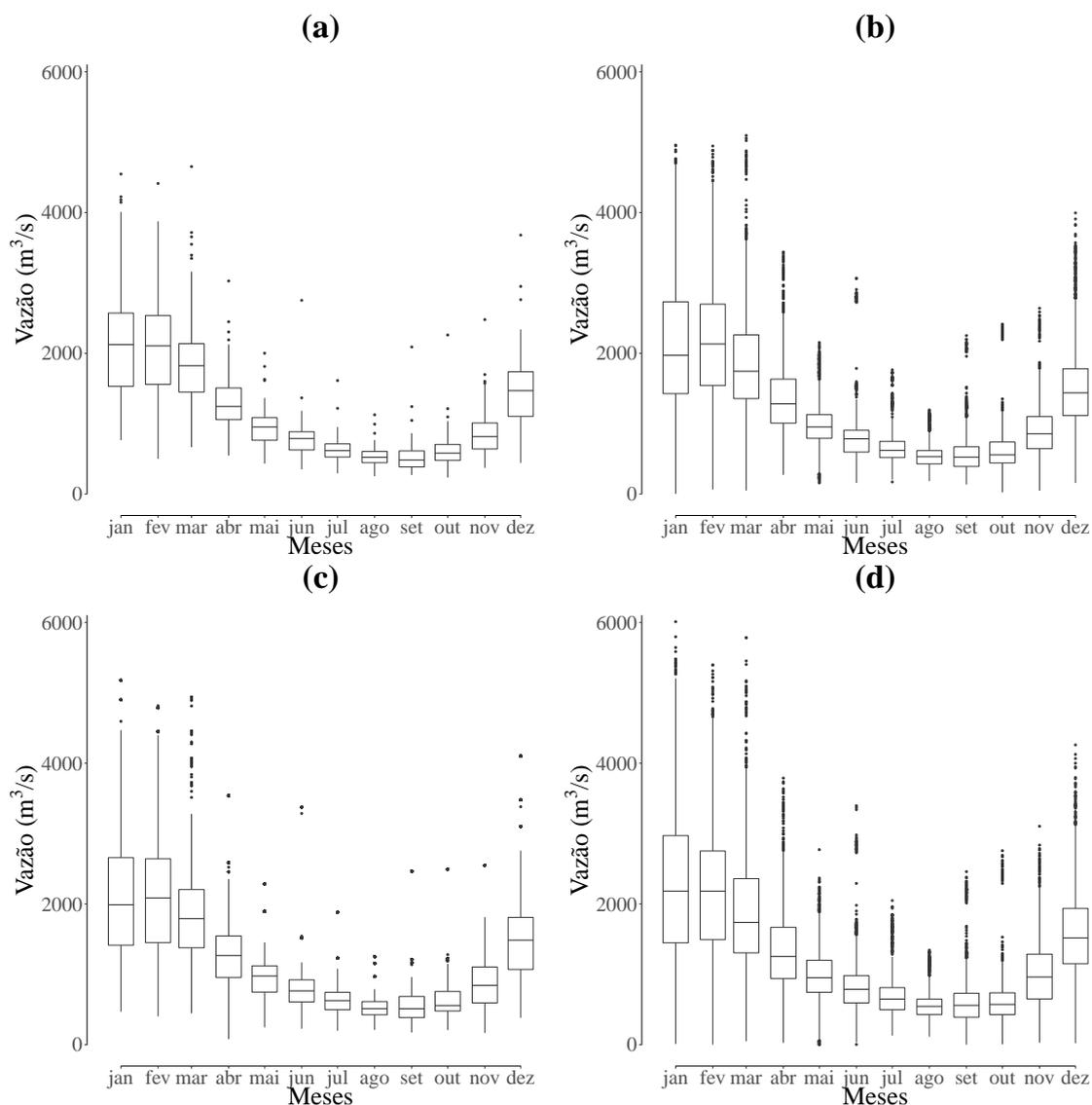


Fonte: Elaborada pelo autor.

Através do box-plots das vazões geradas pelos modelos exibidos nas Figuras 78c e 78d, nota-se que o modelo com CA TEMP gera proporcionalmente menos *outliers* em todos os meses, exceto no mês de março, mas os *outliers* superiores não são tão bem representados quanto no modelo com CA ALEAT. Quando comparado ao de CA ALEAT (ver Figura 78b) nota-se que o modelo de CA ALEAT consegue representar melhor a amplitude da série na maior

parte dos meses, além de não gerar nenhum *outlier* inferior, sendo fiel ao que ocorre na série histórica (ver Figura 78a). Já o modelo com CA EST produziu *outliers* inferiores com valores próximos aos presentes na série histórica, assim como o modelo de CA ALEAT, porém também houve ocorrência *outliers* inferiores nos meses de maio e junho, além de ter representado muito mal a amplitude da série em praticamente todos os meses (ver Figura 78d).

Figura 78 – Box-plots das vazões geradas pelos modelos RF para usina de Volta Grande



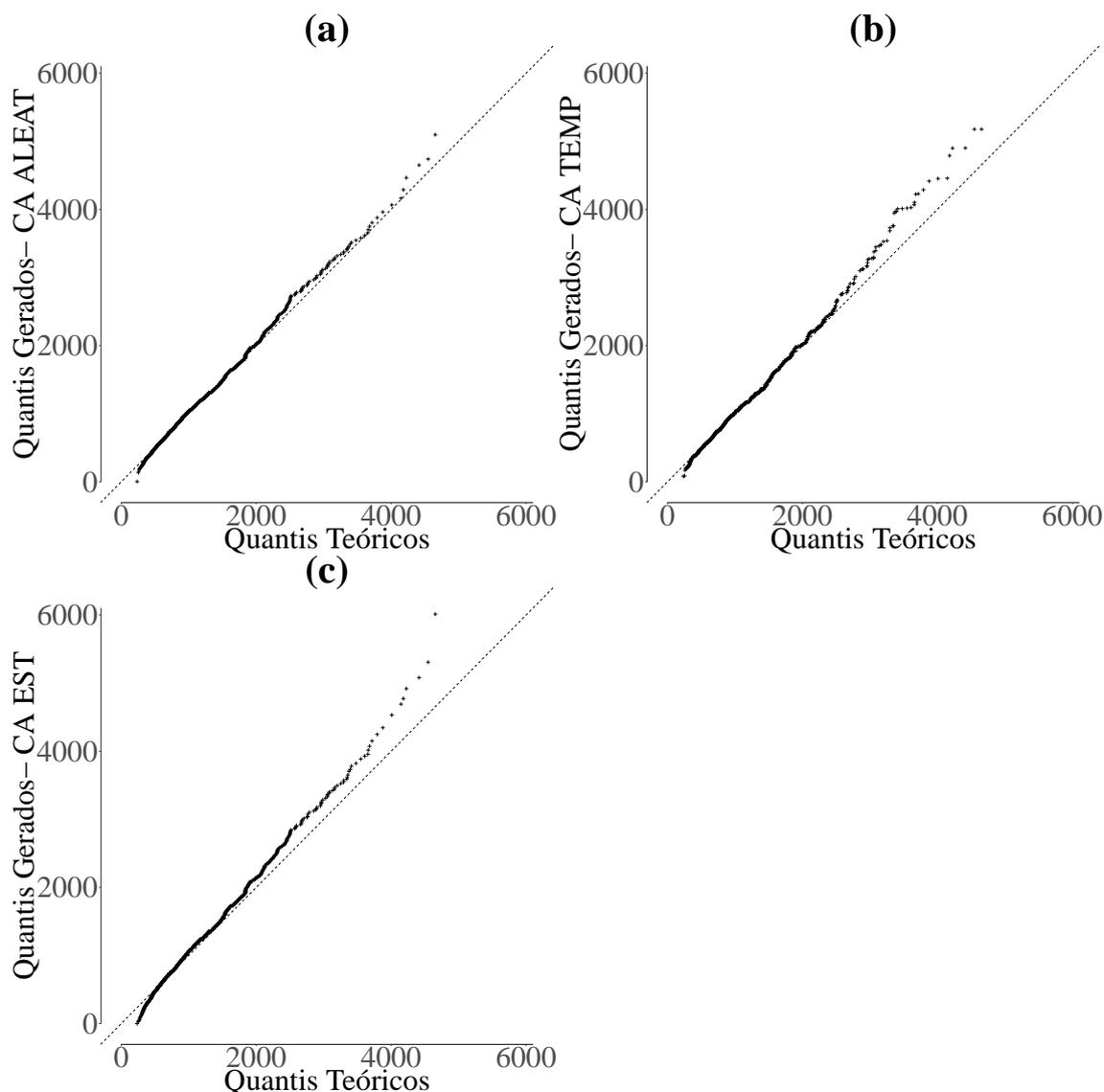
Legenda: a) Série Histórica
 b) CA ALEAT
 c) CA TEMP
 d) CA EST

Fonte: Elaborada pelo autor.

Através dos q-qplots apresentados nas Figuras 79b e 79c nota-se que os modelos de CA TEMP e EST apresentam um ajuste pior nos extremos em comparação ao modelo de CA

ALEAT (ver Figura 79a) nos quantis mais elevados, mas nos menores, o CA TEMP parece reproduzir um pouco melhor essa característica do que o CA ALEAT.

Figura 79 – Q-plots das vazões geradas pelos modelos RF para usina de Volta Grande



Legenda: a) CA ALEAT
 b) CA TEMP
 c) CA EST

Fonte: Elaborada pelo autor.

Por fim, analisando os RMSE da FAC e FDP desses três modelos (ver Tabela 20), constata-se que o modelo de CA TEMP apresentou um erro menor da FAC em todos os meses. Efeito esse semelhante ao ocorrido no erro da FDP, onde na maior parte dos meses houve uma diminuição desse erro, em especial no mês de outubro. Quanto ao modelo com CA EST, sua FAC teve valores próximos ou um pouco menores que o modelo de CA ALEAT, mas sempre maior que o de CA TEMP. Já o erro da sua FDP foi a maior em todos os meses.

Tabela 20 – RMSE da FAC e FDP mensal dos modelos RF de CA ALEAT, TEMP e EST

RMSE da FAC			
Mês	RF CA ALEAT	RF CA TEMP	RF CA EST
janeiro	0,096	0,080	0,079
fevereiro	0,088	0,064	0,107
março	0,104	0,052	0,101
abril	0,075	0,063	0,084
maio	0,072	0,060	0,071
junho	0,062	0,041	0,077
julho	0,072	0,045	0,069
agosto	0,092	0,062	0,092
setembro	0,090	0,045	0,087
outubro	0,065	0,034	0,055
novembro	0,059	0,031	0,055
dezembro	0,093	0,043	0,082
Média dos meses	0,081	0,052	0,080
RMSE da FDP			
Mês	RF CA ALEAT	RF CA TEMP	RF CA EST
janeiro	0,038	0,026	0,046
fevereiro	0,017	0,022	0,034
março	0,031	0,025	0,044
abril	0,026	0,017	0,046
maio	0,025	0,016	0,044
junho	0,021	0,015	0,055
julho	0,019	0,021	0,057
agosto	0,031	0,017	0,041
setembro	0,027	0,022	0,058
outubro	0,026	0,008	0,047
novembro	0,034	0,002	0,072
dezembro	0,032	0,021	0,040
Média dos meses	0,027	0,019	0,049

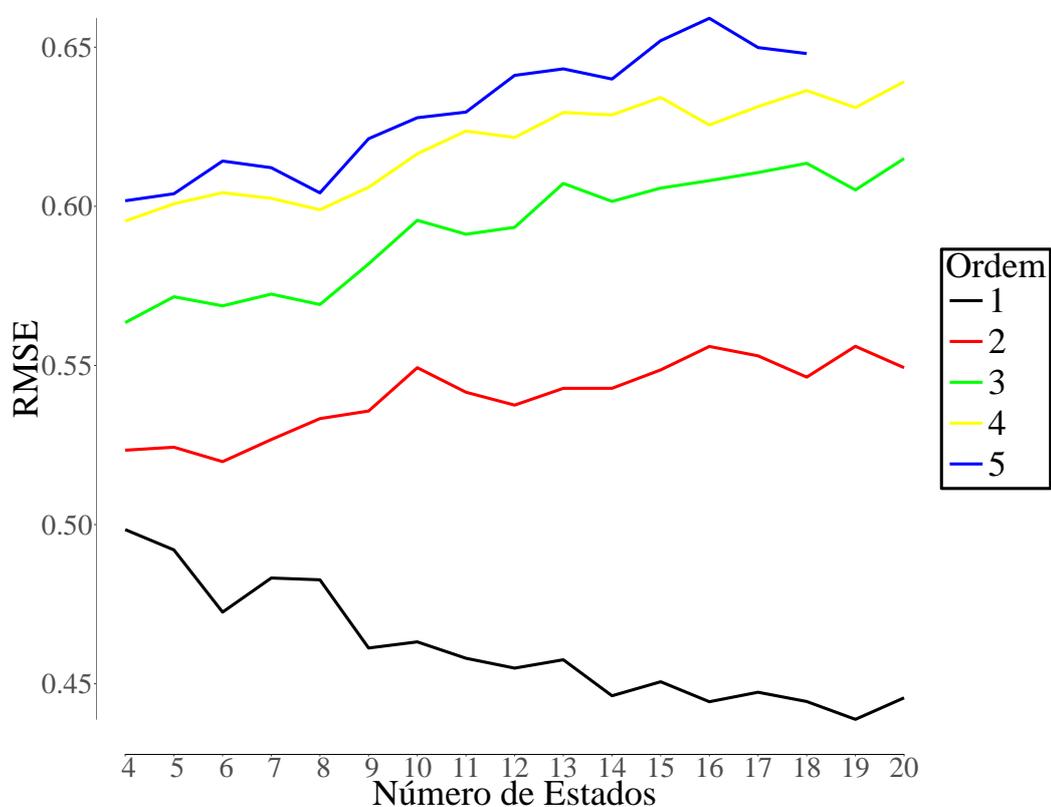
Fonte: Elaborada pelo autor.

Portanto, concluí-se que o modelo RF CA TEMP se mostrou a melhor opção entre os demais modelos, visto que possui menor erro da FDP na maioria dos meses, menor erro da FAC, obteve uma melhor representação da curtose e assimetria além de não produzir *outliers* inferiores.

APÊNDICE J - MODELO SYNTISE PARA USINA DE VOLTA GRANDE

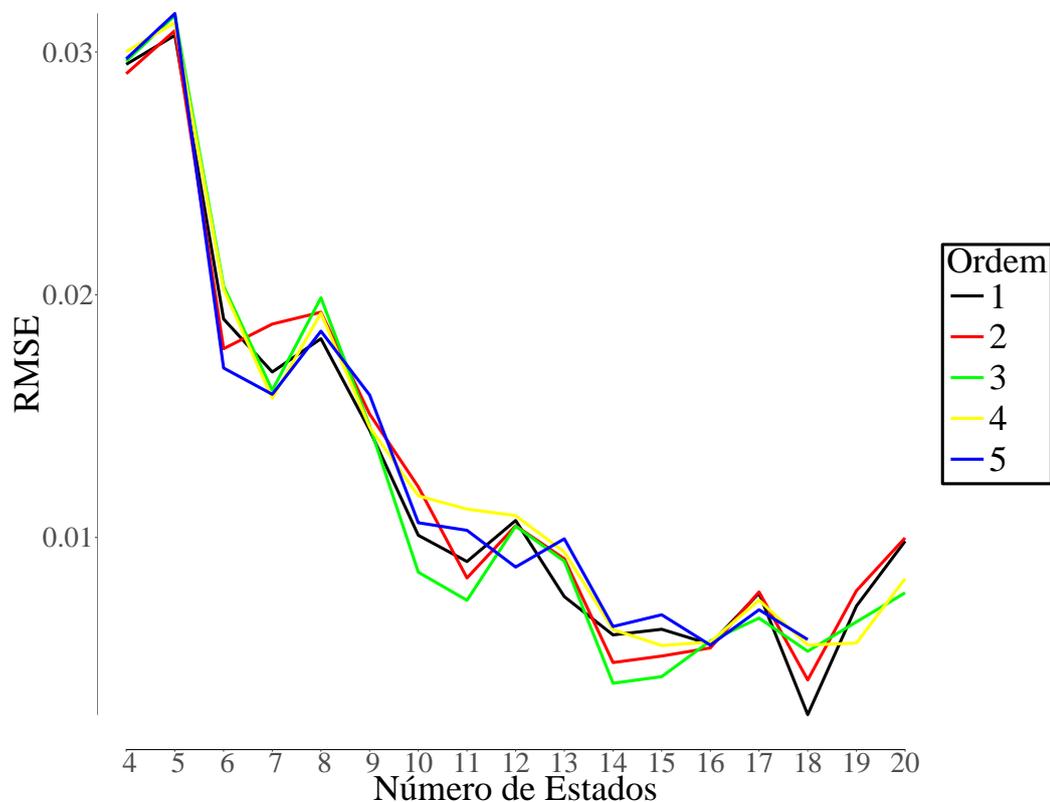
Analisando o erro da FAC através da Figura 80, fica evidente que a ordem mais adequada para o modelo é 1, já que seu erro foi menor que os modelos de demais ordem em todos os estados. Quanto ao estado, através do erro da FDP (ver Figura 81), nota-se que o menor entre os erros aparenta ser o de 18 estados. Portanto assim o modelo gerado para essa usina será de ordem 1 e 18 estados.

Figura 80 – RMSE da FAC dos modelos de série possíveis no SynTise para usina de Volta Grande



Fonte: Elaborada pelo autor.

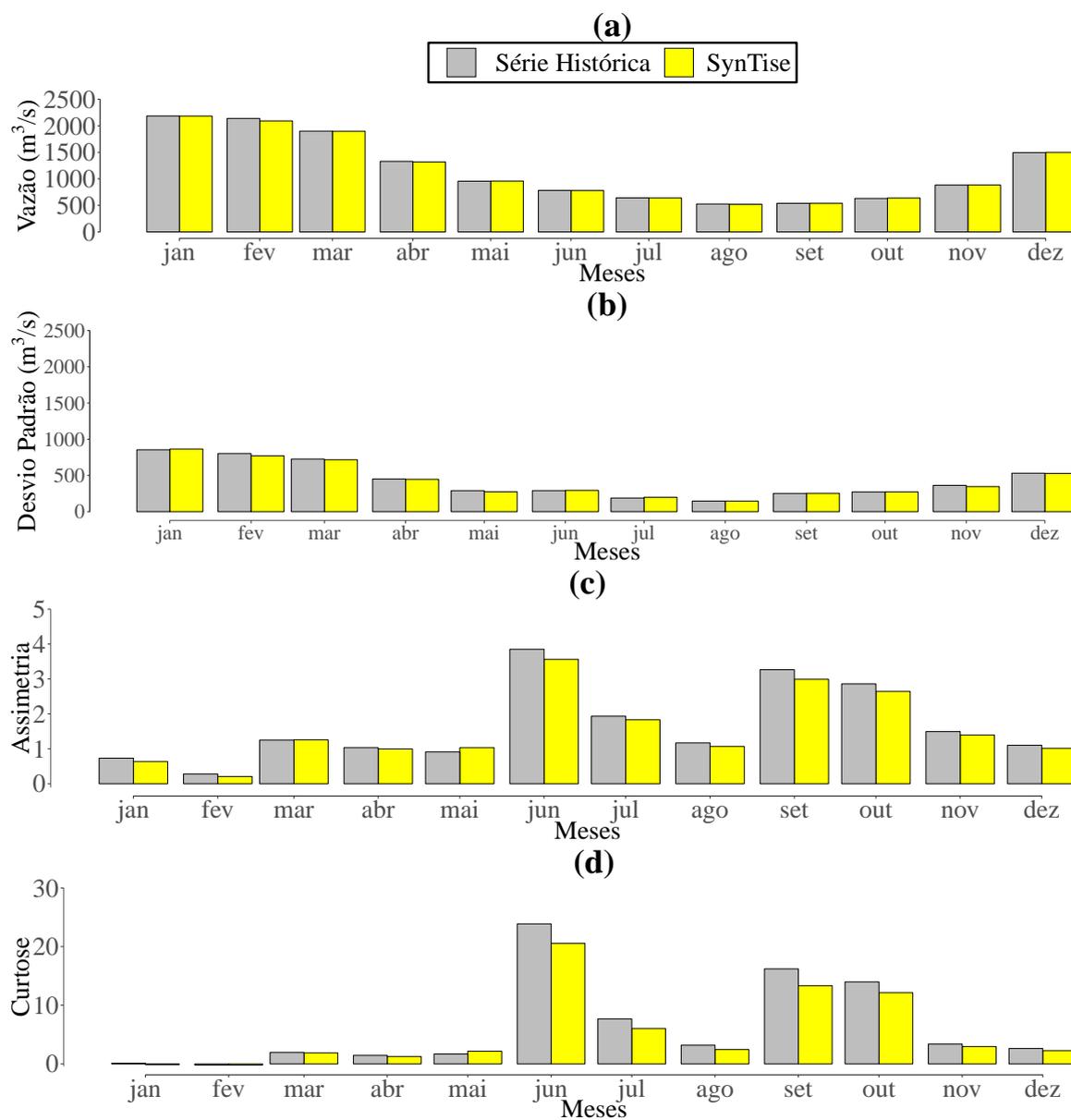
Figura 81 – RMSE da FDP dos modelos de série possíveis no SynTise para usina de Volta Grande



Fonte: Elaborada pelo autor.

Analisando as estatísticas descritivas do modelo, nota-se que as médias e desvios padrão foram muito bem representadas no SynTise (ver Figuras 82a e 82b). Mesmo não tendo a mesma precisão quanto a estimativa da curtose e assimetria, ainda sim essas estatísticas foram bem replicadas, conforme as Figuras 82c e 82d.

Figura 82 – Comparação entre as estatísticas descritivas das vazões mensais geradas pelo modelo SynTise para usina de Volta Grande

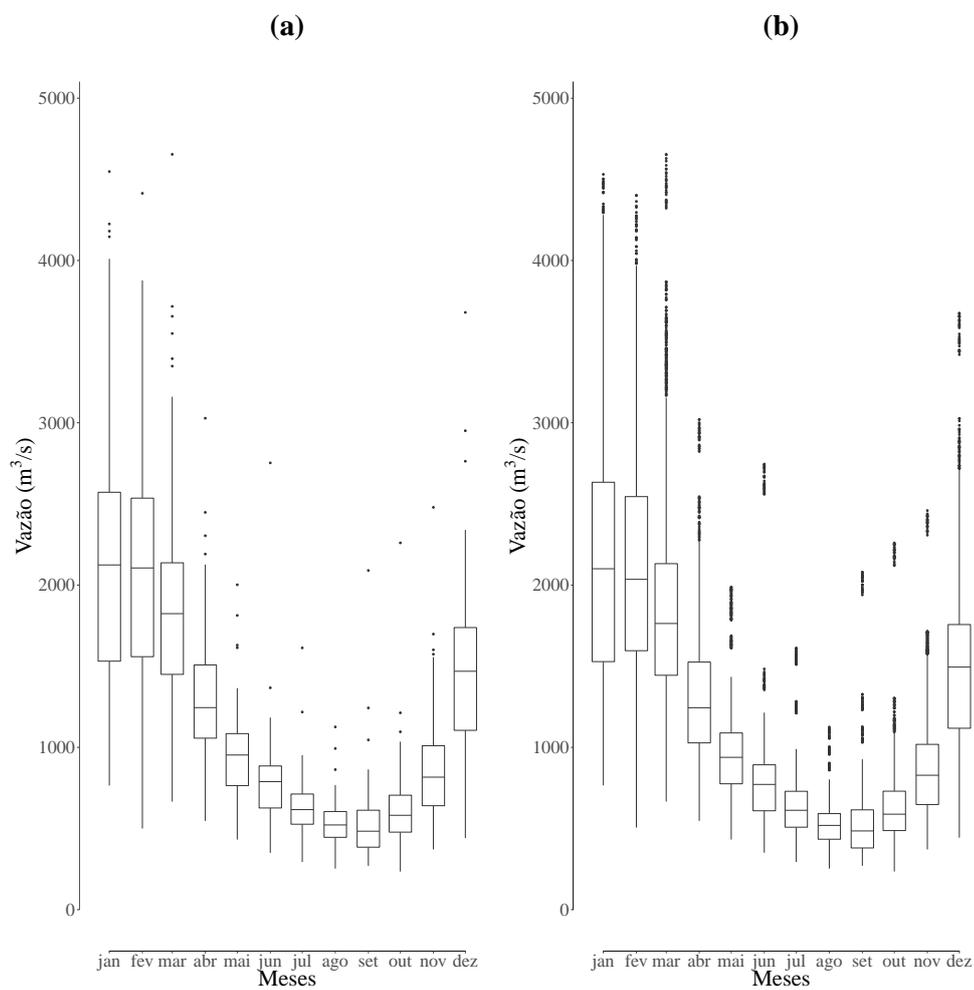


Legenda: a) Média
 b) Desvio Padrão
 c) Assimetria
 d) Curtose

Fonte: Elaborada pelo autor.

O box-plot e q-qplot das Figuras 83b e 84, respectivamente, demonstram como que a estrutura da série histórica foi bem replicada por esse modelo, inclusive nos *outliers*, que em todos os meses foram bem representados.

Figura 83 – Box-plots das vazões geradas pelo SynTise para usina de Volta Grande

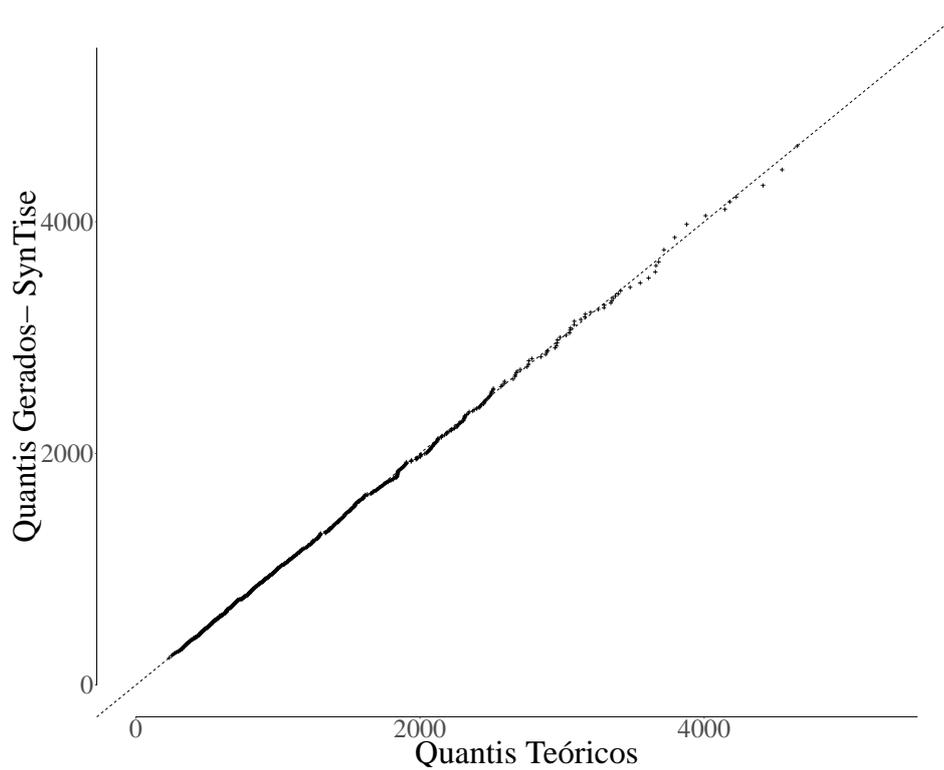


Legenda: a) Série Histórica

b) SynTise

Fonte: Elaborada pelo autor.

Figura 84 – Q-plot das vazões geradas pelo SynTise



Fonte: Elaborada pelo autor.

Portanto, através dessas análises, é possível concluir que o SynTise reproduziu bem as principais características dos dados de vazões mensais da usina de Volta Grande, visto que conseguiu replicar de maneira fidedigna as estatísticas descritivas e os *outliers* da série em todos os meses.