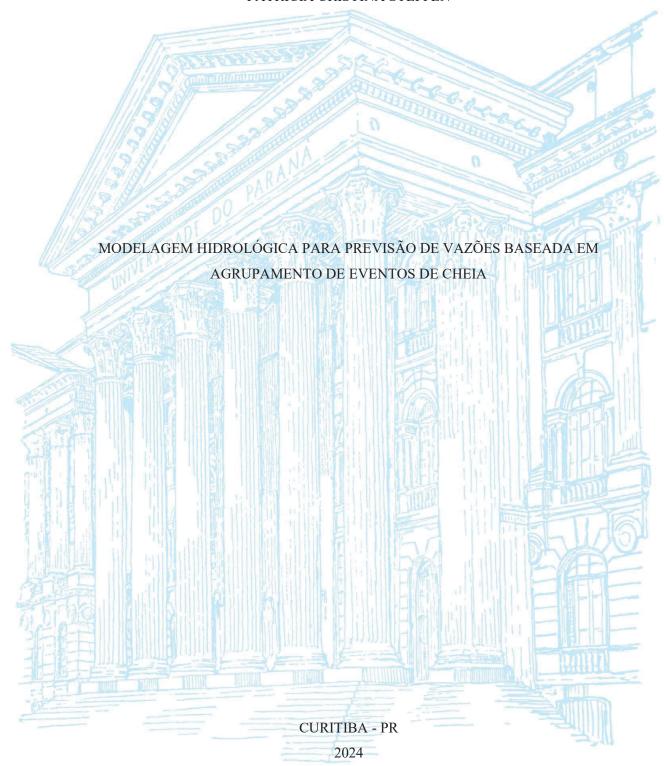
UNIVERSIDADE FEDERAL DO PARANÁ

PATRÍCIA CRISTINA STEFFEN



PATRÍCIA CRISTINA STEFFEN

MODELAGEM HIDROLÓGICA PARA PREVISÃO DE VAZÕES BASEADA EM AGRUPAMENTO DE EVENTOS DE CHEIA

Tese apresentada ao curso de Pós-Graduação em Engenharia de Recursos Hídricos e Ambiental, Setor de Tecnologia, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Doutora em Engenharia de Recursos Hídricos e Ambiental.

Orientador: Prof. Dr. Júlio Gomes

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP) UNIVERSIDADE FEDERAL DO PARANÁ SISTEMA DE BIBLIOTECAS – BIBLIOTECA CIÊNCIA E TECNOLOGIA

Steffen, Patrícia Cristina

Modelagem hidrológica para previsão de vazões baseada em agrupamento de eventos de cheia. / Patrícia Cristina Steffen. — Curitiba, 2024.

1 recurso on-line: PDF.

Tese (Doutorado) - Universidade Federal do Paraná, Setor de Tecnologia, Programa de Pós-Graduação em Engenharia de Recursos Hídricos e Ambiental.

Orientador: Prof. Dr. Júlio Gomes

1. Inteligência artificial. 2. Aprendizado do computador. 3. Regressão por vetores de suporte. I. Universidade Federal do Paraná. II. Programa de Pós-Graduação em Engenharia de Recursos Hídricos e Ambiental. III. Gomes, Júlio. IV. Título.

Bibliotecária: Roseny Rivelini Morciani CRB-9/1585



MINISTÉRIO DA EDUCAÇÃO
SETOR DE TECNOLOGIA
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO ENGENHARIA DE
RECURSOS HÍDRICOS E AMBIENTAL - 40001016021P0

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação ENGENHARIA DE RECURSOS HÍDRICOS E AMBIENTAL da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **PATRICIA CRISTINA STEFFEN** intitulada: **Modelagem hidrológica para previsão de vazões baseada em agrupamento de eventos de cheia**, sob orientação do Prof. Dr. JÚLIO GOMES, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutora está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 25 de Outubro de 2024.

Assinatura Eletrônica 04/11/2024 21:07:37.0 JÚLIO GOMES Presidente da Banca Examinadora

Assinatura Eletrônica 23/11/2024 07:44:38.0 ELOY KAVISKI Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ) Assinatura Eletrônica
04/11/2024 16:09:43.0
TALITA FERNANDA DAS GRAÇAS SILVA
Avaliador Externo (UNIVERSIDADE FEDERAL DE MINAS GERAIS)

Assinatura Eletrônica 30/10/2024 15:03:11.0 DANIEL HENRIQUE MARCO DETZEL Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica
01/11/2024 09:17:51.0
ADRIANO ROLIM DA PAZ
Avaliador Externo (UNIVERSIDADE FEDERAL DA PARAÍBA - UFPB)



AGRADECIMENTOS

A Deus, pelo discernimento na condução desta pesquisa e pelo dom da perseverança.

Ao meu marido, Luan, por seu incansável suporte emocional, e por sua paciência e compreensão nos momentos de minha ausência, permitindo que eu me dedicasse ao Doutorado e o concluísse com êxito. Obrigada por suas aulas de programação, por nossas discussões de melhoria nos códigos e por sua disposição em ajudar, sempre que precisei.

Aos meus pais, Wilson e Ana Lúcia, por terem abdicado de muitos sonhos para que eu pudesse realizar os meus, por permitirem que eu explorasse o mundo, apesar de suas infinitas preocupações, e por me proporcionarem a melhor educação que puderam.

À minha irmã, Ana Paula, por confiar no meu potencial, mais do que eu mesma, pelos momentos de descontração e pela disposição em ajudar. Ao meu cunhado, Eduardo, pelas conversas e por entender que eu e sua excelentíssima esposa passaremos horas conversando.

Ao Prof. Dr. Júlio Gomes, pela orientação, pelos anos que trabalhamos juntos, pela paciência, e por me incentivar a caminhar sozinha como pesquisadora.

Ao Prof. Dr. Eloy Kaviski, pela oportunidade de me aprofundar em Estatística Multivariada e por me auxiliar na compreensão de diversos assuntos relacionados, pelas contribuições durante os Exames de Qualificação e durante a Banca de Defesa, e por estar sempre disposto a esclarecer qualquer dúvida.

Ao Prof. Dr. Daniel Detzel, pelas grandes contribuições nos Exames de Qualificação e na Banca de Defesa, e por aceitar conduzir o grupo de estudos em *Machine Learning*, em 2019, permitindo que muitos alunos, inclusive eu, tivessem a oportunidade de conhecer e aplicar técnicas, até então, desconhecidas por nós.

Ao Prof. Dr. Adriano Rolim da Paz e à Profa. Dra. Talita Silva pelas excelentes contribuições na Banca de Defesa.

À Carolina Natel, pela iniciativa do grupo de *Machine Learning*, o qual foi o ponto de partida para aplicação de muitas técnicas desta pesquisa.

À minha amiga Ana Paula Muhlenhoff, pelo incentivo constante, por estar sempre disposta a ouvir e por compartilhar das mesmas angústias que apenas grandes feitos são capazes de gerar. Obrigada por sua amizade sincera.

À minha amiga Lediane Marcon, pela parceria durante os estudos de *Machine Learning*, pela companhia na sala de estudos do PPGERHA e pelas conversas que tornavam o dia a dia mais leve.

Às minhas amigas Liege, Luziadne e Laís, pela convivência na sala de estudos e pela amizade que cultivamos além da pós-graduação.

Aos professores das disciplinas cursadas durante o Doutorado, em especial àqueles que me incentivaram na mudança para Barra do Garças e permitiram a conclusão de suas disciplinas à distância, em uma época anterior à pandemia, quando aulas remotas ainda não faziam parte do cotidiano.

Aos colegas do PPGERHA com os quais pude compartilhar conhecimento, principalmente durante as disciplinas cursadas.

À Coordenação do PPGERHA, pela compreensão e confiança depositadas, permitindo que o Doutorado fosse conduzido à distância.

À Universidade Federal do Paraná, pela formação proporcionada, desde a Graduação.

À Universidade Federal de Mato Grosso, pela concessão do afastamento durante os últimos meses de Doutorado, e, em especial, à Profa. Dra. Loyse Tussolini, que, como chefia direta na UFMT, incentivou a solicitação de afastamento.



RESUMO

O processo de previsão de vazões contém incertezas inerentes que se propagam pelo sistema, tornando difícil a determinação dos mecanismos de formação de descargas e fazendo com que características distintas entre os eventos extremos sejam negligenciadas, apesar da existência de alternativas que buscam melhorar a acurácia dos resultados. Neste contexto, ferramentas da Inteligência Artificial (AI) vêm sendo utilizadas como opções aos modelos hidrológicos tradicionais, dando margem às aplicações de algoritmos baseados na lógica difusa e nas Máquinas de Vetores de Suporte. Contudo, o que se observa na literatura é que o enfoque das pesquisas, em geral, está na comparação de técnicas e na identificação das melhores ferramentas para as diversas aplicações hidrológicas, e não especificamente desenvolvimento de novas formas de modelagem. Portanto, o presente estudo propõe uma nova abordagem de modelagem hidrológica para previsão de vazões, que se baseia no agrupamento prévio de eventos, e que foi demonstrada por meio da aplicação na área de estudo que compreende a Estação Fluviométrica União da Vitória, em União da Vitória, Paraná, Brasil, a partir das vazões médias diárias de 1964 a 2015, subdivididas entre os períodos de ajuste (até 2004) e previsão (de 2005 a 2015). Para a obtenção de três grupos de cheia, aplicou-se o algoritmo fuzzy c-means (FCM), que evidenciou características similares entre eventos de um mesmo grupo. As cheias pertencentes a cada grupo deram origem às três séries de eventos por grupo, que, juntamente com a série completa de vazões e a série contendo todos os eventos de cheia, permitiram o treinamento de cinco modelos de previsão, por meio do algoritmo de Regressão por Vetores de Suporte (SVR). Os modelos individuais apresentaram melhor desempenho ao prever vazões de mesma categoria para as quais foram treinados, bem como, desempenhos similares para a previsão de todas as séries para um horizonte de previsão de um dia. A partir dos três modelos por grupo, foram criadas, ainda, quatro configurações de previsão: (i) média; (ii) por grupo de pertencimento; (iii) ponderada por distâncias Euclidianas dos eventos aos grupos; e (iv) ponderada por graus de pertencimento dos eventos aos grupos. As configurações apresentaram desempenhos distintos para o referido horizonte de previsão, com melhor performance, em geral, para a previsão média, ponderada pelas distâncias Euclidianas, ponderada pelos graus de pertencimento e por grupo, nesta ordem.

Palavras-chave: aprendizado de máquinas; inteligência artificial; modelos inteligentes orientados a dados; Regressão por Vetores de Suporte.

ABSTRACT

Streamflow forecasting process contains inherent uncertainties that propagate through the system, making it difficult to determine the mechanisms of discharge generation and leading to the potential oversight of distinct characteristics between extreme events. This occurs despite the existence of alternatives aimed at improving result accuracy. In this context, Artificial Intelligence (AI) tools have emerged as alternatives to traditional hydrological models, leading to the application of algorithms based on fuzzy logic and Support Vector Machines. However, a review of the literature shows that research generally focuses on comparing techniques and identifying the best tools for various hydrological applications, rather than specifically developing new modeling approaches. Therefore, this study proposes a new hydrological modeling approach for flow forecasting, based on the previous clustering of events. This approach was demonstrated through its application to the study area encompassing the União da Vitória Gauging Station in União da Vitória, Paraná State, Brazil. The study used mean daily discharge data from 1964 to 2015, divided into an adjustment period (up to 2004) and a forecast period (2005 to 2015). To obtain three flood clusters, the fuzzy c-means (FCM) algorithm was applied, revealing similar characteristics between events within the same group. The floods in each group resulted in three event series per group. These, along with the complete series of discharges and the series containing all flood events, enabled the training of five forecasting models using the Support Vector Regression (SVR) algorithm. The individual models performed best when predicting flows within the same category for which they were trained and showed similar performance when predicting all series one-day ahead. Based on the three models per group, four forecast configurations were created: (i) average; (ii) by group assignment; (iii) weighted by the Euclidean distances of events to the groups; and (iv) weighted by the degrees of belonging of events to the groups. These configurations performed differently for the one-day forecast horizon, with the average forecast generally performing best, followed by the forecasts weighted by Euclidean distances, degrees of belonging, and group assignment, in that order.

Keywords: machine learning; artificial intelligence; data-driven intelligent models; Support Vector Regression.

LISTA DE FIGURAS

FIGURA 2.1	- REPRESENTAÇÃO DE UM SISTEMA DIFUSO46
FIGURA 2.2	- REPRESENTAÇÃO CONCEITUAL DE MÁQUINA DE VETOR DE
	SUPORTE54
FIGURA 2.3	- HIPERPLANO IDEAL DA MÁQUINA DE VETORES DE SUPORTE
	QUE SEPARA DUAS CLASSES54
FIGURA 2.4	- ESTRUTURA DE REDE NEURAL DIRETA DE CAMADA OCULTA
	ÚNICA60
FIGURA 2.5	- ARQUITETURA BÁSICA DO MÉTODO EXTREME LEARNING
	MACHINE63
FIGURA 3.1	– LOCALIZAÇÃO DO MUNICÍPIO DE UNIÃO DA VITÓRIA, NO
	ESTADO DO PARANÁ E NA BACIA DO RIO IGUAÇU65
FIGURA 3.2	– CURVAS DE PERMANÊNCIA DAS ESTAÇÕES
	FLUVIOMÉTRICAS SÃO BENTO (65155000) E RIO DA VÁRZEA
	DOS LIMA (65135000)69
FIGURA 3.3	- LOCALIZAÇÃO DAS ESTAÇÕES FLUVIOMÉTRICAS73
FIGURA 3.4	- CURVA DE PERMANÊNCIA DA SÉRIE HISTÓRICA DE VAZÕES
	DA ESTAÇÃO UNIÃO DA VITÓRIA (65310000)74
FIGURA 3.5	– DISTRIBUIÇÃO TEMPORAL DOS EVENTOS SELECIONADOS 78
FIGURA 3.6	- ESTRATÉGIA DE AMOSTRAGEM DIVIDIDA90
FIGURA 3.7	– HIDROGRAMAS DOS EVENTOS OBSERVADOS E
	SUBDIVIDIDOS CONFORME ESTRATÉGIA DE AMOSTRAGEM –
	AMOSTRA 192
FIGURA 3.8	– HIDROGRAMAS DOS EVENTOS OBSERVADOS E
	SUBDIVIDIDOS CONFORME ESTRATÉGIA DE AMOSTRAGEM –
	AMOSTRA 292
FIGURA 3.9	– HIDROGRAMAS DOS EVENTOS OBSERVADOS E
	SUBDIVIDIDOS CONFORME ESTRATÉGIA DE AMOSTRAGEM –
	AMOSTRA 3
FIGURA 3.10	– HIDROGRAMAS DOS EVENTOS OBSERVADOS E
	SUBDIVIDIDOS CONFORME ESTRATÉGIA DE AMOSTRAGEM –
	AMOSTRA 4

FIGURA 3.11	_	HIDROGRAMAS DOS EVENTOS OBSERVADOS E
		SUBDIVIDIDOS CONFORME ESTRATÉGIA DE AMOSTRAGEM –
		AMOSTRA 5
FIGURA 3.12	_	HIDROGRAMAS DOS EVENTOS OBSERVADOS E
		SUBDIVIDIDOS CONFORME ESTRATÉGIA DE AMOSTRAGEM –
		AMOSTRA 6
FIGURA 3.13	_	HIDROGRAMAS DOS EVENTOS OBSERVADOS E
		SUBDIVIDIDOS CONFORME ESTRATÉGIA DE AMOSTRAGEM –
		AMOSTRA 7
FIGURA 3.14	_	HIDROGRAMAS DOS EVENTOS OBSERVADOS E
		SUBDIVIDIDOS CONFORME ESTRATÉGIA DE AMOSTRAGEM –
		AMOSTRA 894
FIGURA 3.15	_	PROCESSOS DE TREINAMENTO E VALIDAÇÃO95
FIGURA 4.1	_	DISTRIBUIÇÃO TEMPORAL DOS EVENTOS CLASSIFICADOS
		NOS QUATRO GRUPOS
FIGURA 4.2	_	MÁXIMOS GRAUS DE SIMILARIDADE DOS EVENTOS COM OS
		GRUPOS
FIGURA 4.3	_	PERTENCIMENTO DOS EVENTOS DO GRUPO 1 AOS DEMAIS
		GRUPOS
FIGURA 4.4	_	PERTENCIMENTO DOS EVENTOS DO GRUPO 2 AOS DEMAIS
		GRUPOS. 105
FIGURA 4.5	_	PERTENCIMENTO DOS EVENTOS DO GRUPO 3 AOS DEMAIS
		GRUPOS
FIGURA 4.6	_	VARIAÇÃO DAS COORDENADAS NORMALIZADAS DOS
		CENTROIDES DOS GRUPOS
FIGURA 4.7	_	RESULTADOS DE KG, NS, NS MODIFICADO E MARE PARA
		DIFERENTES CONJUNTOS DE PARÂMETROS E MODELOS -
		MÉDIA ENTRE AS AMOSTRAS – ETAPA DE TREINAMENTO 112
FIGURA 4.8	_	RESULTADOS DE KG, NS, NS MODIFICADO E MARE PARA
		DIFERENTES CONJUNTOS DE PARÂMETROS E MODELOS -
		MÉDIA ENTRE AS AMOSTRAS – ETAPA DE TESTE121
FIGURA 4.9	_	HIDROGRAMAS PREVISTO E OBSERVADO DA SÉRIE COM
		TODOS OS EVENTOS – PREVISÃO MÉDIA – 2005 A 2015 134

FIGURA 4.10	– HIDROGRAMAS PREVISTO E OBSERVADO DA SÉRIE COM
	TODOS OS EVENTOS – PREVISÃO POR GRUPO D
	PERTENCIMENTO – 2005 A 201513
FIGURA 4.11	– HIDROGRAMAS PREVISTO E OBSERVADO DA SÉRIE COM
	TODOS OS EVENTOS – PREVISÃO PONDERADA PO
	DISTÂNCIAS EUCLIDIANAS – 2005 A 2015
FIGURA 4.12	– HIDROGRAMAS PREVISTO E OBSERVADO DA SÉRIE COM
	TODOS OS EVENTOS – PREVISÃO PONDERADA POR GRAUS D
	PERTENCIMENTO – 2005 A 2015
FIGURA 4.13	- COMPORTAMENTO DOS COEFICIENTES KG, NS E MARE NA
	PREVISÕES – 2005 A 2015
FIGURA 4.14	- COMPORTAMENTO DO COEFICIENTE NS MODIFICADO NA
	PREVISÕES – 2005 A 2015
FIGURA 4.15	 ERROS RELATIVOS ENTRE AS ESTATÍSTICAS DAS VAZÕE
	OBSERVADAS E PREVISTAS – 2005 A 2015
FIGURA 4.16	– CURVAS DE PERMANÊNCIA DAS VAZÕES OBSERVADA
	PREVISTAS – 2005 A 2015
FIGURA 4.17	- ERROS RELATIVOS DE VAZÕES DE DIFERENTES TEMPOS D
	PERMANÊNCIA – 2005 A 2015
FIGURA 4.18	- HIDROGRAMAS OBSERVADO E PREVISTOS DO EVENTO ID 2
FIGURA 4.19	- HIDROGRAMAS OBSERVADO E PREVISTOS DO EVENTO ID 3
FIGURA 4.20	- HIDROGRAMAS OBSERVADO E PREVISTOS DO EVENTO ID 1
FIGURA 4.21	- HIDROGRAMAS OBSERVADO E PREVISTOS DO EVENTO ID
FIGURA 4.22	- HIDROGRAMAS OBSERVADO E PREVISTOS DO EVENTO ID 3
FIGURA 4.23	 HIDROGRAMAS OBSERVADO E PREVISTOS DO EVENTO ID 2
	14
FIGURA 4.24	- HIDROGRAMAS OBSERVADO E PREVISTOS DO EVENTO ID 4
	14

FIGURA 4.25	-	HIDROGRAMAS OBSERVADO E PREVISTOS DO EVENTO ID 15
FIGURA B.1	-	EVOLUÇÃO DO COEFICIENTE KG PARA AS AMOSTRAS E
		CONJUNTOS DE PARÂMETROS DO MODELO DE SÉRIE
		COMPLETA – FASE DE TREINAMENTO 167
FIGURA B.2	-	EVOLUÇÃO DO COEFICIENTE NS PARA AS AMOSTRAS E
		CONJUNTOS DE PARÂMETROS DO MODELO DE SÉRIE
		COMPLETA – FASE DE TREINAMENTO
FIGURA B.3	-	EVOLUÇÃO DO COEFICIENTE NS MODIFICADO PARA AS
		AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DE
		SÉRIE COMPLETA – FASE DE TREINAMENTO 168
FIGURA B.4	_	EVOLUÇÃO DO COEFICIENTE MARE PARA AS AMOSTRAS E
		CONJUNTOS DE PARÂMETROS DO MODELO DE SÉRIE
		COMPLETA – FASE DE TREINAMENTO
FIGURA B.5	_	EVOLUÇÃO DO COEFICIENTE KG PARA AS AMOSTRAS E
		CONJUNTOS DE PARÂMETROS DO MODELO DE TODOS OS
		EVENTOS – FASE DE TREINAMENTO169
FIGURA B.6	_	EVOLUÇÃO DO COEFICIENTE NS PARA AS AMOSTRAS E
		CONJUNTOS DE PARÂMETROS DO MODELO DE TODOS OS
		EVENTOS – FASE DE TREINAMENTO169
FIGURA B.7	_	EVOLUÇÃO DO COEFICIENTE NS MODIFICADO PARA AS
		AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DE
		TODOS OS EVENTOS – FASE DE TREINAMENTO 170
FIGURA B.8	_	EVOLUÇÃO DO COEFICIENTE MARE PARA AS AMOSTRAS E
		CONJUNTOS DE PARÂMETROS DO MODELO DE TODOS OS
		EVENTOS – FASE DE TREINAMENTO170
FIGURA B.9	_	EVOLUÇÃO DO COEFICIENTE KG PARA AS AMOSTRAS E
		CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 1 –
		FASE DE TREINAMENTO
FIGURA B.10	_	EVOLUÇÃO DO COEFICIENTE NS PARA AS AMOSTRAS E
		CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 1 –
		FASE DE TREINAMENTO

FIGURA B.11	_	EVOLUÇÃO DO COEFICIENTE NS MODIFICADO PARA	AS
		AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO	DO
		GRUPO 1 – FASE DE TREINAMENTO	172
FIGURA B.12	_	EVOLUÇÃO DO COEFICIENTE MARE PARA AS AMOSTRA	S E
		CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO	1 –
		FASE DE TREINAMENTO	172
FIGURA B.13	_	EVOLUÇÃO DO COEFICIENTE KG PARA AS AMOSTRAS	S E
		CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO	2 –
		FASE DE TREINAMENTO	173
FIGURA B.14	_	EVOLUÇÃO DO COEFICIENTE NS PARA AS AMOSTRAS	S E
		CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO	2 –
		FASE DE TREINAMENTO	173
FIGURA B.15	_	EVOLUÇÃO DO COEFICIENTE NS MODIFICADO PARA	AS
		AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO	DO
		GRUPO 2 – FASE DE TREINAMENTO	174
FIGURA B.16	_	EVOLUÇÃO DO COEFICIENTE MARE PARA AS AMOSTRA	S E
		CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO	2 –
		FASE DE TREINAMENTO	174
FIGURA B.17	_	EVOLUÇÃO DO COEFICIENTE KG PARA AS AMOSTRAS	S E
		CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO	3 –
		FASE DE TREINAMENTO	175
FIGURA B.18	-	EVOLUÇÃO DO COEFICIENTE NS PARA AS AMOSTRAS	S E
		CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO	3 –
		FASE DE TREINAMENTO	
FIGURA B.19	-	EVOLUÇÃO DO COEFICIENTE NS MODIFICADO PARA	AS
		AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO	
		GRUPO 3 – FASE DE TREINAMENTO	176
FIGURA B.20	-	EVOLUÇÃO DO COEFICIENTE MARE PARA AS AMOSTRA	S E
		CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO	3 –
		FASE DE TREINAMENTO	176
FIGURA C.1	-	DISTRIBUIÇÃO DOS COEFICIENTES KG, NS, NS MODIFICA	DO
		E MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETE	
		DO MODELO DE SÉRIE COMPLETA – FASE DE TREINAMEN	VTO
			177

FIGURA C.2	– DISTRIBUIÇÃO DOS COEFICIENTES KG, NS, NS MODIFICADO
	E MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS
	DO MODELO DE TODOS OS EVENTOS – FASE DE
	TREINAMENTO178
FIGURA C.3	– DISTRIBUIÇÃO DOS COEFICIENTES KG, NS, NS MODIFICADO
	E MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS
	DO MODELO DO GRUPO 1 – FASE DE TREINAMENTO 179
FIGURA C.4	– DISTRIBUIÇÃO DOS COEFICIENTES KG, NS, NS MODIFICADO
	E MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS
	DO MODELO DO GRUPO 2 – FASE DE TREINAMENTO 180
FIGURA C.5	– DISTRIBUIÇÃO DOS COEFICIENTES KG, NS, NS MODIFICADO
	E MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS
	DO MODELO DO GRUPO 3 – FASE DE TREINAMENTO 181
FIGURA D.1	– DISTRIBUIÇÃO DOS COEFICIENTES KG, NS, NS MODIFICADO
	E MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS
	DO MODELO DE SÉRIE COMPLETA – FASE DE TESTE182
FIGURA D.2	– DISTRIBUIÇÃO DOS COEFICIENTES KG, NS, NS MODIFICADO
	E MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS
	DO MODELO DE TODOS OS EVENTOS – FASE DE TESTE 183
FIGURA D.3	 DISTRIBUIÇÃO DOS COEFICIENTES KG, NS, NS MODIFICADO
	E MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS
	DO MODELO DO GRUPO 1 – FASE DE TESTE184
FIGURA D.4	 DISTRIBUIÇÃO DOS COEFICIENTES KG, NS, NS MODIFICADO
	E MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS
	DO MODELO DO GRUPO 2 – FASE DE TESTE185
FIGURA D.5	 DISTRIBUIÇÃO DOS COEFICIENTES KG, NS, NS MODIFICADO
	E MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS
	DO MODELO GRUPO 3 – FASE DE TESTE186
FIGURA E.1	– EVOLUÇÃO DO COEFICIENTE KG PARA AS AMOSTRAS E
	CONJUNTOS DE PARÂMETROS DO MODELO DE SÉRIE
	COMPLETA – FASE DE TESTE187
FIGURA E.2	– EVOLUÇÃO DO COEFICIENTE NS PARA AS AMOSTRAS E
	CONJUNTOS DE PARÂMETROS DO MODELO DE SÉRIE
	COMPLETA – FASE DE TESTE187

FIGURA E.3	_	EVOLUÇÃO DO COEFICIENTE NS MODIFICADO PARA AS
		AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DE
		SÉRIE COMPLETA – FASE DE TESTE
FIGURA E.4	_	EVOLUÇÃO DO COEFICIENTE MARE PARA AS AMOSTRAS E
		CONJUNTOS DE PARÂMETROS DO MODELO DE SÉRIE
		COMPLETA – FASE DE TESTE
FIGURA E.5	_	EVOLUÇÃO DO COEFICIENTE KG PARA AS AMOSTRAS E
		CONJUNTOS DE PARÂMETROS DO MODELO DE TODOS OS
		EVENTOS – FASE DE TESTE
FIGURA E.6	_	EVOLUÇÃO DO COEFICIENTE NS PARA AS AMOSTRAS E
		CONJUNTOS DE PARÂMETROS DO MODELO DE TODOS OS
		EVENTOS – FASE DE TESTE
FIGURA E.7	-	EVOLUÇÃO DO COEFICIENTE NS MODIFICADO PARA AS
		AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DE
		TODOS OS EVENTOS – FASE DE TESTE
FIGURA E.8	_	EVOLUÇÃO DO COEFICIENTE MARE PARA AS AMOSTRAS E
		CONJUNTOS DE PARÂMETROS DO MODELO DE TODOS OS
		EVENTOS – FASE DE TESTE
FIGURA E.9	_	EVOLUÇÃO DO COEFICIENTE KG PARA AS AMOSTRAS E
		CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 1 –
		FASE DE TESTE
FIGURA E.10	-	EVOLUÇÃO DO COEFICIENTE NS PARA AS AMOSTRAS E
		CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 1 –
		FASE DE TESTE
FIGURA E.11	-	EVOLUÇÃO DO COEFICIENTE NS MODIFICADO PARA AS
		AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO
		GRUPO 1 – FASE DE TESTE
FIGURA E.12	-	EVOLUÇÃO DO COEFICIENTE MARE PARA AS AMOSTRAS E
		CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 1 –
		FASE DE TESTE
FIGURA E.13	-	EVOLUÇÃO DO COEFICIENTE KG PARA AS AMOSTRAS E
		CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 2 –
		FASE DE TESTE

FIGURA E.14	_	EVOLUÇÃO DO COEFICIENTE NS PARA AS AMOSTRAS	E
		CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 2	_
		FASE DE TESTE	3
FIGURA E.15	_	EVOLUÇÃO DO COEFICIENTE NS MODIFICADO PARA A	S
		AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO	С
		GRUPO 2 – FASE DE TESTE	4
FIGURA E.16	_	EVOLUÇÃO DO COEFICIENTE MARE PARA AS AMOSTRAS	E
		CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 2	_
		FASE DE TESTE	4
FIGURA E.17	_	EVOLUÇÃO DO COEFICIENTE KG PARA AS AMOSTRAS	E
		CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 3	_
		FASE DE TESTE	5
FIGURA E.18	_	EVOLUÇÃO DO COEFICIENTE NS PARA AS AMOSTRAS	E
		CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 3	_
		FASE DE TESTE	5
FIGURA E.19	_	EVOLUÇÃO DO COEFICIENTE NS MODIFICADO PARA A	S
		AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO	С
		GRUPO 3 – FASE DE TESTE	6
FIGURA E.20	_	EVOLUÇÃO DO COEFICIENTE MARE PARA AS AMOSTRAS	E
		CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 3	_
		FASE DE TESTE	6

LISTA DE TABELAS

TABELA 2.1	 PESQUISAS COM APLICAÇÕES NA HIDROLOGIA DE MODELOS
	INTELIGENTES ORIENTADOS A DADOS49
TABELA 3.1	– PRINCIPAIS ESTAÇÕES FLUVIOMÉTRICAS IDENTIFICADAS A
	MONTANTE DE UNIÃO DA VITÓRIA68
TABELA 3.2	– ESTATÍSTICA DESCRITIVA DAS ESTAÇÕES FLUVIOMÉTRICAS
	ENCONTRADAS69
TABELA 3.3	– CORRELAÇÕES ENTRE AS SÉRIES DE VAZÕES MÉDIAS
	DIÁRIAS DE UNIÃO DA VITÓRIA (65310000) E AS DAS ESTAÇÕES
	A MONTANTE, PARA DIFERENTES <i>LAGS</i> DE TEMPO70
TABELA 3.4	– AUTOVALORES E AUTOVETORES DA ANÁLISE DOS
	COMPONENTES PRINCIPAIS DAS ESTAÇÕES72
TABELA 3.5	– VARIÂNCIA EXPLICADA DAS ESTAÇÕES, EM ORDEM
	DECRESCENTE72
TABELA 3.6	– ESTATÍSTICA DESCRITIVA DAS VAZÕES MÉDIAS DIÁRIAS NA
	ESTAÇÃO FLUVIOMÉTRICA UNIÃO DA VITÓRIA (65310000) -
	1964 A 201574
TABELA 3.7	– PARÂMETROS MULTIPLICADORES, LIMITES INFERIORES DE
	VAZÃO E TEMPOS DE PERMANÊNCIA75
TABELA 3.8	– CARACTERÍSTICAS DOS EVENTOS PARA OS DIFERENTES
	LIMITES MÍNIMOS DE VAZÃO76
TABELA 3.9	– FREQUÊNCIA ABSOLUTA DOS PICOS DE EVENTOS PARA
	DIFERENTES LIMITES MÍNIMOS DE VAZÕES77
TABELA 3.10	– FREQUÊNCIA DOS PICOS DE EVENTOS PARA DIFERENTES
	INTERVALOS DE VAZÕES78
TABELA 3.11	– ESTATÍSTICAS DA SÉRIE COMPLETA OBSERVADA E
	SUBDIVIDIDA CONFORME ESTRATÉGIA DE AMOSTRAGEM -
	PERÍODOS DE TREINAMENTO91
TABELA 3.12	– ESTATÍSTICAS DA SÉRIE COMPLETA OBSERVADA E
	SUBDIVIDIDA CONFORME ESTRATÉGIA DE AMOSTRAGEM -
	PERÍODOS DE TESTE 91

TABELA 3.13	-	ESTATÍSTICAS DA SÉRIE OBSERVADA DE TODOS OS EVENTOS
		E SUBDIVIDIDA CONFORME ESTRATÉGIA DE AMOSTRAGEM –
		PERÍODOS DE TREINAMENTO91
TABELA 3.14	_	ESTATÍSTICAS DA SÉRIE OBSERVADA DE TODOS OS EVENTOS
		E SUBDIVIDIDA CONFORME ESTRATÉGIA DE AMOSTRAGEM –
		PERÍODOS DE TESTE92
TABELA 4.1	_	FREQUÊNCIAS DE INTERVALOS DE MÁXIMOS GRAUS DE
		SIMILARIDADE, POR SÉRIE DE EVENTOS103
TABELA 4.2	_	PARÂMETROS ESTATÍSTICOS DOS MÁXIMOS GRAUS DE
		SIMILARIDADE
TABELA 4.3	_	VALORES BRUTOS DAS COORDENADAS DOS CENTROIDES DOS
		GRUPOS
TABELA 4.4	_	PARÂMETROS ESTATÍSTICOS DOS EVENTOS DE TODO O
		PERÍODO DE DADOS – 1964 A 2015107
TABELA 4.5	_	PARÂMETROS ESTATÍSTICOS DOS EVENTOS DO PERÍODO DE
		TREINAMENTO E TESTE – 1964 A 2004
TABELA 4.6	_	PARÂMETROS ESTATÍSTICOS DOS EVENTOS DO PERÍODO DE
		PREVISÃO – 2005 A 2015
TABELA 4.7	_	RELAÇÕES ENTRE OS PARÂMETROS ESTATÍSTICOS DOS
		PERÍODOS DE TREINAMENTO E TESTE, PARA A ESTRATÉGIA
		DE AMOSTRAGEM – TODOS OS EVENTOS108
TABELA 4.8	_	RELAÇÕES ENTRE OS PARÂMETROS ESTATÍSTICOS DOS
		PERÍODOS DE TREINAMENTO E TESTE, PARA A ESTRATÉGIA
		DE AMOSTRAGEM – GRUPO 1109
TABELA 4.9	-	RELAÇÕES ENTRE OS PARÂMETROS ESTATÍSTICOS DOS
		PERÍODOS DE TREINAMENTO E TESTE, PARA A ESTRATÉGIA
		DE AMOSTRAGEM – GRUPO 2109
TABELA 4.10	_	RELAÇÕES ENTRE OS PARÂMETROS ESTATÍSTICOS DOS
		PERÍODOS DE TREINAMENTO E TESTE, PARA A ESTRATÉGIA
		DE AMOSTRAGEM – GRUPO 3110
TABELA 4.11	-	AVALIAÇÃO MÉDIA DO COMPORTAMENTO DE KG PARA
		DIFERENTES CONJUNTOS DE PARÂMETROS – FASE DE
		TREINAMENTO

TABELA 4.12	– AVALIAÇÃO MÉDIA DO COMPORTAMENTO DE NS PARA
	DIFERENTES CONJUNTOS DE PARÂMETROS – FASE DE
	TREINAMENTO
TABELA 4.13	– AVALIAÇÃO MÉDIA DO COMPORTAMENTO DE NS
	MODIFICADO PARA DIFERENTES CONJUNTOS DE
	PARÂMETROS – FASE DE TREINAMENTO113
TABELA 4.14	– AVALIAÇÃO MÉDIA DO COMPORTAMENTO DE MARE PARA
	DIFERENTES CONJUNTOS DE PARÂMETROS – FASE DE
	TREINAMENTO
TABELA 4.15	- RANQUEAMENTO DAS MAIORES E MENORES VAZÕES DE PICO
	DO GRUPO 1, EM UNIÃO DA VITÓRIA115
TABELA 4.16	– ESTRATÉGIA DE AMOSTRAGEM DOS EVENTOS DO GRUPO 1 –
	FASE DE TREINAMENTO
TABELA 4.17	– FREQUÊNCIAS ABSOLUTAS (À ESQUERDA) E RELATIVAS (À
	DIREITA) DOS MAIORES E MENORES PICOS DO GRUPO 1 – FASE
	DE TREINAMENTO116
TABELA 4.18	- RANQUEAMENTO DAS MAIORES E MENORES VAZÕES DE PICO
	DO GRUPO 2, EM UNIÃO DA VITÓRIA117
TABELA 4.19	– ESTRATÉGIA DE AMOSTRAGEM DOS EVENTOS DO GRUPO 2 –
	FASE DE TREINAMENTO
TABELA 4.20	– FREQUÊNCIAS ABSOLUTAS (À ESQUERDA) E RELATIVAS (À
	DIREITA) DOS MAIORES E MENORES PICOS DO GRUPO 2119
TABELA 4.21	– RANQUEAMENTO DAS VAZÕES DE PICO DO GRUPO 3, EM
	UNIÃO DA VITÓRIA119
TABELA 4.22	– ESTRATÉGIA DE AMOSTRAGEM DOS EVENTOS DO GRUPO 3 –
	FASE DE TREINAMENTO
TABELA 4.23	– AVALIAÇÃO MÉDIA DO COMPORTAMENTO DE KG, PARA
	DIFERENTES CONJUNTOS DE PARÂMETROS – FASE DE TESTE
TABELA 4.24	– AVALIAÇÃO MÉDIA DO COMPORTAMENTO DE NS PARA
	DIFERENTES CONJUNTOS DE PARÂMETROS – FASE DE TESTE

TABELA 4.25	_	AVALIAÇÃO MÉDIA DO COMPORTAMENTO DE NS
		MODIFICADO PARA DIFERENTES CONJUNTOS DE
		PARÂMETROS – FASE DE TESTE
TABELA 4.26	_	AVALIAÇÃO MÉDIA DO COMPORTAMENTO DE MARE PARA
		DIFERENTES CONJUNTOS DE PARÂMETROS – FASE DE TESTE
TABELA 4.27	_	ESTRATÉGIA DE AMOSTRAGEM DOS EVENTOS DO GRUPO 1 –
		FASE DE TESTE
TABELA 4.28	_	FREQUÊNCIAS ABSOLUTAS (À ESQUERDA) E RELATIVAS (À
		DIREITA) DOS MAIORES E MENORES PICOS DO GRUPO 1 – FASE
		DE TESTE 124
TABELA 4.29	_	ESTRATÉGIA DE AMOSTRAGEM DOS EVENTOS DO GRUPO 2 –
		FASE DE TESTE
TABELA 4.30	_	FREQUÊNCIAS ABSOLUTAS (À ESQUERDA) E RELATIVAS (À
		DIREITA) DOS MAIORES E MENORES PICOS DO GRUPO 2 – FASE
		DE TESTE
TABELA 4.31	_	ESTRATÉGIA DE AMOSTRAGEM DOS EVENTOS DO GRUPO 3 –
		FASE DE TESTE
TABELA 4.32	_	COEFICIENTES KG, NS, NS MODIFICADO E MARE DAS
		SIMULAÇÕES DAS SÉRIES – 1964 A 2004
TABELA 4.33	_	ERROS RELATIVOS ENTRE OS PARÂMETROS ESTATÍSTICOS
		DAS SÉRIES SIMULADAS E OBSERVADAS – 1964 A 2004 128
TABELA 4.34	_	ERROS RELATIVOS DOS PERCENTIS DE VAZÕES DAS SÉRIES
		SIMULADAS – 1964 A 2004
TABELA 4.35	_	COEFICIENTES KG, NS, NS MODIFICADO E MARE DAS
		PREVISÕES DAS SÉRIES – 2005 A 2015
TABELA 4.36	_	ERROS RELATIVOS ENTRE OS PARÂMETROS ESTATÍSTICOS
		DAS SÉRIES SIMULADAS E OBSERVADAS – 2005 A 2015 131
TABELA 4.37	-	ERROS RELATIVOS DOS PERCENTIS DE VAZÕES DAS SÉRIES
		SIMULADAS – 2005 A 2015
TABELA 4.38	-	MEDIDAS DE QUALIDADE DA PREVISÃO MÉDIA – 2005 A 2015
		PARÂMETROS ESTATÍSTICOS DAS SÉRIES OBSERVADA E
		PREVISTAS PELA MÉDIA _ 2005 A 2015 133

TABELA 4.40	_	MEDIDAS DE QUALIDADE DA PREVISÃO POR GRUPO DE
		PERTENCIMENTO – 2005 A 2015
TABELA 4.41	_	PARÂMETROS ESTATÍSTICOS DAS SÉRIES DE EVENTOS
		OBSERVADA E PREVISTAS POR GRUPO DE PERTENCIMENTO –
		2005 A 2015
TABELA 4.42	_	MEDIDAS DE QUALIDADE DA PREVISÃO PONDERADA POR
		DISTÂNCIAS EUCLIDIANAS – 2005 A 2015135
TABELA 4.43	_	PARÂMETROS ESTATÍSTICOS DAS SÉRIES DE EVENTOS
		OBSERVADA E PREVISTAS PELA PONDERAÇÃO POR
		DISTÂNCIAS EUCLIDIANAS – 2005 A 2015136
TABELA 4.44	_	MEDIDAS DE QUALIDADE DA PREVISÃO PONDERADA POR
		GRAUS DE PERTENCIMENTO – 2005 A 2015
TABELA 4.45	_	PARÂMETROS ESTATÍSTICOS DAS SÉRIES OBSERVADA E
		PREVISTAS PELA PONDERADA POR GRAUS DE
		PERTENCIMENTO – 2005 A 2015
TABELA 4.46	_	INTERVALOS DE RESULTADOS DO COEFICIENTE KG PARA AS
		PREVISÕES DE EVENTOS DE 2005 A 2015
TABELA 4.47	-	INTERVALOS DE RESULTADOS DO COEFICIENTE NS PARA AS
		PREVISÕES DE EVENTOS DE 2005 A 2015142
TABELA 4.48	-	INTERVALOS DE RESULTADOS DO COEFICIENTE NS
		MODIFICADO PARA AS PREVISÕES DE EVENTOS DE 2005 A 2015
TABELA 4.49	_	INTERVALOS DE RESULTADOS DO COEFICIENTE MARE PARA
EADEL 4 4 50		AS PREVISÕES DE EVENTOS DE 2005 A 2015
TABELA 4.50	_	DISTRIBUIÇÃO DOS ERROS RELATIVOS DAS VAZÕES
TABLE A 4.51		MÁXIMAS PREVISTAS – 2005 A 2015
TABELA 4.51	_	DISTRIBUIÇÃO DOS ERROS RELATIVOS DAS VAZÕES MÉDIAS
TADELA 452		PREVISTAS – 2005 A 2015
TABELA 4.52	_	DISTRIBUIÇÃO DOS ERROS RELATIVOS DOS DESVIOS-PADRÃO DAS VAZÕES PREVISTAS – 2005 A 2015
TADELA 4 52		DISTRIBUIÇÃO DOS ERROS RELATIVOS DOS VOLUMES
IADELA 4.53	_	ESCOADOS, COM BASE NAS VAZÕES PREVISTAS – 2005 A 2015
TARELA A 1	_	AGRUPAMENTO DE EVENTOS
		110110111111111111111111111111111111111

TABELA F.1	- VALORES DE KG PARA OS EVENTOS PREVISTOS - 2005 A 2015
TABELA F.2	- VALORES DE NS PARA OS EVENTOS PREVISTOS - 2005 A 2015
TABELA F.3	- VALORES DE NS MODIFICADO PARA OS EVENTOS PREVISTOS
	– 2005 A 2015199
TABELA F.4	- VALORES DE MARE PARA OS EVENTOS PREVISTOS - 2005 A
	2015200
TABELA G.1	– ERROS RELATIVOS DAS VAZÕES MÍNIMAS PREVISTAS DOS
	EVENTOS – 2005 A 2015201
TABELA G.2	- ERROS RELATIVOS DAS VAZÕES MÁXIMAS PREVISTAS DOS
	EVENTOS – 2005 A 2015202
TABELA G.3	– ERROS RELATIVOS DAS VAZÕES MÉDIAS PREVISTAS DOS
	EVENTOS – 2005 A 2015203
TABELA G.4	– ERROS RELATIVOS DOS DESVIOS-PADRÃO DAS VAZÕES
	PREVISTAS DOS EVENTOS – 2005 A 2015204
TABELA G.5	– ERROS RELATIVOS DAS ASSIMETRIAS DAS VAZÕES
	PREVISTAS DOS EVENTOS – 2005 A 2015205
TABELA G.6	– ERROS RELATIVOS DOS VOLUMES PREVISTOS DOS EVENTOS –
	2005 A 2015206

LISTA DE ABREVIATURAS E SIGLAS

AI — Inteligência Artificial (Artificial Intelligence)

ANA – Agência Nacional de Águas

ANFIS – Adaptive Neuro-Fuzzy Inference System

ANN – Redes Neurais Artificiais (Artificial Neural Networks)

AR – Modelo Autorregressivo

ARMA – Modelo Autorregressivo de Médias Móveis

ARIMA – Modelo Autorregressivo Integrado de Médias Móveis

BPNN – Back Propagation Neural Networks

CI – Inteligência Computacional ((Computational Intelligence)

CNN – Convolutional Neural Network

DM – Mineração de Dados (*Data Mining*)

EC – Evolutionary Computational
ELM – Extreme LearningMachine

EM-DAT — Emergency Disasters Data Base

FCM – fuzzy c-means

FFNN – feed-forward neural networks

FL – Fluviópolis

GA – Algoritmos Genéticos (Genetic Algorithms)

IBGE – Instituto Brasileiro de Geografia e Estatística

IAT – Instituto Água e Terra

IDA – Análise Inteligente de Dados (*Intelligent Data Analysis*)

KDD – Descoberta do Conhecimento em Banco de Dados (Knowledge Discovery in

Databases)

KG – Eficiência de Kling-Gupta

K-NN – *K-Nearest Neighbour*

LSTM – Long Short-Term Memory

MA – Modelo de Médias Móveis

MARE – Erro Relativo Absoluto Médio

MARS – Multivariate Adaptive Regression Splines

MGGP – Multi-Gene Genetic Programming

ML – Aprendizado de Máquinas (*Machine Learning*)

MLP – Multilayer Perceptron

MSE – Erro Quadrado Médio

NS – Eficiência de Nash-Sutcliffe

NSM – Eficiência de Nash-Sutcliffe Modificado

PSO – Particle Swarm Optimization

RBF — Radial Basis Function

RF – Random Forest

RN – Rio Negro

RNN – Recurrent Neural Network

SB – São Bento

SC – Computação Flexível (Soft Computing)

SEMA – Secretaria de Estado do Meio Ambiente e Recursos Hídricos

SLFN – Rede Neural de Camada Oculta Única

SOM – Self Organizing Map

SVM – Máquinas de Vetores de Suporte (Support Vector Machines)

SVR — Regressão por Vetores de Suporte (Support Vector Regression)

UV – União da Vitória

LISTA DE SÍMBOLOS

b — deslocamento que permite aumentar a margem de separação

 $f(\mathbf{x})$ – hiperplano separador, de dimensão (p)

 $k(\mathbf{x}_i, \mathbf{x}_i)$ — função kernel e seus pontos recebidos no espaço de entradas

m – tamanho da série de dados

w – vetor p-dimensional, perpendicular ao hiperplano separador

 w_i – vetor aleatório de pesos que conecta o j-ésimo neurônio da camada oculta aos

neurônios da camada de entrada:

X – vetor multidimensional de entrada dos modelos

x – observação do vetor multidimensional de entrada dos modelos

x – representação vetorial de x no espaço de características

Y – vetor de saída dos modelos

y – observação do vetor de saída dos modelos

ε – tolerância

 ξ_i – variáveis de folga

Φ – produto escalar

 ξ_i^* - variável de folga que especifica o menor erro de treinamento, sujeito à

tolerância

 α_i^* , α_i — multiplicadores de Lagrange

L – número de neurônios na camada oculta

β – vetor de pesos entre as camadas oculta e de saída

b_i – tendência aleatória do *j*-ésimo neurônio da camada oculta

 $h(x_i)$ – vetor de saída da camada oculta relacionado ao vetor de entrada X

H – matriz de saída da camada oculta

 Q_{ij} – vazão tratada da *i*-ésima observação da *j*-ésima estação

 Q^{θ}_{ii} – vazão não tratada da *i*-ésima observação da *j*-ésima estação

 $Q^0_{j,m\acute{a}x}$ — máxima vazão não tratada da j-ésima estação $Q^0_{j,m\acute{a}n}$ — mínima vazão não tratada da j-ésima estação

Qp - vazão de pico em União da Vitória
 tp - tempo de pico em União da Vitória
 Oto - vazão inicial em União da Vitória

 $(\mathbf{Q_{tp-1}})_{FL}$ – vazão em Fluviópolis defasada em um dia ($lag\ I$) do pico em União da Vitória

 $(\mathbf{Q}_{\mathsf{tp-1}})_{\mathsf{SB}}$ - vazão em São Bento defasada em um dia $(lag\ l)$ do pico em União da Vitória

(Q_{tp-2})_{RN} – vazão em Rio Negro defasada em dois dias (*lag 2*) do pico em União da Vitória

Q70 - vazão de 70% de permanência
Q35 - vazão de 35% de permanência
Q20 - vazão de 20% de permanência
Q5 - vazão de 5% de permanência

 σ_o — variância das vazões observadas

n – tamanho da amostra

 Q_{o_i} — vazões observadas

 Q_{s_i} – vazões simuladas

 $\overline{Q_o}$ — vazão média observada

 r_{KG} — coeficiente de correlação

 α_{KG} — variabilidade da modelagem

 β_{KG} — tendência da modelagem

 σ_{s} — desvio-padrão simulado

 σ_o — desvio-padrão observado

 μ_s — média simulada

 μ_o — média observada

 ε_r – erro relativo, expresso em percentual

 P_e — parâmetro estimado

 P_o — parâmetro observado

C – parâmetro de regularização

ν – limite superior para a fração de erros de treinamento e um limite inferior para

a fração de vetores de suporte

γ – coeficiente *kernel*

K – número de amostras da estratégia de amostragem

X – matriz de indicadores do agrupamento

V – matriz de centroides dos grupos

U – matriz de partição difusa

D – matriz de distâncias Euclidianas

 μ_{ki} — grau de similaridade da *i*-ésima observação com o *k*-ésimo grupo; elemento da

matriz de partição difusa (U)

- distância Euclidiana da i-ésima observação ao k-ésimo grupo; elemento da d_{ik} matriz de distâncias Euclidianas (D)

– distância Euclidiana da t-ésima observação ao k-ésimo grupo; elemento da d_{tk} matriz de distâncias Euclidianas (**D**)

número total de grupos С

– parâmetro de incerteza

-i-ésima observação da amostra (i = 1, 2, ..., n)

- número total de observações da amostra n

- valor tratado da i-ésima observação do j-ésimo indicador x_{ij}

- coordenada do k-ésimo centroide, correspondente ao j-ésimo indicador v_{ki}

d - dimensões de X; número total de indicadores de cheia para cada classe; número de características de X

- função objetivo do algoritmo FCM

- erro máximo tolerado

J

 $(\mu_{ki})^p$ – grau de similaridade da *i*-ésima observação ao *k*-ésimo grupo no passo (*p*)

 $(\mu_{ki})^{p+1}$ - grau de similaridade da *i*-ésima observação ao *k*-ésimo grupo no passo (p + I)

 P_i – peso do *i-ésimo* modelo

 ε_t

 d_i – distância do evento até o *i-ésimo* centroide, correspondente ao *i-ésimo* modelo

SUMÁRIO

1	INTRODUÇÃO	31
1.1	PROBLEMA	34
1.2	HIPÓTESES	36
1.3	OBJETIVOS	36
1.3.1	Objetivo geral	36
1.3.2	Objetivos específicos	36
1.4	JUSTIFICATIVA	36
1.5	ESTRUTURA DO TRABALHO	38
2	REVISÃO DE LITERATURA	39
2.1	CLASSES DE MODELOS HIDROLÓGICOS PARA PREVISÃO	39
2.2	APRENDIZADO DE MÁQUINAS	42
2.3	LÓGICA DIFUSA	45
2.4	ANÁLISE DE AGRUPAMENTO	46
2.5	APLICAÇÕES DA INTELIGÊNCIA ARTIFICIAL NA HIDROLOGIA	48
2.6	PRINCIPAIS FERRAMENTAS PARA PREVISÃO de vazões	51
2.6.1	Teoria do caos	51
2.6.2	SVM – Support Vector Machine	53
2.6.3	ELM – Extreme Learning Machine	60
3	MATERIAIS E MÉTODOS	65
3.1	ÁREA DE ESTUDO	65
3.1.1	Coleta e análise de dados	67
3.1.2	Seleção de eventos de cheia	74
3.1.3	Indicadores de cheia	78
3.2	MÉTODO DE AGRUPAMENTO	 7 9
3.3	MODELO HIDROLÓGICO PARA PREVISÃO DE VAZÕES	82
3.3.1	Dados de entrada dos modelos	83
3.3.2	Critérios de qualidade de ajuste	83
3.3.2.1	Índices de performance	83
3.3.2.2	Assinaturas hidrológicas	87
3.3.2.3	Medidas de erros relativos.	87
3.3.3	Algoritmo para previsão de vazões	87

3.3.4	Calibração dos parâmetros do modelo	90
3.3.5	Ajuste final dos modelos de previsão	95
3.4	CONFIGURAÇÕES DE MODELAGEM PARA PREVISÃO DE VAZÕES	96
3.4.1	Configuração 1 – Previsão Média	97
3.4.2	Configuração 2 – Previsão por Grupo de Pertencimento	97
3.4.3	Configuração 3 – Previsão Ponderada por Distâncias Euclidianas	98
3.4.4	Configuração 4 – Previsão Ponderada por Graus de Pertencimento	100
4	RESULTADOS E DISCUSSÕES	102
4.1	ANÁLISE DE AGRUPAMENTO	102
4.1.1	Agrupamento de eventos de cheia	102
4.1.2	Caracterização dos grupos de eventos	107
4.2	GERAÇÃO DOS MODELOS DE PREVISÃO	110
4.2.1	Treinamento	111
4.2.2	Teste	120
4.2.3	Ajuste final dos modelos	126
4.2.4	Validação dos modelos	129
4.3	PREVISÃO DE VAZÕES	133
4.3.1	Configuração 1 – Previsão Média	133
4.3.2	Configuração 2 – Previsão por Grupo de Pertencimento	134
4.3.3	Configuração 3 – Previsão Ponderada por Distâncias Euclidianas	135
4.3.4	Configuração 4 – Previsão Ponderada por Graus de Pertencimento	136
4.4	COMPARATIVO ENTRE AS CONFIGURAÇÕES DE PREVISÃO	138
4.5	AVALIAÇÃO DA PREVISÃO POR EVENTOS	141
5	CONCLUSÕES E RECOMENDAÇÕES	151
	REFERÊNCIAS BIBLIOGRÁFICAS	154
	APÊNDICE A – EVENTOS AGRUPADOS	166
	APÊNDICE B – VARIAÇÃO DE MÉTRICAS NA FASE DE TREINAME	NTO
		167
	APÊNDICE C – DISTRIBUIÇÃO DAS MÉTRICAS NA FASE	DE
	TREINAMENTO	177
	APÊNDICE D – DISTRIBUIÇÃO DAS MÉTRICAS NA FASE DE TESTE	182
	APÊNDICE E – VARIAÇÃO DAS MÉTRICAS NA FASE DE TESTE	187
	APÊNDICE F – COEFICIENTES KG, NS, NS MODIFICADO E MARE	DOS
	EVENTOS PREVISTOS	197

APÊNDICE G -ERROS RELATIVOS DAS ESTATÍSTICAS DOS EVENTOS
201
ANEXO 1 – AVALIAÇÃO DO COMPORTAMENTO DE MODELOS DE
PREVISÃO DE CHEIAS NO RIO IGUAÇU, EM UNIÃO DA VITÓRIA,
PARANÁ, UTILIZANDO O AJUSTE COM BASE EM AGRUPAMENTO207
ANEXO 2 – CARACTERÍSTICAS DE CHEIAS OBSERVADAS NA BACIA DO
RIO IGUAÇU – PARTE 1: AGRUPAMENTO DE EVENTOS218
ANEXO 3 – CARACTERÍSTICAS DE CHEIAS OBSERVADAS NA BACIA DO
RIO IGUAÇU – PARTE 2: AVALIAÇÃO DE EVENTOS OBSERVADOS 229
ANEXO 4 – EVALUATING DIFFERENT STRATEGIES FOR MACHINE
LEARNING TRAINING APPLIED TO FLOW FORECASTING BASED ON
CLUSTERING OF FLOOD EVENTS240

1 INTRODUÇÃO

O modelo hidrológico consiste em uma das ferramentas desenvolvidas pela Ciência para entender, representar e prever condições diante das características de uma bacia hidrográfica (Tucci, 2005). Os objetivos do modelo são alcançados por meio das simulações, limitadas pela heterogeneidade física da bacia e dos processos envolvidos, o que propiciou o surgimento de modelos que se diferenciam a partir dos dados de entrada, da discretização, de suas prioridades de representação e, obviamente, pelos próprios objetivos a serem atingidos (Tucci, 2005). No processo de simulação, contudo, existem três fases até a obtenção dos resultados desejados: o ajuste, a verificação e a previsão; a etapa de previsão consiste na simulação propriamente dita, onde o modelo, calibrado e verificado, é aplicado para definir a saída do sistema para as condições requeridas (Tucci, 2005).

A modelagem hidrológica registrou grande avanço desde o início da revolução computacional, em que o poder de processamento dos computadores favoreceu, dentre outros fatores, a utilização de grandes séries de dados (Singh, 2018). Para Tucci (2005), a Hidrologia passou "de uma ciência preponderantemente descritiva e qualitativa, para uma área de conhecimento onde métodos quantitativos têm sido explorados, através de metodologias matemáticas e estatísticas"; o desenvolvimento dos modelos representa esta evolução, em que, a partir do uso de computadores na década de 50, desenvolveram-se os modelos semiconceituais e, em paralelo, os modelos empíricos.

Neste contexto, a previsão hidrológica tem sido efetuada de duas formas principais (Ding *et al.*, 2020; Mohammadi *et al.*, 2020; Snieder *et al.*, 2020; Wang *et al.*, 2021; Difi *et al.*, 2023): (i) pelo uso de modelos hidrológicos tradicionais, e (ii) pelo uso de modelos de orientados a dados (*data-driven models*), os quais podem ser elaborados com base nas técnicas de Inteligência Artificial e, ainda, combinadas aos algoritmos de otimização (Zhang *et al.*, 2018; Mohammadi *et al.*, 2020; Ibrahim *et al.*, 2022; Difi *et al.*, 2023; Zakhrouf *et al.*, 2023).

Na modelagem hidrológica clássica, modelos tradicionais analisam características hidrológicas e descrevem fisicamente a confluência de vazões, atribuindo a seus parâmetros certa relevância física (Ding *et al.*, 2020). No entanto, a calibração destes modelos se baseia em dados observados, requer certa familiaridade do pesquisador com o modelo (Ding *et al.*, 2020; Ebtehaj e Bonakdari, 2022), e, ainda, estimativas incorretas dos parâmetros podem aumentar, de maneira significativa, os erros desta classe de modelos (Luppichini *et al.*, 2019; Luppichini *et al.*, 2022).

O processo de geração de previsões engloba muitas incertezas, que, por se propagarem pelo sistema, deveriam ser quantificadas (Fan *et al.*, 2014) a fim de se obter uma previsão mais acurada. Da mesma forma, a complexidade do processo de formação de vazões é de difícil determinação. Para contornar estas dificuldades, podem ser gerados diferentes cenários de previsão, que formam um conjunto (*ensemble*), onde cada um de seus membros representa uma possível trajetória dos processos ao longo do horizonte de previsão (Fan *et al.*, 2014).

Ebtehaj e Bonakdari (2022), embora reconheçam que os modelos tradicionais produziram bons resultados para a previsão de cheias, também destacam várias limitações. Estas incluem a coleta manual de dados, que são frequentemente armazenados de forma desorganizada, conduzindo a séries de dados de dimensão insuficiente — especialmente problemáticas em regiões remotas, onde a disponibilidade de dados já é limitada. Luppichini *et al.* (2022) afirmam, ainda, que os modelos de base física enfrentam desafios devido à heterogeneidade inerente aos sistemas naturais.

Por outro lado, com o crescimento da infraestrutura hidrológica, o aumento da disponibilidade de dados e os avanços no desenvolvimento computacional (Singh, 2018; Bai *et al.*, 2021), a capacidade dos modelos empíricos se expandiu, levando ao surgimento dos chamados modelos orientados por dados. Esses modelos utilizam dados do sistema para identificar conexões entre variáveis (de entrada, internas e de saída) sem explicar o comportamento físico do sistema. Em outras palavras, a fonte de informação para esses modelos é a própria série de dados (Adnan et al., 2019; Fathian et al., 2019; Ding et al., 2020; Bai et al., 2021; Ebtehaj e Bonakdari, 2022).

Os modelos orientados a dados captam relações lineares e não lineares entre as descargas e seus parâmetros matemáticos (Zhang et al., 2018; Adnan et al., 2019). As abordagens comumente utilizadas incluem o método dos mínimos quadrados, a regressão linear múltipla e modelos como os autorregressivos (AR), de médias móveis (MA) e autorregressivos de médias móveis (ARMA). No entanto, estes métodos normalmente consideram apenas as relações lineares entre as entradas e saídas, resultando frequentemente em saídas insuficientes e que não consideram a não-linearidade inerente aos sistemas (Khosravi et al., 2021).

Para enfrentar estas desvantagens foram desenvolvidos e explorados modelos automatizados de séries temporais utilizando Inteligência Artificial (IA) (Bai et al., 2021; Khosravi et al., 2021), denominados modelos inteligentes orientados por dados (Ding et al., 2020; Khosravi et al., 2021). Estes esforços se concentram na Inteligência Computacional (IC) (Fathian et al., 2019; Bai et al., 2021), sendo seus modelos notáveis por sua capacidade de lidar com séries extensas de dados e acomodar observações em diferentes escalas (Bai et al., 2021;

Khosravi et al., 2021). Entre as técnicas de IC mais populares estão as redes neurais, os sistemas baseados em lógicas difusas e os algoritmos genéticos (Mahdavi-Meymand et al., 2023).

Dentre os métodos da Inteligência Computacional, as técnicas de Aprendizagem de Máquinas (ML) envolvem a detecção automática de padrões significativos em séries de dados. Estas técnicas são categorizadas com base em diferentes objetivos de aprendizagem (Xu e Liang, 2021), incluindo classificação, regressão e agrupamento (Ibrahim et al., 2022; Mahdavi-Meymand et al., 2023). Nas últimas duas décadas, essas técnicas produziram resultados promissores para problemas hidrológicos (Yu et al., 2004; Lin et al., 2006; Lohani et al., 2011; Lohani et al., 2012; Yaseen et al., 2016; Moeeni et al., 2017; Zhang et al., 2018; Li et al., 2019; Saadi et al., 2019; Ding et al., 2020; Schoppa et al., 2020; Snieder et al., 2020; Chen et al., 2021; Desai e Ouarda, 2021; Niu e Feng, 2021; Kim et al., 2022; Difi et al., 2023; Sharma et al., 2023; Zakhrouf et al., 2023).

A maioria das aplicações envolve ferramentas das Redes Neurais Artificiais (RNA) (Zhang et al., 2018; Fathian et al., 2019; Snieder et al., 2020; Brêda et al., 2021; Chen et al, 2021; Lima e Scofield, 2021) e suas variações, como as *Extreme Learning Machines* (ELM) (Li et al., 2019; Ribeiro et al., 2020; Yaseen et al, 2020; Niu e Feng, 2021; Ebtehaj e Bonakdari, 2022; Feng et al., 2022; Difí et al., 2023) e *Long Short-Term Memory* (LSTM) (Ding et al., 2020; Chen et al., 2021; Kim et al., 2022; Sharma et al., 2023; Zakhrouf et al., 2023), bem como as Máquinas de Vetores de Suporte (SVM) e a sua variante de regressão, Regressão por Vetores de Suporte (SVR) (Li et al., 2019; Chen et al., 2021; Niu e Feng, 2021; Kim et al., 2022; Difí et al., 2023; Sharma et al, 2023). De um modo geral, a investigação tem enfoque na comparação destas técnicas. Recentemente, LSTM, ELM e SVM (ou SVR) têm sido frequentemente comparados em estudos (Li et al., 2019; Chen et al., 2021; Niu e Feng, 2021; Kim et al., 2022; Difí et al., 2023; Sharma et al., 2023).

Para a previsão de vazões, Adnan et al. (2019) destacam que, nas últimas décadas, tem-se observado aplicações bem-sucedidas de métodos como ANN, ANFIS, MARS, SVM, algoritmos genéticos (GA) e seus modelos híbridos. Estes métodos são eficazes na identificação da não-linearidade nos processos de formação de cheias. Neste trabalho, o algoritmo SVR foi escolhido pelos seus resultados satisfatórios e pela sua implementação computacional simples. No entanto, o modelo é apenas uma ferramenta a ser aplicada, sendo necessário considerar outros pormenores na modelagem hidrológica para previsão.

Desta forma, o objetivo deste trabalho é desenvolver uma nova abordagem de modelagem hidrológica para a previsão de vazões com base no agrupamento prévio de eventos de cheia. A principal finalidade da análise de agrupamento é utilizar as variáveis caraterísticas

dos dados para classificá-los de forma que objetos semelhantes estejam no mesmo conjunto, reduzindo o número de observações numa amostra (Morettin e Singer, 2023). Este processo requer também a definição da similaridade dentro de uma classe e da dissimilaridade entre classes, o que é uma tarefa complexa (Ezugwu et al., 2022). Nesta pesquisa, o algoritmo Fuzzy C-Means (FCM) foi utilizado como método de agrupamento para atribuir cada evento de cheia a um grupo, com base em um número predefinido de grupos (Mosavi et al., 2021), a partir dos quais foram gerados modelos hidrológicos por meio do algoritmo SVR.

A abordagem de modelagem hidrológica proposta foi demonstrada através de uma aplicação no mundo real; foram utilizadas as vazões médias diárias de 1964 a 2015 na Estação Fluviométrica União da Vitória (65310000), na bacia do rio Iguaçu, Estado do Paraná, Brasil. O algoritmo FCM agrupou os eventos de cheia em um número de grupos predefinido, e, após, o algoritmo SVR foi aplicado para desenvolver modelos de Inteligência Artificial para previsão de vazão com horizonte de 1 dia. Assim, foram treinados os modelos para a série completa, para todos os eventos de cheia e um para cada grupo.

Ainda, uma vez que cada grupo de eventos resulta em um modelo hidrológico, foram realizadas previsões em quatro configurações de modelagem, que simulam um sistema de previsão em tempo real: (i) média; (ii) por grupo de pertencimento; (iii) ponderada pelas distâncias Euclidianas; e (iv) ponderada pelos graus de pertencimento. Desta forma, tomando como base a classificação de eventos com características semelhantes, a abordagem proposta tende a restringir as características do evento previsto àquelas do grupo ao qual ele pertence, reduzindo as possibilidades de previsão e, consequentemente, os modelos utilizados.

1.1 PROBLEMA

De acordo com Fan *et al.* (2016), no Brasil, os impactos econômicos e sociais advindos de ocorrências de cheias são frequentes, tanto devido às cheias de rios maiores, quanto daqueles de menores dimensões. Meller (2012) afirma que, historicamente, a redução dos impactos oriundos de cheias tem sido idealizada a partir de intervenções estruturais na bacia. Porém, apesar da significativa contribuição para a diminuição da vulnerabilidade da região protegida, sempre há chances de superação do risco estimado em projeto. Por isso, tem-se questionado a implantação de medidas que permitam antecipar os eventos extremos de cheia, favorecendo uma diminuição nos impactos. Neste contexto, os sistemas de previsão e alerta têm sido reconhecidos como medidas efetivas na minimização dos impactos causados por cheias, pois,

após emitidos alertas pelo sistema, torna-se possível a aplicação dos planos de ação a fim de evacuar áreas de risco.

A concepção dos sistemas de previsão e alerta de cheias e sua gestão adequada consistem em desafios para as autoridades locais com eventos recorrentes e, na tentativa de se executar um sistema adequadamente, Meller (2012) afirma que um dos principais elementos de um sistema de alerta de cheias é o sistema de previsão hidrológica, cujo objetivo é diminuir as incertezas relativas às condições futuras do rio em pontos considerados vulneráveis, para então subsidiar a tomada de decisão de emissão dos alertas. Previsão acurada de vazões, por sua vez, possibilita o controle sustentável das cheias e uma minimização dos esforços demandados pelos hidrólogos e gestores de recursos hídricos (Khosravi *et al.*, 2021).

Desta forma, apesar da existência de alternativas como as previsões por conjunto (ensemble), não se tem completa compreensão do benefício e da necessidade de se considerar todas as incertezas dos sistemas nas previsões hidrológicas (Fan et al., 2014), o que reforça a existência das incertezas das previsões e a necessidade de implantação de novas formas de previsão hidrológica. Portanto, ainda é desafiador desenvolver modelos únicos de previsão de vazão que sejam capazes de mapear adequadamente as variáveis envolvidas para os diferentes regimes de escoamento (Chu et al., 2020).

Para Tarasova *et al.* (2019), uma vez que os mecanismos de geração de eventos não estão bem delimitados a nível de bacia hidrográfica, as diferenças entre suas características são negligenciadas. Os autores afirmam que existem três categorias principais para os mecanismos de geração de cheias: hidroclimática, caracterizada por padrões de circulação em grande escala e condições atmosféricas no início do evento; hidrológica, definida pelos padrões de precipitação da bacia hidrográfica e suas condições antecedentes; e baseada no hidrograma, que considera os mecanismos de formação através dos seus efeitos nas caraterísticas do hidrograma (Tarasova et al., 2019).

Portanto, entender, avaliar e prever eventos de cheia e seus impactos é fundamental na tentativa de prevenir perdas humanas e danos às propriedades locais. Neste contexto, a modelagem hidrológica pode fornecer informações capazes de gerar previsões confiáveis em intervalos de tempo adequados às tomadas de decisões. Contudo, em virtude das limitações existentes nos processos conhecidos e atualmente utilizados, outras abordagens podem colaborar na melhor compreensão dos mecanismos de geração de eventos e aprimorar a realização de previsões em tempo real, como proposto nesta pesquisa.

1.2 HIPÓTESES

Verifica-se, inicialmente, a aplicabilidade da análise de agrupamento como subsídio à modelagem hidrológica, de modo que se ajuste um modelo para cada grupo de eventos, capaz de simular adequadamente diferentes séries de vazões e com desempenho superior ao prever a mesma categoria de série para a qual o modelo foi treinado. Após, são concebidos sistemas de previsão de vazões em tempo real – baseados nos modelos por grupo pré-estabelecidos –, cujas capacidades de performance devem melhorar com o aumento de sua complexidade.

1.3 OBJETIVOS

Após a contextualização do tema e as motivações à execução do trabalho, apresentamse os objetivos geral e específicos da pesquisa.

1.3.1 Objetivo geral

O objetivo geral da pesquisa é desenvolver uma nova abordagem de modelagem hidrológica para previsão de vazões, baseada em grupos previamente definidos de eventos observados.

1.3.2 Objetivos específicos

Como objetivos específicos deste trabalho podem ser destacados:

- a) Delimitar e selecionar eventos de cheia.
- Definir grupos de eventos observados, com base em características específicas de cheias.
- c) Calibrar e validar modelos hidrológicos para diferentes grupos de eventos observados.
- d) Desenvolver uma abordagem de modelagem hidrológica de previsão de vazões em tempo real, a partir de modelos por grupos.

1.4 JUSTIFICATIVA

Com o intuito de melhor representarem processos de diferentes áreas, com diversos níveis de robustez e limitações inerentes, têm se desenvolvido várias técnicas de Aprendizado

de Máquinas (*Machine Learning*) (Sharma et al., 2023), com aplicações na Hidrologia e resultados satisfatórios (Solomatine *et al.*, 2008; Zhang *et al.*, 2018; Liang *et al.*, 2019; Ding *et al.*, 2020; Kurian *et al.*, 2020). No entanto, em geral, o que se observa é a comparação de diferentes técnicas – modelos – capazes de bem reproduzir um sistema de vazões (Zhang *et al.*, 2018; Chu *et al.*, 2020; Ding *et al.*, 2020; Kim *et al.* 2022), enquanto os processos de modelagem se baseiam em abordagens convencionais ou por *ensemble*, ainda em crescimento (Khosravi *et al.*, 2021; Das *et al.*, 2022).

Da mesma forma, Xu e Liang (2021) observaram que os métodos de aprendizagem automática para agrupamento e classificação são frequentemente utilizados para apoiar os métodos de regressão. No entanto, estas abordagens servem, normalmente, como alternativas às técnicas de regressão, ao invés de integrarem uma modelagem hidrológica por grupos, por exemplo. Portanto, as técnicas de agrupamento aplicadas à Hidrologia possuem como objetivo a caracterização dos eventos em si (Wang *et al.*, 2015; Liu *et al.*, 2017; Joo *et al.*, 2021; Fischer e Schumann, 2022) – como nas análises de frequência de cheias, na regionalização de vazões e na classificação de características de bacias hidrográficas –, sem direcionar a uma utilização mais prática e direta dos resultados obtidos a partir da técnica.

Neste contexto, o presente trabalho propõe uma nova abordagem para modelagem hidrológica, na qual o agrupamento de cheias é utilizado como etapa prévia ao processo de previsão de vazões, enfatizando caraterísticas semelhantes entre eventos dentro de um mesmo grupo. Ao agrupar eventos com caraterísticas semelhantes, a abordagem restringe as caraterísticas do evento previsto às do seu respectivo grupo (Joo *et al.*, 2021), reduzindo a abrangência das previsões produzidas por um único modelo. Assim, quanto mais distante um evento se apresentar de determinado grupo de cheias observadas, menor tende a ser a semelhança do evento com o modelo do referido grupo.

Desta forma, o processo de modelagem por grupos favorece um sistema de previsão de vazões em tempo real – simulado nesta pesquisa – reduzindo as possibilidades de previsão, de forma que, ao identificar a ocorrência de um evento, o sistema captura suas características, como tempo e vazão de pico atuais, e produz previsões a partir dos modelos por grupo mais próximos às características capturadas; com o avançar do evento, as características são atualizadas a cada instante de tempo e as previsões se tornam restritas a um grupo em particular, com o qual o evento compartilha maior grau de similaridade.

Por fim, reforça-se que a abordagem proposta e as contribuições científicas da pesquisa não se restringem aos algoritmos adotados – FCM e SVR –, mas, sim, baseiam-se na concepção do sistema descrito. Ou seja, diferentes categorias de modelos hidrológicos poderiam ser

adotadas, beneficiando diferentes áreas de estudo, com resultados que dependeriam do agrupamento em si, do próprio modelo hidrológico e do horizonte de previsão considerado.

1.5 ESTRUTURA DO TRABALHO

O presente trabalho está subdividido em cinco Capítulos, que compreendem, (i) no Capítulo 1, a apresentação do contexto em que se insere a nova abordagem de modelagem hidrológica para previsão de vazões, bem como, a relevância do trabalho proposto, (ii) no Capítulo 2, a apresentação da revisão de literatura relacionada à modelagem hidrológica, sua evolução ao longo do tempo, bem como, a identificação das principais técnicas aplicadas para previsão de vazões, (iii) no Capítulo 3, os materiais e métodos do trabalho, partindo do estudo da bacia hidrográfica e dos eventos de cheia delimitados, avançando pelo método de agrupamento aplicado e finalizando com os processos de modelagem desenvolvidos, (iv) no Capítulo 4, a apresentação dos resultados obtidos e suas respectivas discussões, e (v) no Capítulo 5, as principais conclusões e recomendações desta pesquisa.

2 REVISÃO DE LITERATURA

Este capítulo apresenta o embasamento teórico relacionado aos fundamentos deste trabalho – modelagem hidrológica para previsão de vazões, aliada à técnica da estatística multivariada para caracterização dos eventos de cheia. Inicialmente, são apresentados os principais conceitos, abordagens e técnicas da modelagem hidrológica, perpassando pelos modelos tradicionais e pelos modelos orientados a dados, proporcionados pela evolução computacional. Na sequência, adentra-se às definições pertinentes da análise de agrupamentos – estatística multivariada –, para o entendimento da classificação de cheias e das ferramentas computacionais existentes para aplicação, atingindo as técnicas de Aprendizado de Máquinas – *Machine Learning*.

2.1 CLASSES DE MODELOS HIDROLÓGICOS PARA PREVISÃO

Os modelos hidrológicos podem ser classificados como físicos, matemáticos e empíricos. Ao contrário dos modelos físicos e matemáticos, os modelos empíricos envolvem equações matemáticas que não são derivadas de processos físicos na bacia, mas sim da análise das séries históricas dos dados, como, por exemplo, o modelo de Hidrograma Unitário (Solomatine *et al.*, 2008).

No contexto clássico, pode-se dizer que os modelos hidrológicos tradicionais analisam características hidrológicas e descrevem a confluência de vazões fisicamente e, apesar de suas diferentes categorias e aplicações, os seus parâmetros apresentam uma relevância física (Solomatine et al., 2008; Ding *et al.*, 2020). A calibração dos modelos, contudo, pode ser efetuada de acordo com a observação ou dados analisados, o que exige alta familiaridade do pesquisador com o modelo (Tucci, 2005; Ding et al., 2020; Ebtehaj e Bonakdari, 2022).

Ding et al. (2020) afirmam que, no início dos estudos de previsão de cheias, hidrólogos estabeleceram modelos de bacias hidrográficas a partir da hidrodinâmica, e apesar da boa performance dos modelos, os pesquisadores ainda julgaram difícil a sua calibração diante das condições existentes; para Adnan et al. (2019), a performance confiável de um modelo não pode ser garantida, dependendo da área de estudo e dos objetivos específicos, diante da eventual difículdade para obtenção de parâmetros. Modelos físicos exigem uma grande quantidade de dados e uma variedade de parâmetros baseados nas alturas, na intensidade e na distribuição da

precipitação, nas características fisiográficas da bacia hidrográfica e no uso e ocupação do solo (Adnan *et al.*, 2019).

Por sua vez, Ebtehaj e Bonakdari, 2022, apesar de afirmarem que os modelos tradicionais têm produzido bons resultados para previsão de cheias, listaram algumas de suas limitações, e dentre elas está a coleta de dados majoritariamente manual, cujo armazenamento se dá de maneira desorganizada e resulta em uma série de dados, muitas vezes, de tamanho insuficiente, fazendo com que a disponibilidade de dados para regiões mais afastadas seja ainda mais dificultosa.

Com o crescimento da infraestrutura hidrológica, aumento na disponibilidade de dados e os avanços no desenvolvimento computacional (Solomatine *et al.*, 2008; Singh, 2018), a capacidade dos modelos empíricos tem se expandido, dando origem aos chamados *data-driven models* (modelos orientados a dados ou modelos automatizados de séries temporais), que utilizam os dados de um sistema para encontrar conexões entre suas variáveis (de entrada, internos e de saída), sem explicitar características de seu comportamento físico; ou seja, sua única fonte de informação é a própria série (Solomatine *et al.*, 2008; Detzel, 2015; Adnan *et al.*, 2019; Fathian *et al.*, 2019; Ding *et al.*, 2020; Ebtehaj e Bonakdari, 2022).

Série temporal é "um encadeamento de observações de uma variável no tempo", considerando, majoritariamente, elementos igualmente espaçados (Detzel, 2015). Modelos automatizados de séries temporais possuem vastas aplicações em previsões de cheias e têm ganhado popularidade diante de sua simplicidade, menor demanda de dados e fácil implementação para aplicações em tempo real (Zhang et al., 2015; Adnan et al., 2019; Ebtehaj e Bonakdari, 2022).

Liang et al. (2019) afirmam que existem duas categorias para os data-driven models, e são classificados de acordo com os fatores que utilizam para previsão. A primeira corresponde aos métodos que utilizam elementos únicos na previsão (ou análise de séries temporais) baseada na própria série hidrológica, enquanto a segunda aborda a previsão com base em múltiplos fatores, sendo construído um modelo de regressão linear múltipla ou um modelo autorregressivo, capaz de determinar os mecanismos associados às interações entre os elementos hidrológicos.

Os modelos de séries temporais apresentam relações lineares e não-lineares entre as vazões e seus parâmetros matemáticos (Yaseen et al., 2016; Zhang et al., 2018; Adnan et al., 2019). O método dos mínimos quadrados, regressão linear múltipla, bem como os modelos autorregressivos (AR), de médias móveis (MA) e autorregressivos de médias móveis (ARMA), são abordagens frequentemente utilizadas (Fan et al., 2016), mas que consideram apenas a

relação linear entre as suas entradas e saídas; o resultado de suas saídas são, em geral, insuficientes desconsiderando a relação não-linear (Adnan *et al.*, 2019).

Ainda, modelos como os Autorregressivos (AR), de Médias Móveis (MA) e Autorregressivos de Médias Móveis (ARMA) requerem longas séries de dados estacionários para retornarem uma previsão robusta (Zhang *et al.*, 2018). No entanto, séries de vazões possuem fortes características não lineares, inconsistentes e não estacionárias, o que dificulta a previsão de vazões pelos modelos estatísticos de séries temporais (Zhu *et al.*, 2016; Khosravi *et al.*, 2021).

Na tentativa de eliminar estas desvantagens, modelos inteligentes orientados a dados utilizando Redes Neurais Artificiais e modelos baseados na lógica difusa têm sido explorados e desenvolvidos para modelagem hidrológica e previsão de vazões (Khosravi *et al.*, 2021), além de gerar contribuições para diversas áreas da Inteligência Artificial.

Para contextualizar as áreas da Inteligência Artificial, Solomatine *et al.* (2008) apresenta as relações entre elas:

- Inteligência artificial (AI Artificial Intelligence): estudo de como a inteligência humana pode ser incorporada aos computadores;
- 2. Inteligência computacional (CI *Computational Intelligence*): composta por redes neurais, sistemas difusos (*fuzzy systems*), computação evolucionária, inteligência artificial e aprendizado de máquinas;
- Computação flexível (SC Soft Computing): similar à inteligência computacional, mas com ênfase em sistemas baseados em regras difusas induzidas pelos dados;
- Aprendizado de máquinas (ML Machine Learning): subárea da inteligência artificial, que prioriza os fundamentos da inteligência computacional e da computação flexível;
- 5. Mineração de dados (DM Data Mining) e descoberta do conhecimento em banco de dados (KDD Knowledge Discovery in Databases): para bancos de dados de grandes dimensões, com aplicações de métodos estatísticos e de aprendizagem de máquinas;
- 6. Análise inteligente de dados (IDA *Intelligent Data Analysis*): focado em análise de dados para medicina e pesquisa, incorporando métodos estatísticos e de aprendizagem de máquinas.

Ante o exposto, tem-se que os modelos orientados a dados tiveram seu enfoque em métodos da Inteligência Computacional e do Aprendizado de Máquinas que pudessem ser aplicados para complementar ou substituir modelos físicos (Solomatine *et al.*, 2008; Fathian *et al.*, 2019), sendo denominados de *data-driven intelligent models*, ou modelos de Inteligência Artificial para séries temporais (Ding *et al.*, 2020; Khosravi *et al.*, 2021), que se destacam por sua habilidade em lidar com grandes séries de dados e aceitar dados em diferentes escalas (Khosravi *et al.*, 2021).

Ainda, a estrutura de "caixa-preta" dos modelos de Inteligência Artificial, juntamente com (i) sua capacidade de trabalhar como funções fortes, auto adaptativas e de autoaprendizado, (ii) sua capacidade de lidar com características não-lineares utilizando funções também não-lineares, e (iii) sua habilidade de analisar múltiplas entradas com atributos variáveis, facilitaram sua disseminação (Khosravi *et al.*, 2021).

2.2 APRENDIZADO DE MÁQUINAS

Shalev-Shwartz e Ben-David (2014) afirmam que o aprendizado de máquinas se refere à detecção automática de significativos padrões em séries de dados. Para Cristianini e Shawe-Taylor (2000), a construção de máquinas capazes de aprender a partir de experiências foi, por um longo período, objetivo de discussões técnicas e filosóficas. A questão técnica do debate se desenvolveu desde a invenção dos computadores, que demonstraram que máquinas podem evidenciar uma significativa capacidade de aprendizado, apesar dos limites desta capacidade não serem bem definidos (Cristianini e Shawe-Taylor, 2000).

Cristianini e Shawe-Taylor (2000) ainda afirmam que a disponibilidade de sistemas de aprendizagem confiáveis é de importância estratégica, pois, para problemas mais complexos, quando o modelo matemático não pode ser precisamente definido, a solução não pode ser obtida por técnicas clássicas de programação. Nesta situação, uma alternativa é fazer com que o computador aprenda as operações de entrada e saída a partir de exemplos, denominados como dados de treinamento (Cristianini e Shawe-Taylor, 2000).

Shalev-Shwartz e Ben-David (2014), por sua vez, afirmam que além da alta complexidade, problemas que apresentam necessidade de adaptação também requerem programas que aprendam e melhorem com base na experiência. A rigidez é uma limitação de

ferramentas computacionais, ao contrário das ferramentas de aprendizagem de máquinas, cujo comportamento se adapta aos dados disponíveis.

Por se tratar de uma área bastante ampla, o aprendizado de máquinas se divide em categorias, que abordam diferentes objetivos de aprendizado (Shalev-Shwartz e Ben-David, 2014). A principal categoria diferencia o aprendizado em "supervisionado" e "não supervisionado" (Cristianini e Shawe-Taylor, 2000; Shalev-Shwartz e Ben-David, 2014; Ibrahim *et al.*, 2022), sendo que no primeiro caso há uma resposta associada (variável de saída) para cada observação das variáveis preditoras (variáveis de entrada) na etapa de treinamento, com o objetivo de obter respostas para as futuras predições (entradas da etapa de teste); as variáveis de saída da etapa de treinamento funcionam como guias do processo de aprendizado (Azevedo, 2018). No aprendizado não supervisionado não há variáveis de saída na etapa de treinamento e, neste caso, o desafio da aprendizagem está em entender o processo de geração dos dados (Cristianini e Shawe-Taylor, 2000).

Ainda, há uma categoria intermediária entre "supervisionado" e "não supervisionado", chamada de aprendizado de reforço, na qual o aprendiz tem um intervalo de respostas associadas (saída) capazes de gerar resultados satisfatórios (Cristianini e Shawe-Taylor, 2000; Shalev-Shwartz e Ben-David, 2014).

Na aprendizagem supervisionada, um conjunto de dados possui uma dependência desconhecida (y = f(x)) entre um vetor de entrada multidimensional (x) e uma saída (y). As informações disponíveis compõem a amostra de treinamento (x_i, y_i) , em que $x \in \mathcal{X}$, $y \in Y$, i = 1, ..., m, e m representa o tamanho da amostra (Lorena e Carvalo, 2007). A função f(x) definida no treinamento deve ser capaz de prever as classes (y) de novos dados de entrada (x) na fase de teste (Lorena e Carvalo, 2007), definindo o critério de generalização (Cristianini e Shawe-Taylor, 2000; Schölkopf e Smola, 2002). As classes representam o fenômeno de interesse sobre o qual se deseja efetuar previsões, e a subdivisão do conjunto de dados entre treinamento e teste se dá a fim de mensurar os erros e acertos do modelo (Lorena e Carvalo, 2007). De acordo com o tipo de variável de saída (y), são definidos diferentes problemas de aprendizagem supervisionada (Lorena e Carvalho, 2007; Moura et al., 2010; Ibrahim et al., 2022): (i) classificação; ou (ii) regressão. Ainda, Solomatine et al. (2008) afirmam que métodos aprendizagem de máquinas para desenvolvimento de agrupamento e classificação são frequentemente utilizados para apoiar métodos de regressão, como o SVM e a teoria do caos.

A **classificação** ocorre quando um objeto precisa ser atribuído a uma categoria (ou classe) pré-definida, baseada em um conjunto de características observadas relacionadas ao objeto (Ibrahim *et al.*, 2022), o que significa dizer que *y* assume valores discretos, os quais

representam as categorias; se apenas duas categorias são consideradas ($y \in Y, Y = \{-1; +1\}$), o problema consiste em uma classificação binária; caso contrário, é um problema multiclasses (Lorena e Carvalho, 2007; Moura *et al.*, 2010). Ibrahim *et al.* (2022) destacaram alguns modelos de classificação do Aprendizado de Máquinas: *Artificial Neural Networks* (ANN), *Back Propagation Neural Networks* (BPNN), *Multilayer Perceptron* (MLP), *Recurrent Neural Network* (RNN), *Convolutional Neural Network* (CNN), *Support Vector Machine* (SVM), *Radial Basis Function* (RBF), *K-Nearest Neighbour* (K-NN), dentre outros.

A **regressão**, por sua vez, é uma ferramenta matemática para avaliar a intensidade e natureza da conexão entre a variável dependente (y) e as demais (variáveis independentes) (Ibrahim *et al.*, 2022); ou seja, y assume valores reais (contínuos) $(y \in \mathbb{R})$ e sua relação com o vetor de entrada (x) é dado pela função y = f(x) (Lorena e Carvalho, 2007; Moura *et al.*, 2010). Dentre os modelos de regressão destacados por Ibrahim *et al.* (2022) estão: *Artificial Neural Networks* (ANN), *Support Vector Regression* (SVR), *Random Forest* (RF), *Adaptive Neuro-Fuzzy Inference System* (ANFIS), além da regressão linear.

A aprendizagem não-supervisionada, em contrapartida, contempla as técnicas de agrupamento (*clustering*). A diferença entre a classificação e o agrupamento está na não existência de classes pré-definidas deste método; neste caso, grandes bancos de dados são divididos em subgrupos, nos quais os dados que os compõem apresentam alta similaridade entre si (Ibrahim *et al.*, 2022). Ibrahim *et al.* (2022) evidenciam que, dentre as técnicas de agrupamento, o método *k-means* é o mais difundido, devido à sua simplicidade e à alta velocidade de convergência, porém, também apresentam outras ferramentas, como os modelos de lógica difusa (*fuzzy logic*), *Artificial Neural Networks* (ANN) e *Self-Organizing Map* (SOM).

Ibrahim *et al.* (2022) ainda apresentam técnicas de **otimização** – utilizadas para estimativa dos parâmetros dos modelos –, em que o *processamento evolucionário* é um campo da ciência da computação que utiliza conceitos da evolução biológica para resolver problemas complexos, como, por exemplo, os algoritmos genéticos (GA) e *Particle Swarm Optimization* (PSO).

Também, é importante observar a tendência de *overfitting* dos modelos, que ocorre quando suas hipóteses se tornam muito mais complexas, mas ainda assim, consistentes, representando uma taxa de acerto bastante elevada (Cristianini e Shawe-Taylor, 2000). Em outras palavras, o *overfitting* ocorre quando o algoritmo é muito especializado em uma amostra de dados (D), mas não se adequa tão bem a dados que não estão nesta mesma amostra (D) (Moura *et al.*, 2010). Cristianini e Shawe-Taylor (2000) também afirmam que uma forma de tentar controlar este problema é restringindo a abrangência das hipóteses.

Hammer e Villmann (2007) afirmam que um dos principais problemas em aprendizado de máquinas é a existência de incertezas, que pode ser devido à: (i) ruídos, *outliers*, e erros nos dados de entrada; porém, um método de aprendizado de máquina tem que lidar com as informações difusas, mostrando robustez em relação aos distúrbios, que podem resultar em um efeito positivo no comportamento de generalização do método de aprendizado de máquina, uma vez que é forçado a desenvolver formas de invariância e a abstrair do ruído; (ii) as decisões de saída devem ser acompanhadas de uma medida que permita julgar a incerteza de saída; (iii) representação de informações em um sistema de aprendizado de máquina é distribuída e difusa. Diante disso, Lorena e Carvalho (2007) afirmam que um requisito importante para a aprendizagem de máquinas é que sejam capazes de lidar com a existência destes ruídos, *outliers* e erros. Neste contexto, insere-se a análise difusa, que permite a representação das incertezas inerentes aos dados reais (Tao, 2002; Ross, 2004; Sato-Ilic e Jain, 2006; Hüllermeier, 2015).

2.3 LÓGICA DIFUSA

O conceito de conjuntos difusos foi introduzido em 1965, por Lofti Zadeh, sob influência da consideração das incertezas inerentes aos fenômenos em geral, conceituadas, até então, pela teoria de probabilidades. A ideia apresentada por Zadeh sugere que um nível de pertencimento é o ponto principal de uma decisão frente às incertezas, o que desafiou, inclusive, toda a base da teoria de probabilidades: a classificação binária (Ross, 2004).

Isto apresenta uma distinção entre a lógica difusa e probabilidade: enquanto a lógica difusa demonstra a distinção entre os fenômenos, a probabilidade descreve a incerteza quanto à ocorrência dos fenômenos. Ou seja, para a probabilidade, o evento pode acontecer, ou não, mas na lógica difusa observa-se se a descrição do evento é clara o suficiente para *medir* sua ocorrência ou não ocorrência (Ross, 2004). Direcionada aos princípios de classificação ou agrupamento, a análise difusa tem como finalidade estabelecer as semelhanças que um dado da amostra compartilha com cada grupo, a partir da minimização de uma função objetivo (Bezdek *et al.*, 1984). A Figura 2.1 apresenta um sistema difuso genérico e seus componentes: *fuzzificação*, base da regra difusa, mecanismo de inferência difusa e a *desfuzzificação*.

A *fuzzificação* converte cada observação dos dados de entrada em graus de pertinência, por meio da pesquisa em uma ou mais funções de pertinência. Ao contrário do conceito clássico de conjuntos – ou grupos –, que determina o pertencimento completo (ou o não pertencimento)

de certo dado da amostra a um determinado conjunto, a lógica difusa permite o pertencimento parcial de uma observação a diferentes subconjuntos (Sato-Ilic e Jain, 2006; Turan e Yurdusev, 2009). Ou seja, o pertencimento parcial a um conjunto pode ser descrito numericamente por uma função de pertinência, que assume valores maiores que 0 e menores ou iguais a 1 (Turan e Yurdusev, 2009), permitindo que os objetos se relacionem com todos os grupos, em maior ou menor grau, dependendo do grau de pertinência apresentado; quando os valores das associações são elevados (próximos de 1), existe grande semelhança entre a amostra e o grupo, enquanto aqueles que estão perto de zero, implicam em pouca semelhança (Bezdek *et al.*, 1984).

Fuzzy Rule Base

| Defuzzification | Defuzzification | Defuzzification | Fuzzy Inference | Engine | Fuzzy Inference | Fuzzy Infer

FIGURA 2.1 – R EPRESENTAÇÃO DE UM SISTEMA DIFUSO

FONTE: Turan e Yurdusev (2009)

A base de regras difusas contém critérios que incluem todas as possíveis relações difusas entre entradas e saídas; essas regras são expressas no formato *IF-THEN*, incluídas nos mecanismos de inferência difusa; o mecanismo de inferência mapeia amostras difusas em subconjuntos também difusos, trabalhando com a combinação das regras. A *desfuzzificação* converte as saídas difusas resultantes dos mecanismos de inferência em um número (Turan e Yurdusev, 2009).

2.4 ANÁLISE DE AGRUPAMENTO

Dentre as técnicas da estatística multivariada, encontra-se a análise de agrupamentos (Mingoti, 2005), que consiste em uma descoberta importante, sob aspectos conceituais e algorítmicos, para a análise de dados em diversas áreas (Sato-Ilic e Jain, 2006; Ezugwu *et al.*,

2022). Ezugwu *et al.* (2022) descrevem a análise de agrupamentos como uma técnica ativa de classificação de dados, de forma que os elementos pertencentes a um mesmo grupo apresentem características semelhantes entre si.

Em geral, o objetivo da análise de agrupamento é utilizar variáveis características dos dados para planejar a maneira mais adequada de agrupá-los em classes, impondo que os objetos similares estejam na mesma classe, reduzindo, portanto, a quantidade de observações de uma amostra (Bezdek *et al.*, 1984; Lattin *et al.*, 2011; Han *et al.*, 2012; Morettin e Singer, 2023). Portanto, o agrupamento é um método para categorizar os dados em classes e atribuir vetores de dados a estas classes, que são criadas de forma que estejam distantes umas das outras segundo as características de classificação, mas que as observações dentro delas estejam o próximas possível do ponto central deste grupo, chamado de centroide (Solomatine *et al.*, 2008; Lattin *et al.*, 2011; Han *et al.*, 2012), e que caracteriza cada um dos grupos formados (Morettin e Singer, 2023).

A análise de agrupamentos se insere ao conjunto de técnicas de aprendizagem não supervisionada – já citada na seção 2.2 –, na qual todas as variáveis são ditas como preditoras e previstas, sem distinção, na tentativa de identificar padrões entre as características que definem os dados de uma amostra. Dentre as categorias de algoritmos de agrupamento existentes, destacam-se os algoritmos combinatórios, que analisam os dados diretamente, não havendo referência à sua distribuição de probabilidades. Esta categoria, por sua vez, ainda se subdivide entre os algoritmos hierárquicos ou de partição (Morettin e Singer, 2023).

Assim, visto que se trata de uma técnica não supervisionada, uma dificuldade para sua aplicação está na definição dos números de grupos a serem utilizados, uma vez que é complexo caracterizar a similaridade entre os dados de uma mesma classe, e a dissimilaridade entre as classes (Kaviski, 1992; Ezugwu *et al.*, 2022). Ou seja, o número de grupos depende da semelhança entre os dados, tornando-se uma definição complexa (Bezdek *et al.*, 1984; Ren *et al.*, 2016; Yang e Nataliani, 2017). Para Ren *et al.* (2016), o modo mais simples de se determinar o número de grupos é a partir de uma visualização gráfica: se os dados da amostra puderem ser dispostos em um espaço bidimensional, o número de grupos pode ser inferido por observação. Porém, para amostras de mais dimensões, a abordagem não é aplicável (Ren *et al.*, 2016).

Ademais, segundo Lattin *et al.* (2011), existem inúmeras maneiras diferentes para se subdividir uma amostra em grupos. Não é computacionalmente viável buscar, por meio de todas as possíveis partições, aquela que se designe como a melhor dentre todas. Por isso, a análise de agrupamentos adota, em grande parte, uma abordagem heurística, baseada em algoritmos fáceis

de programar, extremamente eficientes e que forneçam soluções alinhadas ao objetivo (Lattin *et al.*, 2011).

As técnicas de agrupamento têm sido amplamente empregadas em diversas áreas, na estatística, no reconhecimento de padrões e no *Machine Learning* (Ezugwu *et al.*, 2022). No entanto, a escolha destas técnicas precisa ser feita com base nas peculiaridades de cada problema (Hair Junior *et al.*, 2009; Ezugwu *et al.*, 2022).

Um dos métodos mais populares de agrupamento e que se utiliza da lógica difusa, é o algoritmo *fuzzy c-means* (FCM), com aplicações em diversas áreas (Tao, 2002; Silva Filho *et al.*, 2015; Ren *et al.*, 2016; Yang e Nataliani, 2017). A técnica consiste em um método de classificação não supervisionado, em que, a partir de um número de grupos previamente definido, o algoritmo determina a qual grupo cada elemento pertence (Ren *et al.*, 2016; Yang e Nataliani, 2017).

No algoritmo FCM, a lógica difusa está inserida na definição de um parâmetro de incerteza, utilizado na determinação do grau de similaridade – ou pertencimento – dos eventos com cada grupo; o agrupamento é considerado difuso quando o grau de incerteza é maior que 1 (Yang e Nataliani, 2017). O intervalo recomendado por Ross (2004) para o parâmetro é de $1,25 \le m \le 2,00$.

2.5 APLICAÇÕES DA INTELIGÊNCIA ARTIFICIAL NA HIDROLOGIA

Redes Neurais, sistemas baseados em regras difusas e algoritmos genéticos são as técnicas mais populares da inteligência computacional (Solomatine *et al.*, 2008; Liang *et al.*, 2018; Ding *et al.*, 2020). Além destas técnicas, existem outros modelos inteligentes orientados a dados que têm produzido resultados adequados para problemas hidrológicos, como: regressão evolucionária (Solomatine *et al.*, 2008); teoria do caos e dinâmica não-linear (Solomatine *et al.*, 2008; Liang *et al.*, 2019); Máquinas de Vetores de Suporte (SVM – Support Vector Machines) (Solomatine *et al.*, 2008; Ding *et al.*, 2020); e os algoritmos Random Forest e Adaptive Neuro-Fuzzy Inference System (ANFIS) (Liang *et al.*, 2019).

A TABELA 2.1 sintetiza as pesquisas encontradas com aplicações de técnicas conhecidas como modelos inteligentes orientados a dados hidrológicos. As principais aplicações observadas são para previsão de vazões e de cheias.

TABELA 2.1 – PESQUISAS COM APLICAÇÕES NA HIDROLOGIA DE MODELOS INTELIGENTES ORIENTADOS A DADOS.

Autor	Ferramentas	Aplicação
Atiquzzaman e Kandasamy (2016)	ELM, ANN, EC-SVM	Previsão de vazões
Brêda et al. (2021)	ANN	Previsão de vazões
Chen et al. (2021)	SVM, ANN, LSTM, outros	Previsão de vazões
Desai e Ouarda (2021)	RF	Análise de frequência de cheias
Difi et al. (2023)	ELM, SVR, outros	Previsão de vazões
Ding et al. (2020)	LSTM, outros	Previsão de cheias
Ebtehaj e Bonakdari (2022)	ELM, outros	Previsão de cheias
Eddine et al. (2024)	SVR, outros	Previsão de vazões
Fathian <i>et al.</i> (2019)	ANN, MARS, RF, outros	Previsão de vazões
Feng et al. (2022)	ELM, outro	Previsão de vazões
Gizaw e Gan (2016)	SVR	Análise de frequência de cheias
Kim et al. (2022)	SVM, LSTM	Previsão de nível d'água
Li et al. (2019)	ELM, ELM-kernel, SVR, outros	Previsão de vazões
Lima e Scofield (2021)	ANN	Previsão de nível d'água
Lin et al. (2006)	ANN, SVM, ARMA	Previsão de vazões
Lohani et al. (2011)	lógica difusa, ANN, outros	Modelagem chuva-vazão
Lohani et al. (2012)	ANFIS, ANN, AR	Previsão de vazões
Moeeni et al. (2017)	SARIMA, ANFIS, ANN, outros	Previsão de vazões
Mohammadi et al. (2020)	ANFIS	Previsão de vazões
Niu e Feng (2021)	ANN, ANFIS, ELM, SVM, outro	Previsão de vazões
Ribeiro et al. (2020)	ELM, ARIMA, outros	Previsão de vazões
Samantaray et al. (2023a)	ANFIS-PSOSMA	Previsão de cheias
Samantaray et al. (2023b)	SVM, PSO-SVM, outro	Previsão de cheias
Schoppa et al. (2020)	RF	Previsão de cheias
Sharma et al. (2023)	SVM, RF, LSTM, MARS	Previsão de vazões
Shukla et al. (2022)	ANFIS	Previsão de vazões
Wu et al. (2019)	SVR	Previsão de cheias
Yaseen et al. (2016)	ELM, SVR, outro	Previsão de vazões
Yaseen et al. (2018b)	ANN, ELM	Previsão de vazões
Zakhrouf et al. (2023)	LSTM, outros	Previsão de vazões
Zhang et al. (2018)	ANN, outros	Previsão de vazões

ANFIS – Adaptive Nero-Fuzzy Inference System; ANN – Artificial Neural Network; AR – Modelo Autorregressivo; ARIMA – Modelo Autorregressivo Integrado de Médias Móveis; ARMA – Modelo Autorregressivo de Médias Móveis; EC-SVM – Evolutionary Computational-SVM; ELM – Extreme Learning Machine; LSTM – Long Short-Term Memory; MARS – Multivariate Adaptive Regression Splines; PSO-SVM – Particle Swarm Optimization-SVM; RF – Random Forest; SARIMA – Modelo Autorregressivo Integrado de Médias Móveis com Sazonalidade; SVM – Support Vector Machine SVR – Support Vector Regression.

Da Tabela 2.1, foram destacadas algumas pesquisas ao longo das últimas duas décadas. Lohani *et al.* (2011) fizeram uma comparação entre três métodos: *Artificial Neural Network* (ANN), lógica difusa, e uma função de transferência linear, além de investigarem o potencial de aplicação do modelo difuso de Takagi-Sugeno e seu impacto no modelo chuva-vazão, considerando as condições anteriores de umidade do solo. Os resultados mostraram que a modelagem baseada na lógica difusa apresenta melhor performance que as demais, inclusive, melhor que o ANN. Lohani *et al.* (2012), por sua vez, compararam os modelos ANFIS, ANN e AR para previsão de vazões afluentes em reservatório e destacaram que o modelo ANFIS apresentou maior acurácia, quando comparado aos demais, para previsão de vazões extremas.

Em seu estudo, Gizaw e Gan (2016) desenvolveram um modelo de análise regional de frequência de cheias baseado no SVR, que demonstrou resultados satisfatórios para estimar quantis de vazões (Q_{10} , Q_{25} , Q_{50} e Q_{100}) para duas bacias hidrográficas homogêneas, apesar das diferentes regiões climáticas.

Yaseen *et al.* (2018b), por sua vez, utilizaram os métodos ANN e ELM para demonstrar a confiabilidade e efetividade da previsão de vazões em diferentes escalas de tempo. ELM comparado ao ANN apresentou melhores resultados e menores tempos de processamento para a previsão, para todas as escalas. Para Yaseen *et al.* (2018b), comparando-o a outras técnicas da Inteligência Artificial, ELM oferece boas vantagens para a modelagem, como, por exemplo, a atribuição aleatória dos pesos internos dos parâmetros, rápido processamento da aprendizagem e uma arquitetura muito simples da rede neural.

Ribeiro et al. (2020) testaram os métodos citados na TABELA 2.1 para diferentes horizontes de previsão, em séries de dados de 5 hidrelétricas brasileiras. Os resultados mostraram que ELM performou melhor que os demais. Niu e Feng (2021) avaliaram 5 métodos, para dois grandes reservatórios de hidrelétricas na China, e, mais uma vez, ELM se destacou, apresentando melhor desempenho, assim como SVM, enquanto ANN e ANFIS tiveram resultados menos acurados. Ainda se ressalta que, há pelo menos uma década, autores já evidenciavam o SVM como uma das ferramentas mais efetivas para previsão, atuando como uma alternativa ao ANN (Guo *et al.*, 2011) e, atualmente, com o crescimento das técnicas de ELM, o SVM se mantém como uma ferramenta em destaque (Ibrahim *et al.*, 2022).

Ainda, na literatura, encontrou-se o trabalho de Yaseen *et al.* (2016), que elaboraram uma revisão bibliográfica a respeito das técnicas de Inteligência Artificial para modelos de previsão de vazões, com trabalhos de 2000 a 2015. Os autores categorizam as ferramentas em quatro classes distintas: (i) classificação e regressão baseada em séries temporais; (ii) conjuntos difusos; (iii) evolução computacional; e (iv) modelos híbridos da Inteligência Artificial.

Outras aplicações de modelos híbridos na Hidrologia podem ser encontradas em Yaseen *et al.* (2018a), Adnan *et al.* (2019), Kurian *et al.* (2020), Islam *et al.* (2021) e Ebthaj *et al.* (2022).

Especificamente para previsão de vazões, Adnan *et al.* (2019) afirmam que, nas últimas décadas, métodos como *Artificial Neural Networks* (ANN), *Adaptive Nero-Fuzzy Inference System* (ANFIS), *Multivariate Adaptive Regression Splines* (MARS), Máquinas de Vetores de Suporte (SVM), algoritmos genéticos (GA) e seus modelos híbridos têm demonstrado resultados de sucesso, pois ajudam a identificar a não-linearidade nos processos de formação de cheias.

2.6 PRINCIPAIS FERRAMENTAS PARA PREVISÃO DE VAZÕES

Notadamente, a partir da Tabela 2.1, a maioria das ferramentas aplicadas para previsão de vazões envolve ELM, com base em ANN, e SVR, partindo dos princípios do SVM. Neste contexto, são abordadas nas próximas subseções os principais conceitos relativos ao SVM e ao ELM, bem como à *Teoria do Caos* – inerente aos fenômenos hidrológicos.

2.6.1 Teoria do caos

Apesar da conotação de *desordem* ou *confusão* que a palavra *caos* pode sugerir, a ciência trata este conceito como "as mudanças dos fenômenos ocorridas ao longo do tempo", o que é identificado como a base da teoria (Williams, 1997).

Liang *et al.* (2019) apresentam um histórico do desenvolvimento da *Teoria do Caos*, desde os princípios desenvolvidos pelo cientista francês Henri Poincar, em 1903, perpassando pelos conceitos do "*atrator estranho*" de Ruelle e Takens, em 1976, que efetivamente contribuiu com o surgimento da teoria do caos pela utilização do expoente de Lyapunov, da entropia de Kolmogorov e da dimensão fractal, que viabilizaram as aplicações práticas.

Prever as tendências de evolução dos sistemas caóticos a partir de observações se tornou um desafio indispensável, uma vez que o movimento caótico é amplamente encontrado em vários sistemas não lineares complexos, incluindo a Hidrologia. Desta forma, Sivakumar (2017) explica que, tradicionalmente, duas abordagens têm sido amplamente empregadas em modelagens hidrológicas: determinística e estocástica; ambas são complementares uma à outra

para a maioria dos sistemas e processos hidrológicos, e possuem seus méritos diante dos resultados que apresentam. Contudo, identificar qual das abordagens é mais aplicável a cada sistema hidrológico não é tão relevante quanto conseguir acoplá-las, a ponto de aumentar suas vantagens e reduzir suas limitações para as situações de interesse. É neste contexto que os princípios da dinâmica determinística não linear e da teoria do caos podem ser úteis, a fim de preencher a lacuna existente (Sivakumar, 2017).

Sivakumar (2000) afirma que as aplicações da teoria do caos na hidrologia, inicialmente, tinham como principal objetivo investigar a existência do caos nos processos hidrológicos, encontrando-se poucas pesquisas relacionadas à previsão. O início da busca por evidências do caos na Hidrologia se dá em 1987, cujos resultados foram convincentes de que existe, de fato, caos em conjuntos de dados hidrológicos (Sivakumar, 2000), já que os processos geralmente exibem complexos padrões temporais e espaciais, possivelmente devido às suas interações altamente não lineares com o ambiente (Sivakumar e Singh, 2012). Ou seja, entendese que para se estabelecer o comportamento de um sistema complexo, deve-se estudar não apenas comportamentos isolados do problema, mas sim, a atuação completa dos elementos envolvidos (Sivakumar, 2017).

Após o início dos estudos do caos na Hidrologia, Sivakumar (2000) também apresenta uma série de trabalhos que deram maior evidência a esta linha de pesquisa. Porém, foi após 1996 que as aplicações se voltaram à modelagem hidrológica como forma de comparação com os resultados produzidos por modelos estocásticos.

De acordo com Yasmin e Sivakumar (2018), a modelagem de vazões é um dos problemas centrais na Hidrologia, uma vez que é fundamental para o planejamento e gerenciamento de recursos hídricos. No entanto, os desafios da área ainda são notáveis, visto que, mais uma vez, as vazões são resultado de interações complexas e não lineares entre variáveis climáticas e características da bacia, em uma larga escala temporal e espacial (Sivakumar, 2017).

Na literatura são encontrados exemplos de aplicações de modelos baseados na teoria do caos, como, por exemplo, no trabalho de Yu *et al.* (2004), no qual foi desenvolvida uma ferramenta denominada EC-SVM, que combina a teoria do caos e o SVM, utilizando um algoritmo de otimização (*evolutionary algorithm*) para estimação dos parâmetros do modelo. Para fins de comparação, outros modelos foram testados e o método proposto demonstrou melhor desempenho que os demais, incluindo a teoria do caos convencional.

Ainda, Tongal (2013) avalia a performance de dois modelos de previsão com base na teoria do caos: *k-nearest neighbor* (k-NN) e *feed-forward neural networks* (FFNN). Os métodos

foram aplicados ao rio mais extenso da Turquia (rio Kızılırmak). O método FFNN apresentou melhores resultados, principalmente quando se espera representar bem os picos de cheia. No entanto, os autores não realizaram comparação com métodos tradicionais.

Ghorbani *et al.* (2018) integraram a teoria do caos à técnica *Multi-Gene Genetic Programming* (MGGP), da Inteligência Artificial, a fim de formar um novo modelo de previsão de vazões, denominando-o como "Caos-MGGP" e comparando-o ao modelo de previsão local (LPM) e ao MGGP isoladamente. As previsões obtidas pelo modelo que acoplou a teoria do caos foram mais acuradas. Mihailović *et al.* (2019), por sua vez, investigaram o comportamento caótico e a previsibilidade da vazão diária do rio Brazos, no Texas. Portanto, tem-se que os conceitos e as técnicas relacionados à teoria do caos são utilizados como forma de subsidiar outros princípios e ferramentas.

2.6.2 SVM – Support Vector Machine

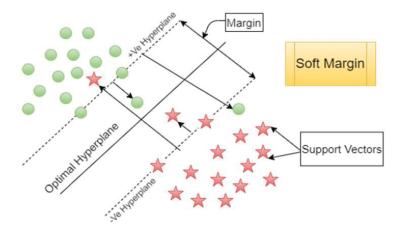
A abordagem das Máquinas de Vetores de Suporte tem se popularizado nas últimas quatro décadas como uma forma moderna de aprendizado estatístico, bem como, por sua velocidade de processamento e eficiência (Ibrahim *et al.*, 2022). Com base na literatura, Freitas (2016) apresentou um histórico do desenvolvimento do método SVM, desde os primeiros algoritmos para reconhecimento de padrões (na década de 50), até chegar ao seu formato mais atual, baseado na Teoria de Aprendizagem Estatística.

Máquinas de Vetores de Suporte (SVM – Support Vector Machines) consistem em um método de classificação baseado na teoria de Vapnik-Chervonenski, desenvolvido com o intuito de solucionar problemas de aprendizagem supervisionada e que pertence à família de classificação linear generalizada (Bhavsar e Panchal, 2012; Ibrahim et al., 2022). Cristianini e Shawe-Taylor (2000), por sua vez, definem as Máquinas de Vetores de Suporte como "sistemas de aprendizagem que utilizam um espaço hipotético de funções lineares em um espaço multidimensional, treinado com um algoritmo de aprendizagem baseado na teoria de otimização, que implementa uma tendência originada na teoria estatística de aprendizagem" (traduzido de Cristianini e Shawe-Taylor, 2000).

Em outras palavras, SVM é um método que divide a amostra de dados em duas classes por meio da determinação de um hiperplano definido por vetores de suporte, a partir da máxima margem de distância entre as classes (Tomar e Agarwal, 2015; Kim *et al.*, 2022), conforme representa a FIGURA 2.2: geometricamente, a distância entre os hiperplanos paralelos – denominada *margem* – é delimitada pelos vetores **x** (pontos) mais próximos; os vetores **x** são

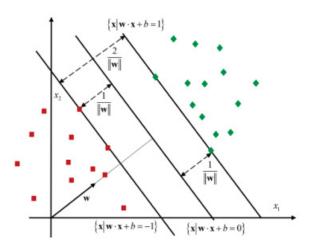
denominados de *vetores de suporte* e a margem é equivalente a $\frac{2}{\|\mathbf{w}\|}$ (Bhavsar e Panchal, 2012). A Figura 2.3 representa a margem de $\frac{2}{\|\mathbf{w}\|}$, onde \mathbf{w} indica o vetor p-dimensional, perpendicular ao hiperplano ótimo, indicado na Figura 2.2.

FIGURA 2.2 – REPRESENTAÇÃO CONCEITUAL DE MÁQUINA DE VETOR DE SUPORTE



FONTE: Ibrahim et al. (2022)

FIGURA 2.3 – HIPERPLANO IDEAL DA MÁQUINA DE VETORES DE SUPORTE QUE SEPARA DUAS CLASSES



FONTE: Roy e Chakraborty (2023)

Portanto, o interesse do SVM é minimizar o valor de **w** (Smola e Schölkopf, 2004), pois assume-se que quanto maior a margem entre os hiperplanos paralelos, menor será o erro de generalização do classificador (Lorena e Carvalho, 2007; Bhavsar e Panchal, 2012). Ou seja, o que torna *eficaz* um método, originalmente, de reconhecimento de padrões, como a SVM é

sua capacidade de generalização e de discriminação das classes: a minimização do erro estrutural é o que melhora a capacidade de generalização, enquanto a maximização da distância entre as margens contribui com a discriminação entre as classes (Freitas, 2016).

Para Moura *et al.* (2010), a minimização do erro estrutural da SVM é uma das razões pelas quais o método performa melhor que o ANN, por exemplo, pois este se baseia na minimização do risco empírico, que apenas reduz os erros de treinamento. Para Tomar e Agarwal (2015), esta característica no processo da SVM também possui a vantagem de solucionar problemas a partir de pequenas amostras.

Desde sua proposição, o método SVM vêm sendo aplicado a diversos tipos de dados com resultados satisfatórios para problemas de séries temporais em diversas áreas (Moura *et al.* 2010; Kim et al., 2022; Ibrahim *et al.*, 2022), inclusive na Hidrologia, como demonstraram os trabalhos apresentados na TABELA 2.1. Nos últimos anos, o método é aplicado com as finalidades de reconhecimento de padrões (classificação) e regressão (Moura *et al.*, 2010; Bhavsar e Panchal, 2012; Liang *et al.*, 2018).

Classificação

As Máquinas de Vetores de Suporte para classificação se enquadram na interseção de duas áreas de pesquisa: classificadores de grandes margens e métodos *kernel*. A classificação, por sua vez, pode ser linear ou não-linear (Bhavsar e Panchal, 2012). Para classificação linear, ainda existem as Máquinas de Vetores de Suporte com Margens Rígidas e com Margens Suaves (Lorena e Carvalho, 2007).

A equação do hiperplano é representada pela Equação (2.1) (Schölkopf e Smola, 2002; Lorena e Carvalho, 200; Bhavsar e Panchal, 2012; Freitas, 2016):

$$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b = 0,$$
 (2.1)
 $\mathbf{w} \in \mathcal{X}, b \in \mathbb{R}$

Onde:

 $f(\mathbf{x})$ é o hiperplano separador, de dimensão (p);

 \mathbf{x} é a representação vetorial de \mathbf{x} no espaço de características;

w é o vetor p-dimensional, perpendicular ao hiperplano separador;

b é o deslocamento que permite aumentar a margem de separação; parâmetro escalar.

Smola e Schölkopf (2003) apresentam um método para estimativa de *b*; a Equação (2.1) divide o espaço dos dados em duas regiões (uma maior que *zero*, e outra menor) (Lorena

e Carvalho, 2007), gerando as equações dos hiperplanos paralelos, representadas pelas Equações (2.2) e (2.3) (Schölkopf e Smola, 2002; Lorena e Carvalho, 2007; Bhavsar e Panchal, 2012; Freitas, 2016):

$$\mathbf{w} \cdot \mathbf{x} + b = +1 \tag{2.2}$$

$$\mathbf{w} \cdot \mathbf{x} + b = -1 \tag{2.3}$$

O problema de otimização a ser resolvido é equivalente à minimização da Equação (2.4) (Schölkopf e Smola, 2002):

$$\tau(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2 \tag{2.4}$$

Sujeito a:

$$y_i(\mathbf{w}.\mathbf{x}_i + b) - 1 \ge 0 \tag{2.5}$$

$$\|\mathbf{w}\|^2 = \mathbf{w} \cdot \mathbf{w} \tag{2.6}$$

Sendo que:

$$i = 1, ..., m;$$

Problemas reais apresentam alta complexidade (Freitas, 2016), devido à sua própria natureza, ou à existência de ruídos e *outliers* (Lorena e Carvalho, 2007), fazendo com que as SVMs precisem considerar (Bhavsar e Panchal, 2012; Freitas, 2016): (i) mais de duas variáveis preditoras; (ii) separação de dados a partir de curvas não-lineares; (iii) casos em que os dados não podem ser completamente separáveis entre as classes; (iv) manuseio de mais de duas classes.

Assim, tem-se os SVMs com margens suaves que se definem pela extensão dos conceitos dos SVMs de Margens Rígidas, com a introdução de variáveis de folga (ξ_i), em que i=1,...,m, sendo m o tamanho da amostra de treinamento.

A restrição ao problema de otimização, portanto, é apresentado na Equação (2.7) (Lorena e Carvalho, 2007; Freitas, 2016):

$$y_i(\mathbf{w}.\mathbf{x}_i + b) \ge 1 - \xi_i,\tag{2.7}$$

$$\xi_i \ge 0$$

A introdução das folgas faz com que alguns dados permaneçam entre os hiperplanos paralelos, uma vez que suaviza as margens do classificador linear, provocando erros de classificação, identificados por valor de folga superior a 1. A somatória das folgas representa o limite no número de erros no processo de treinamento (Lorena e Carvalho, 2007; Freitas, 2016).

Desta forma, o problema de otimização deixa de ser resolvido pela minimização da Equação (2.4) e passa a ser solucionado pela minimização da Equação (2.8) (Schölkopf e Smola, 2002; Lorena e Carvalho, 2007), onde *C* é um termo de regularização que "*impõe um peso à minimização dos erros no conjunto de treinamento em relação à minimização da complexidade do modelo*" (Lorena e Carvalho, 2007), valor constante que determina o grau de penalização de custo quando o erro de treinamento ocorre (Yu *et al.*, 2008):

$$\tau(\mathbf{w}, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + C(\sum_{i=1}^m \xi_i)$$
 (2.8)

No entanto, apesar do modo mais simples de dividir dois grupos ser por meio de uma linha reta, um plano ou um hiperplano *p*-dimensional, devido à eventual incidência de ruídos e *outliers*, torna-se mais efetiva a utilização de uma fronteira curva, denominada *região não-linear* (Schölkopf e Smola, 2002; Freitas, 2016); a SVM se comporta bem para a classificação de amostra de dados que *não* são classificados linearmente, por meio da teoria da função *kernel*, que transforma a amostra em um espaço multidimensional, por meio do cálculo do produto escalar no espaço de características (Schölkopf e Smola, 2002; Tomar e Agarwal, 2015; Freitas, 2016; Sharma *et al.*, 2023) (Equação 2.9):

$$k(\mathbf{x}_i, \mathbf{x}_i) \equiv \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_i) \tag{2.9}$$

Onde:

 \mathbf{x}_i e \mathbf{x}_j são os pontos recebidos pela função kernel no espaço de entradas.

Este processo torna possível a separação dos dados, de forma que os vetores de treinamento são mapeados no espaço dimensional superior pela função não-linear dada pela Equação (2.10) (Yu *et al.*, 2008), e a SVM encontra um hiperplano linear de separação, com a máxima margem no espaço dimensional superior (Hsu *et al.*, 2016).

$$f(\mathbf{x}) = \mathbf{w} \cdot \Phi(\mathbf{x}) + b \tag{2.10}$$

$$\Phi(\mathbf{x}) = \Phi(x_1, x_2) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$$
(2.11)

Onde:

$$i, j = 1, ..., m$$
.

A escolha da função *kernel* depende do problema a ser modelado (Bhavsar e Panchal, 2012). Exemplos de funções *kernel* são apresentados:

1. Linear: consiste em um caso especial da função *Kernel* RBF, e é o mais simples; na Equação (2.12), *c* é uma constante opcional (Bhavsar e Panchal, 2012; Hsu *et al.*, 2016):

$$k(\mathbf{x}_i, \mathbf{x}_i) = \mathbf{x}_i \cdot \mathbf{x}_i + c \tag{2.12}$$

2. Polinomial: função *Kernel* não-estacionária, adequada para problemas em que todos os dados de treinamento são normalizados; na Equação (2.13), γ, r e d são parâmetros de *Kernel* (Bhavsar e Panchal, 2012; Hsu *et al.*, 2016):

$$k(\mathbf{x}_i, \mathbf{x}_j) = \gamma (\mathbf{x}_i, \mathbf{x}_j + r)^d,$$

$$\gamma > 0$$
(2.13)

3. Funções de base radial (RBF) ou Gaussiano: é a principal função *Kernel*, uma vez que (i) permite mapear amostras não lineares em um espaço dimensional superior; (ii) possui menos parâmetros a serem calibrados do que função polinomial; (iii) apresenta menor dificuldade numérica (Bhavsar e Panchal, 2012; Freitas, 2016; Hsu *et al.*, 2016):

$$k(\mathbf{x}_{i}, \mathbf{x}_{j}) = \exp(-\gamma \|\mathbf{x}_{i} - \mathbf{x}_{j}\|^{2}),$$

$$\gamma > 0$$
(2.14)

4. Sigmoidal: caso especial da função Kernel RBF (Freitas, 2016):

$$k(\mathbf{x}_i, \mathbf{x}_i) = \tanh(\gamma \mathbf{x}_i, \mathbf{x}_i + r)$$
(2.15)

Apesar do SVM poder utilizar as variadas funções, Barzegar *et al.* (2017) afirmam que a função RBF performa melhor que as demais.

Regressão

A regressão pode ser entendida como um processo ligeiramente mais genérico do que a classificação (reconhecimento de padrões). Ao invés de lidar com saídas binárias ($y \in Y, Y = \{-1; +1\}$), são estimados valores reais ($y \in \mathbb{R}$) (Schölkopf e Smola, 2002).

Para isso, uma situação análoga ao problema de margens suaves é construída no espaço dos valores de saída, a partir da função de custo ε -indiferente de Vapnik, definida na Equação (2.16) (Schölkopf *et al.*, 2000; Schölkopf e Smola, 2002), que não penaliza os erros abaixo de $\varepsilon > 0$, definido *a priori*. Na Equação (2.16), $f(\mathbf{x})$ é dada pela Equação (2.10) (não-linear).

$$|y - f(\mathbf{x})|_{\varepsilon} = \max\{0, |y - f(\mathbf{x})| - \varepsilon\}$$
(2.16)

Formalmente, o problema de regressão pode ser expresso como um problema de otimização convexo, pela minimização da função da Equação (2.17) (Smola e Schölkopf, 2003; Yu *et al.*, 2008).

$$\tau(\mathbf{w}, \xi, \xi^*) = \frac{1}{2} \|\mathbf{w}\|^2 + C \left(\sum_{i=1}^m (\xi_i + \xi_i^*) \right)$$
 (2.17)

Sujeita a:

$$y_{i} - f(\mathbf{x}_{i}) \leq \varepsilon + \xi_{i}$$

$$f(\mathbf{x}_{i}) - y_{i} \leq \varepsilon + \xi_{i}^{*}, \qquad (2.18)$$

$$\xi_{i}, \xi_{i}^{*} \geq 0$$

$$i = 1, 2, ..., m$$

Onde:

 ξ_i é a variável de folga que especifica o maior erro de treinamento, sujeito à tolerância; ξ_i^* é a variável de folga que especifica o menor erro de treinamento, sujeito à tolerância; ε é a tolerância;

C é a constante que define o grau de penalização de custo quando há erro de treinamento, ou constante de regularização.

Introduzindo os multiplicadores de Lagrange (α_i^* e α_i), a estimativa é dada pela Equação (2.19), para regressão não-linear (Schölkopf e Smola, 2002; Yu *et al.*, 2008):

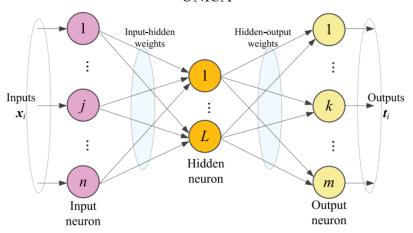
$$f(x) = \sum_{i=1}^{m} (\alpha_i^* - \alpha_i) k(x_i, x) + b$$
 (2.19)

Da mesma forma que a classificação, os princípios da regressão podem ser generalizados a uma estimativa não linear, por meio das funções *kernel*.

2.6.3 ELM – Extreme Learning Machine

ELM pode ser definida como um algoritmo de treinamento para uma rede neural direta de camada oculta única (SLFN) (Wang *et al.*, 2022), conforme Figura 2.4.

FIGURA 2.4 – ESTRUTURA DE REDE NEURAL DIRETA DE CAMADA OCULTA ÚNICA



FONTE: Feng et al. (2022)

A baixa velocidade de aprendizado demonstrada pelas redes neurais diretas (feedforward neural networks) foi uma das motivações de Huang et al. (2004) quando propuseram a ELM; os autores atribuem duas principais razões para esta característica: (i) algoritmos lentos de aprendizado baseados em gradiente decrescente são extensivamente utilizados para treinar redes neurais, e (ii) todos os parâmetros da rede são ajustados iterativamente utilizando tais algoritmos. Desta forma, uma vantagem da ELM é, justamente,

sua convergência mais rápida do que nas redes neurais tradicionais, além de sua promessa de melhor performance (Huang *et al.*, 2004; Alizamir *et al.*, 2018; Wang *et al.*, 2022), comprovada pelos trabalhos apresentados na TABELA 2.1 que fizeram uso da ELM.

As Redes Neurais Artificiais, inspiradas na rede neural biológica, buscam reproduzir, de uma maneira simplificada, o cérebro humano, assemelhando-se em dois aspectos: (i) o conhecimento é adquirido por uma rede através de um processo de aprendizagem, e (ii) as forças de conexão do interneurônio, conhecidas como pesos sinápticos, são utilizadas para armazenar o conhecimento adquirido (Sun *et al.*, 2010). Então, as redes neurais são compostas por neurônios, conexões e pesos: um neurônio recebe sinais de entrada de um neurônio antecedente e envia sinais de saída à neurônios subsequentes; os neurônios são organizados em diferentes camadas (*layers*), que apresentam inúmeras conexões entre si; cada conexão transporta um parâmetro denominado peso (*weight*), que define *como* um neurônio processa os sinais de entrada para gerar os sinais de saída (Wang *et al.*, 2022). A solução das redes neurais parte da definição dos pesos.

Diferentemente do que ocorre com os algoritmos de treinamento que se baseiam no decréscimo do gradiente, a ELM não utiliza iterações durante o aprendizado; seu funcionamento se baseia na definição de matrizes aleatórias que representam os pesos entre as camadas de entrada e oculta, e os vieses (*biases*) da camada oculta. Desta forma, sua solução é analítica e obtida por meio do método Moore-Penrose inversamente generalizado, que resolve um sistema linear. As funções não-lineares de ativação na camada oculta agregam a não-linearidade ao sistema, tornando possível considerá-lo linear e acelerando sua convergência (Huang *et al.*, 2004; Alizamir *et al.*, 2018; Yaseen *et al.*, 2020; Ebtehaj e Bonakdari, 2022; Feng *et al.*, 2022; Wang *et al.*, 2022).

Desta forma, a análise teórica mostrou que ELM possui maior probabilidade de alcançar a solução ótima global com parâmetros aleatórios, do que as redes neurais tradicionais com todos os parâmetros a serem treinados (Huang *et al.*, 2004; Wang *et al.*, 2022); a solução das redes neurais tradicionais – baseadas nos algoritmos de gradiente decrescente – nem sempre retornam a melhor solução global, devido aos parâmetros de inicialização e à complexidade do espaço amostral. Nestes casos, o algoritmo tende a convergir para os mínimos locais, ao invés dos globais (Huang *et al.*, 2004; Wang *et al.*, 2022); a convergência para os mínimos globais da ELM representa sua promessa de melhor performance, o que resulta em uma alta capacidade de generalização do método (Huang *et al.*, 2004; Ebtehaj e Bonakdari, 2022).

Atiquzzaman e Kandasamy (2016), por sua vez, destacam outras vantagens da ELM, além daquelas já citadas, como (i) sua aplicabilidade à muitas funções não-lineares de ativação

e às funções *kernel*, (ii) sua eficiência para aplicações online e em tempo real, e (iii) sua viabilidade de aplicação em computação de larga escala e no *Machine Learning*. Ebtehaj e Bonakdari (2022) evidenciam ainda a limitada intervenção dos usuários e a robustez da ELM.

Apesar dos bons resultados obtidos por meio da ELM, Yaseen *et al.* (2020) afirmam que o algoritmo possui suas limitações e desvantagens, como, por exemplo, a própria aleatoriedade na definição dos pesos entre as camadas de entrada e oculta e das tendências, que podem afetar negativamente a capacidade de generalização da rede. Por isso, muitos autores têm optado pela utilização de modelos híbridos, que se utilizam de algoritmos de otimização para definição dos parâmetros.

Dada uma amostra de treinamento $\{(x_1, y_1), (x_2, y_2), ..., (x_i, y_i), ..., (x_n, y_n), \}$, sendo $X = [x_1, ..., x_i, ..., x_n]^T$ o vetor de entrada e $Y = [y_1, ..., y_i, ..., y_n]^T$ o vetor de resposta do modelo, a função básica de saída da ELM para uma SLFN pode ser expressa pela Equação (2.20) (Huang *et al.*, 2004; Li *et al.*, 2019):

$$f(x_i) = \sum_{i=1}^{L} \beta_i h_i(x_i) = \mathbf{h}(\mathbf{x}_i) \boldsymbol{\beta}$$
 (2.20)

$$\beta = \left[\beta_1, \dots, \beta_j, \dots, \beta_L\right]^T \tag{2.21}$$

$$h(x_i) = [h_1(x_1), \dots, h_i(x_i), \dots, h_L(x_1)]$$
(2.22)

$$i = 1, \dots, n \tag{2.23}$$

Onde:

L é o número de neurônios na camada oculta;

 β é o vetor de pesos entre as camadas oculta e de saída, expresso pela Equação (2.21); $h(x_i)$ é o vetor de saída da camada oculta relacionado ao vetor de entrada X, expresso pela Equação (2.22);

O vetor $h(x_i)$ corresponde à função de ativação do SLFN. Portanto, $h_j(x_i)$ pode ser escrito conforme Equação (2.24):

$$h_j(x_i) = g(\mathbf{w}_j \cdot \mathbf{x}_i + b_j) \tag{2.24}$$

$$\mathbf{w}_{j} = \left[w_{j1}, \dots, w_{ji}, \dots, w_{jn}\right]^{T} \tag{2.25}$$

Onde:

 \mathbf{w}_j é o vetor aleatório de pesos que conecta o *j*-ésimo neurônio da camada oculta aos neurônios da camada de entrada, expresso pela Equação (2.25);

 $\mathbf{w}_j \cdot \mathbf{x}_i$ representa o produto interno de \mathbf{w}_j e \mathbf{x}_i do j-ésimo neurônio da camada oculta; b_j é a tendência aleatória do j-ésimo neurônio da camada oculta.

Cada variável de entrada x_i corresponde a uma equação, então as n equações podem ser escritas por:

$$H\beta = Y \tag{2.26}$$

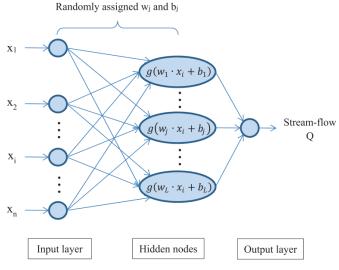
Onde:

H é a matriz correspondente à matriz de saída da camada oculta, expressa pela Equação(2.27):

$$\boldsymbol{H} = \begin{bmatrix} h_1(x_1) & \dots & h_L(x_1) \\ \vdots & \ddots & \vdots \\ h_1(x_n) & \dots & h_L(x_n) \end{bmatrix} = \begin{bmatrix} g(\boldsymbol{w}_1 \cdot \boldsymbol{x}_1 + b_1) & \dots & g(\boldsymbol{w}_L \cdot \boldsymbol{x}_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(\boldsymbol{w}_1 \cdot \boldsymbol{x}_n + b_1) & \dots & g(\boldsymbol{w}_L \cdot \boldsymbol{x}_n + b_L) \end{bmatrix}$$
(2.27)

A arquitetura da ELM é similar à de SLFN, porém a diferença está na atribuição dos valores aleatórios à \mathbf{w}_i e b_i na ELM; a estrutura da ELM é apresentada na Figura 2.5:

FIGURA 2.5 – ARQUITETURA BÁSICA DO MÉTODO *EXTREME LEARNING MACHINE*



FONTE: Li et al. (2019)

Para o desenvolvimento da ELM, o modelo exige a definição das entradas, a determinação do número de neurônios na camada oculta e a seleção de uma função de ativação adequada e infinitamente diferenciável.

3 MATERIAIS E MÉTODOS

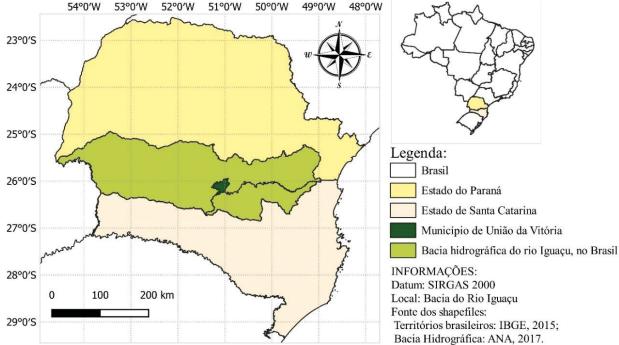
No Capítulo 3, são apresentados os processos metodológicos para desenvolvimento da modelagem hidrológica baseada em agrupamento de cheias, bem como, a área de estudo para sua aplicação.

3.1 ÁREA DE ESTUDO

A região de estudo definida compreende a bacia hidrográfica do rio Iguaçu, considerando como exutório a Estação Fluviométrica União da Vitória (65310000), devido aos recorrentes eventos de cheia registrados historicamente na bacia, principalmente na região do município de União da Vitória, localizado na região sudeste do estado do Paraná. A Figura 3.1 apresenta a localização de União da Vitória na bacia hidrográfica do rio Iguaçu.

FIGURA 3.1 – LOCALIZAÇÃO DO MUNICÍPIO DE UNIÃO DA VITÓRIA, NO ESTADO DO PARANÁ E NA BACIA DO RIO IGUAÇU

54°0′W 53°0′W 52°0′W 51°0′W 50°0′W 49°0′W 48°0′W



O município se desenvolveu às margens do rio Iguaçu, que propiciou o desenvolvimento socioeconômico da região por meio do transporte de passageiros e

mercadorias, da pesca, do lazer, da dessedentação e da geração de energia (Rocha, 2012). A população de União da Vitória é de aproximadamente 55.000 habitantes, distribuídos em 720 km² de território, com população majoritariamente urbana (IBGE, 2022). Segundo a SEMA (2010), toda a área do município está localizada na bacia do rio Iguaçu, representando 1,4% de toda a área da bacia (em seus três trechos).

Desta forma, com a ocorrência das grandes cheias do rio, os impactos gerados sobre a cidade são significativos para a economia local, interrompendo as atividades que a suportam. A população, da mesma forma, por se concentrar na zona urbana, sofre com as consequências oriundas de eventos extremos máximos ocorridos no rio Iguaçu, uma vez que suas casas, atingidas pela água, podem ficar longos períodos inundadas, gerando a perda de bens de consumo e demandando a evacuação das residências atingidas.

Também é importante ressaltar que em União da Vitória (no Paraná), assim como em Porto União (em Santa Catarina), houve uma expansão da área urbana das cidades em direção à planície de inundação do rio Iguaçu (COBRAPE, 20--), agravando as condições da população diante dos eventos de cheia.

A população de toda a bacia do Iguaçu é estimada em 4,5 milhões de habitantes, dos quais cerca de 2,5 milhões vivem na Região Metropolitana de Curitiba (SEMA, 2010), onde se observa intensa atividade industrial, comercial e de prestação de serviços, e, consequentemente, áreas predominantemente urbanizadas (SEMA, 2010; IAT, 2020a). A jusante, grande parcela da área da bacia é ocupada com a classe de uso misto, aparecendo áreas de agricultura intensiva na região de Guarapuava, seguindo no sentido Sudoeste do estado do Paraná, até a divisa com o estado de Santa Catarina; nas regiões do Médio e Baixo Iguaçu, observa-se uma concentração de cobertura florestal (SEMA, 2010).

Ainda, municípios como Porto Amazonas, São Mateus do Sul e Rio Negro, abrangem regiões importantes na bacia, com contribuições significativas para o Iguaçu. Nestes locais, observa-se que a classificação de uso e ocupação do solo identificou uma área urbana de 0,8%, 1,00% e 2,40%, respectivamente, com grandes áreas destinadas à agricultura (IAT, 2020b).

A bacia hidrográfica do rio Iguaçu possui uma área de drenagem total de 70.800 km², dividida entre os estados do Paraná e Santa Catarina, no Brasil, e a Argentina, com uma pequena parcela do território da bacia (Rocha, 2012). Ainda, há uma subdivisão desta área estabelecida pela Resolução Nº 49/2006/CERH/PR, definindo as três Unidades Hidrográficas de Gestão de Recursos Hídricos do Iguaçu: Baixo Iguaçu; Médio Iguaçu; e Alto Iguaçu, agrupado à Bacia do Ribeira.

Juntos, os trechos Alto e Médio Iguaçu contemplam a maior parte da área total da bacia, sendo que na estação fluviométrica União da Vitória (65310000), situada no trecho Médio, a área de drenagem é de, aproximadamente, 24.200 km² (ANA, 2023), mais de um terço da área de toda a bacia. Ainda, na referida estação, segundo COPEL (1995) *apud* Rocha (2012), a vazão média é de cerca de 500 m³/s (20 L/s/km²), tornando-se de quatro a seis vezes maior nos períodos de cheia; na maior cheia registrada na região, em 1983, a vazão de pico atingiu 5.157 m³/s (210 L/s/km²), quase 10 vezes maior que a vazão média da estação.

3.1.1 Coleta e análise de dados

Cruzando praticamente todo o estado do Paraná, no sentido Leste-Oeste, com uma extensão de 1.320 km, o rio Iguaçu deságua no rio Paraná; localiza-se entre os paralelos 25° 00' e 27° 00' de latitude Sul e os meridianos 49° 00' e 54° 30' de longitude Oeste, e apresenta um relevo bastante irregular, com baixas declividades e vales amplos, em seu trecho superior, e altas declividades e vales rochosos estreitos, nos trechos médio e inferior (Mine e Tucci, 2002; Pinto *et al.*, 1976).

O rio Iguaçu é formado pelo encontro dos rios Iraí e Atuba, na Região Metropolitana de Curitiba (SEMA, 2010; IAT, 2023), onde ainda são observados, à margem direita do Iguaçu, os rios Barigui e Passaúna, e os rios Pequeno, Miringuava e Cutia, à margem esquerda. A jusante, até União da Vitória, tem-se como principais contribuintes do Iguaçu, o rio Potinga, à margem direita, e os rios Negro e Timbó, à margem esquerda. Outros afluentes, como os rios Verde, Itaqui, da Vargem, Claro (à margem direita), Passa Dois e Paciência (à margem esquerda), representam contribuições menores ao Iguaçu, a montante de União da Vitória.

A partir dos afluentes do rio Iguaçu identificados, foram investigadas as estações fluviométricas existentes na bacia; dentre elas, foram selecionadas para posterior análise as estações apresentadas na Tabela 3.1. O critério de seleção priorizou as estações nos afluentes com maiores áreas de drenagem, bem como, a consistência dos dados e o tamanho da amostra (Tozzi e Fill, 2020).

Observa-se da Tabela 3.1, que as estações fluviométricas dos formadores do rio Iguaçu (Iraí e Atuba) não foram utilizadas na análise, visto que apresentaram dados não consistidos, com longos períodos de falhas, e/ou tamanhos inadequados para as amostras; na Região Metropolitana de Curitiba, avaliou-se unicamente a Estação Fluviométrica Fazendinha (65010000), no rio Pequeno, de 1976 a 2014. No rio Potinga, por sua vez, encontrou-se a Estação Fluviométrica Pontilhão (65200000), com dados consistidos de 1940 a 1964, que,

apesar de sua contribuição representativa para o rio Iguaçu, foi descartada das análises devido aos longos períodos de falha na série histórica.

TABELA 3.1 – PRINCIPAIS ESTAÇÕES FLUVIOMÉTRICAS IDENTIFICADAS A MONTANTE DE UNIÃO DA VITÓRIA

Código	Nome	Rio	Cidade	UF	Lat. (S)	Long. (W)	Alt. (m)	A (km²)	Início-fim dos dados
65010000	Fazendinha	Pequeno	São José dos Pinhais	PR	-25,5192	-49,1456	875,1	106	1956-2014
65027000	Rodeio	Verde	Balsa Nova	PR	-25,5797	-49,5831	859,6	231	1973-2014
65035000	Porto Amazonas	Iguaçu	Porto Amazonas	PR	-25,5481	-49,8886	780,0	3620	1936-2005
65060000	São Mateus do Sul	Iguaçu	São Mateus do Sul	PR	-25,8758	-50,3897	745,0	6050	1931-2014
65100000*	Rio Negro	Negro	Rio Negro	PR	-26,1097	-49,8011	766,0	3450	1931-2015
65155000**	São Bento	da Várzea	Lapa	PR	-25,9456	-49,7914	750,0	2000	1931-2015
65180000	Salto Canoinhas	Canoinhas	Major Vieira	SC	-26,3722	-50,2881	793,0	772	1981-2014
65220000	Fluviópolis	Iguaçu	São Mateus do Sul	PR	-26,0192	-50,5925	770,0	18600	1964-2015
65295000*	Santa Cruz do Timbó	Timbó	Irineópolis	SC	-26,3839	-50,8783	750,0	2610	1975-2004
65310000	União da Vitória	Iguaçu	União da Vitória	PR	-26,2281	-51,0803	670,0	24200	1931-2015

Lat. (S) – latitude Sul; Long. (W) – latitude Oeste; Alt. – altitude; A – área de drenagem.

FONTE: ANA (2023)

Ainda, dentre as Estações Fluviométricas da Tabela 3.1, observou-se que Santa Cruz do Timbó (65295000), São Bento (65155000) e Rio Negro (65100000) apresentaram períodos de falhas. Do período de dados indicado na Tabela 3.1, foram considerados apenas os anos completos de cada estação, exceto para a Estação São Bento (65155000), cuja série foi preenchida e estendida (até 2015) por meio da curva de permanência da Estação Fluviométrica Rio da Várzea dos Lima (65135000), a montante de São Bento, ambas no rio da Várzea.

Considerando as diferentes magnitudes das áreas de drenagem das Estações São Bento (65155000) e Rio da Várzea dos Lima (65135000) – 2000 e 605 km², respectivamente – e, consequentemente, de suas descargas médias, as curvas de permanência foram representadas em termos da relação entre vazão e vazão média da estação, para os períodos simultâneos de dados. As curvas de permanência são apresentadas na Figura 3.2.

Ainda, vale ressaltar que União da Vitória está sujeita ao remanso oriundo do reservatório da Usina Hidrelétrica de Foz do Areia, a jusante no rio Iguaçu, cuja implantação se deu em 1980. Por isso, Steffen e Gomes (2018) corrigiram os dados de vazão de União da Vitória, do período de 1980 a 2015, por meio dos dados de nível d'água da estação fluviométrica R5 Porto Vitória (65365800), localizada entre União da Vitória e Foz do Areia.

^{*} Estações com falhas; ** Estação preenchida

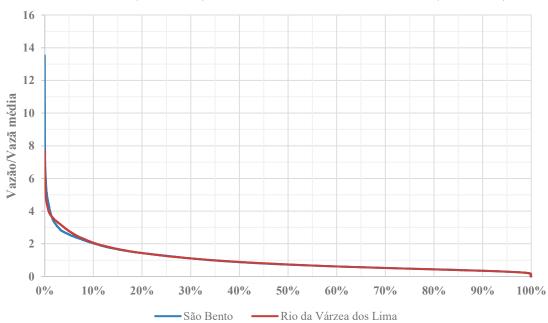


FIGURA 3.2 – CURVAS DE PERMANÊNCIA DAS ESTAÇÕES FLUVIOMÉTRICAS SÃO BENTO (65155000) E RIO DA VÁRZEA DOS LIMA (65135000)

Portanto, considerando apenas os anos completos de dados disponíveis, foram calculados os parâmetros estatísticos das séries de médias diárias de vazões, para cada uma das estações selecionadas. A estatística descritiva é apresentada na TABELA 3.2.

TABELA 3.2 – ESTATÍSTICA DESCRITIVA DAS ESTAÇÕES FLUVIOMÉTRICAS ENCONTRADAS

Código	Nome da estação	n (anos)	μ (m³/s)	q (L/s/km²)	Mín (m³/s)	Máx (m³/s)	σ (m ³ /s)	γ
65010000	Fazendinha	54	2,8	26,4	0,4	21,2	2,31	2,19
65027000	Rodeio	41	4,8	20,7	0,2	49,0	4,65	3,03
65035000	Porto Amazonas	70	68,3	18,9	3,4	1167,0	63,07	3,83
65060000	São Mateus do Sul	84	107,0	17,7	11,7	1533,1	94,68	3,37
65100000	Rio Negro	72	73,0	21,2	7,7	953,9	68,70	3,24
65155000*	São Bento	84	35,8	17,9	6,0	482,3	28,82	2,84
65180000	Salto Canoinhas	34	18,9	24,5	0,8	254,9	22,79	2,83
65220000	Fluviópolis	51	399,7	21,5	34,5	3819,4	346,46	2,53
65295000	Santa Cruz do Timbó	26	78,7	30,2	4,5	706,0	82,61	2,22
65310000	União da Vitória	84	492,2	20,3	46,2	5156,7	439,53	2,51

n – tamanho da série; μ – média; q – vazão específica média; σ – desvio-padrão; γ – assimetria.

A fim de verificar a necessidade de descarte de estações fluviométricas nas análises posteriores, foi realizada a análise de correlações das vazões médias diárias de União da Vitória (65310000) com as vazões médias diárias observadas nas estações fluviométricas a montante,

^{*} Valores correspondentes à série preenchida

para diferentes *lags* de tempo; para cada estação fluviométrica, foram destacadas as maiores correlações encontradas.

TABELA 3.3 – CORRELAÇÕES ENTRE AS SÉRIES DE VAZÕES MÉDIAS DIÁRIAS DE UNIÃO DA VITÓRIA (65310000) E AS DAS ESTAÇÕES A MONTANTE, PARA DIFERENTES *LAGS* DE TEMPO

	Estação Fluviométrica									
lag (dia)	65010000	65027000	65035000	65060000	65100000	65155000	65180000	65220000	65295000	
	Fazendinha	Rodeio	Porto Amazonas	São Mateus do Sul	Rio Negro	São Bento	Salto Canoinhas	Fluviópolis	Santa Cruz do Timbó	
0	0,5406	0,6015	0,7845	0,8988	0,8414	0,8239	0,8192	0,9781	0,9359	
1	0,5706	0,6331	0,8081	0,8968	0,8575	0,8329	0,8380	0,9697	0,9450	
2	0,5875	0,6521	0,8190	0,8871	0,8620	0,8318	0,8430	0,9525	0,9340	
3	0,5962	0,6623	0,8217	0,8718	0,8591	0,8244	0,8398	0,9289	0,9148	
4	0,5990	0,6663	0,8190	0,8522	0,8506	0,8127	0,8309	0,9007	0,8904	
5	0,5974	0,6658	0,8117	0,8294	0,8375	0,7977	0,8173	0,8691	0,8625	

Na Tabela 3.3, observa-se que Fluviópolis (65220000) atingiu a maior correlação com União da Vitória, dentre todas as estações e para todos os *lags* – inclusive para o *lag 0* –, mas sua correlação diminui gradativamente conforme aumenta a defasagem; localizada no próprio rio Iguaçu, Fluviópolis apresenta a maior área de drenagem dentre as estações avaliadas e é a segunda estação mais próxima de União da Vitória, sendo Santa Cruz do Timbó (65295000) a estação com menor distância até o exutório em estudo. Situada no rio Timbó, afluente do rio Iguaçu, Santa Cruz do Timbó (65295000) obteve a segunda maior correlação, com máximo para o *lag 1*, e, apesar da maior proximidade com União da Vitória, foi descartada das análises por apresentar a série de dados mais curta (1975 a 2004), com falhas que incluem o período da cheia de 1983 – maior evento histórico em União da Vitória.

Em geral, quanto maior a distância entre a estação e União da Vitória, maior o *lag* onde se observa a correlação mais alta, o que se justifica pelo tempo de viagem da vazão entre a estação e União da Vitória. Em São Mateus do Sul (65060000) esta situação não se verificou, o que é explicado por sua localização no próprio rio Iguaçu e por sua maior proximidade com Fluviópolis (65220000).

Porto Amazonas (65035000), também localizada no rio Iguaçu, apresentou máxima correlação para o *lag 3*, com valor inferior a praticamente todas as correlações das demais estações, com exceção de Fazendinha (65010000) e Rodeio (65027000). Porto Amazonas, apesar da extensa série de dados (70 anos), não apresentou informações observadas após 2005 (Tabela 3.1); dentre as estações do rio Iguaçu, esta é a estação mais distante de União da Vitória.

As séries das Estações Fluviométricas Fazendinha (65010000) e Rodeio (65027000) estão localizadas em afluentes do rio Iguaçu, situadas a uma maior distância de União da

Vitória, apresentaram menores correlações para todos os *lags* de tempo (inferiores a 0,67). As referidas estações representam as menores contribuições (pequenas áreas de drenagem) do rio Iguaçu dentre os afluentes analisados.

A estação Salto Canoinhas (65180000), apesar de seus valores de correlação com União da Vitória (Tabela 3.3), foi descartada das análises por possuir a segunda menor série de dados (Tabela 3.2) e, portanto, reduzido período simultâneo entre as estações e área de estudo.

Rio Negro (65100000) apresentou maior correlação com União da Vitória para o *lag* 2, o que se justifica pela distância entre os postos e por sua localização em um afluente do rio Iguaçu (rio Negro). No entanto, apesar da maior defasagem observada, Rio Negro (65100000) obteve correlações superiores à São Bento (65155000) para todos os *lags*, inclusive para o *lag* 1, no qual São Bento apresentou sua maior correlação com União da Vitória. Além de sua distância até União da Vitória, a estação São Bento se localiza no rio da Várzea, afluente do rio Negro, o que justifica a menor correlação. Reforça-se que a correlação de São Bento (65155000) corresponde à sua série preenchida, que, após o preenchimento das falhas, passou a ter uma série de vazões de mesma extensão que União da Vitória (86 anos).

Portanto, a montante de União da Vitória, as estações Fazendinha (65010000), Rodeio (65027000), Porto Amazonas (65035000), São Mateus do Sul (65060000), Rio Negro (65100000), São Bento (65155000) e Fluviópolis (65220000) foram selecionadas para o prosseguimento das análises. Devido às diferentes magnitudes das séries, o conjunto de dados foi, primeiramente, normalizado pelo método de padronização linear (Equação 3.1) (Wang *et al.*, 2015; Steffen e Gomes, 2018; Khosravi *et al.*, 2021; Mosavi *et al.*, 2021), a fim de tornar os dados adimensionais.

$$Q_{ij} = \frac{Q_{i,min}^0 - Q_{j,min}^0}{Q_{j,mix}^0 - Q_{j,min}^0}$$
(3.1)

Onde:

 Q_{ij} é a vazão tratada da i-ésima observação da j-ésima estação;

 Q^{θ}_{ij} é a vazão não tratada da *i*-ésima observação da *j*-ésima estação;

 $Q_{j,m\acute{a}x}^0$ é a máxima vazão não tratada da *j*-ésima estação;

 $Q_{j,min}^0$ é a mínima vazão não tratada da *j*-ésima estação.

Após, as séries das referidas estações foram submetidas à Análise dos Componentes Principais, resultando em uma redução no número de postos a serem considerados no processo de modelagem (Betancur *et al.*, 2020). A técnica da estatística multivariada teve como objetivo encontrar aquelas que mais representam um determinado conjunto de dados, neste caso, composto pelas séries de médias diárias das estações fluviométricas, para o período simultâneo entre elas.

Para realização da Análise dos Componentes Principais, foram utilizadas as funções *PCA* e *eig* das bibliotecas *sklearn* e *numpy*, respectivamente, do Python, para obtenção dos autovalores e autovetores. Os autovalores e autovetores, bem como a variância explicada acumulada das estações, são apresentados nas Tabelas 3.4 e 3.5. Na Tabela 3.4, os valores em destaque indicam a estação mais representativa de cada componente principal, identificada pelo maior valor, em módulo. Por meio das referidas tabelas, observa-se que Fluviópolis (65220000) detém a maior representatividade da amostra, com 79%, enquanto São Bento (65155000) e Rio Negro (65100000) representam 5 e 10%, respectivamente; as demais estações possuem menos de 3% de representatividade cada. Acumulando-se estas porcentagens, tem-se que Fluviópolis, São Bento e Rio Negro representam, juntas, quase 95% da amostra. Portanto, as três estações foram selecionadas para utilização neste trabalho.

TABELA 3.4 – AUTOVALORES E AUTOVETORES DA ANÁLISE DOS COMPONENTES PRINCIPAIS DAS ESTAÇÕES

Variân explica		79,42%	10,01%	5,52%	2,30%	1,49%	0,63%	0,63%	$\Sigma = 100,00\%$
AUTO	VALORES	0,0390	0,0049	0,0027	0,0011	0,0007	0,0003	0,0003	$\Sigma = 0.0491$
7.0	65220000	0,4668	0,4251	-0,3114	-0,3855	0,3389	-0,4347	-0,2278	
TORES	65155000	0,3837	0,1609	0,7759	-0,3461	-0,2904	0,0192	0,1428	
2	65100000	0,2789	-0,8317	-0,0404	0,4897	0,0237	-0,4158	0,0996	
VE)	65060000	0,3679	0,0328	0,2982	0,5307	0,5319	0,3262	-0,3220	
<u></u>	65035000	0,3985	0,2744	-0,4325	-0,0110	-0,3190	0,6593	0,2056	
AUTO	65027000	0,2943	-0,1527	-0,0862	0,3581	-0,6272	-0,3074	-0,5163	
7	65010000	0,4199	0,0451	-0,1264	-0,2859	0,1492	0,0458	0,7108	

TABELA 3.5 – VARIÂNCIA EXPLICADA DAS ESTAÇÕES, EM ORDEM DECRESCENTE

Estação	Variância explicada	Acumulada		
65220000	79,42%	79,42%		
65100000	10,01%	89,43%		
65155000	5,52%	94,95%		
65060000	2,30%	97,25%		
65027000	1,49%	98,74%		
65010000	0,63%	99,37%		
65035000	0,63%	100,00%		

Na Figura 3.3, foram representadas as estações fluviométricas Rio Negro, São Bento, Fluviópolis e União da Vitória, indicadas por seus IDs, que correspondem a, respectivamente: **ID** 1, **ID** 2, **ID** 3 e **ID** 4.

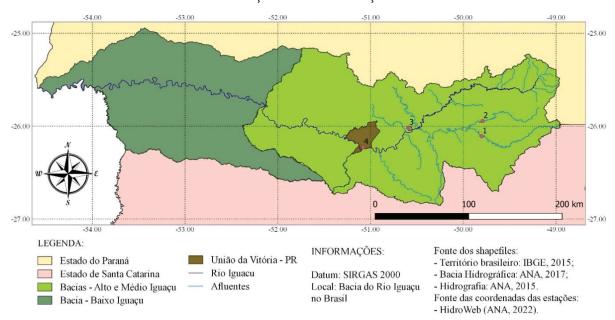


FIGURA 3.3 – LOCALIZAÇÃO DAS ESTAÇÕES FLUVIOMÉTRICAS

Considerando as estações selecionadas a partir da Análise dos Componentes Principais e suas maiores correlações (Tabela 3.3) com a estação Fluviométrica União da Vitória (65310000), foram definidas as defasagens entre cada posto e a área de estudo. Fluviópolis (65220000) atingiu a maior correlação com União da Vitória no *lag 0*, diminuindo sua correlação gradativamente conforme se aumenta a defasagem; Fluviópolis (**ID 3**) é a estação mais próxima de União da Vitória (Figura 3.3) e está localizada no rio Iguaçu, além de apresentar a série de dados mais curta dentre as estações selecionadas, de 1964 a 2015. No entanto, visto que, seguindo o curso do rio, há uma distância de aproximadamente 100 km entre as estações Fluviópolis e União da Vitória, entende-se que a defasagem entre os postos é maior que zero e menor que 1 dia, tornando-se imperceptível na escala diária. Desta forma, o *lag* considerado para Fluviópolis nas demais análises foi igual a 1. Rio Negro (65100000) apresentou maior correlação com União da Vitória para o *lag 2*, enquanto São Bento (65155000), para o *lag 1*. Estas foram as defasagens consideradas para a realização do agrupamento de eventos.

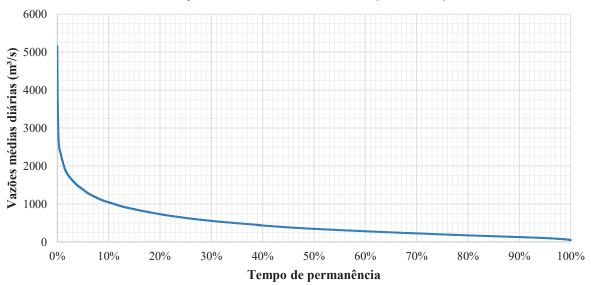
3.1.2 Seleção de eventos de cheia

Para realização do agrupamento, fez-se necessária a delimitação e seleção dos eventos de cheia a serem utilizados na pesquisa. Os eventos de cheia foram selecionados a partir da série de vazões médias diárias da Estação Fluviométrica União da Vitória (65310000), para o período de 1964 a 2015, simultâneo a todas as estações fluviométricas consideradas no estudo. A delimitação das cheias se deu com base na estatística descritiva (Tabela 3.6) e na Curva de Permanência (Figura 3.4) das vazões médias diárias em União da Vitória.

TABELA 3.6 – ESTATÍSTICA DESCRITIVA DAS VAZÕES MÉDIAS DIÁRIAS NA ESTAÇÃO FLUVIOMÉTRICA UNIÃO DA VITÓRIA (65310000) – 1964 A 2015

n (anos)	Média (m³/s)	Mínima (m³/s)	Máxima (m³/s)	Desvio-padrão (m³/s)	Assimetria
52	544,9	46,2	5156,7	473,1	2,41

FIGURA 3.4 – CURVA DE PERMANÊNCIA DA SÉRIE HISTÓRICA DE VAZÕES DA ESTAÇÃO UNIÃO DA VITÓRIA (65310000)



A partir da Equação 3.2, foi estabelecido um limite inferior de vazão (L), a partir do qual se considera um evento válido:

$$L = M.\sigma \tag{3.2}$$

Onde:

L é o limite inferior de vazão (m^3/s);

 σ é o desvio-padrão da amostra (m³/s), definido na Tabela 3.6; M é o parâmetro multiplicador, sujeito a Equação 3.3:

$$\frac{\mu}{\sigma} \le M \le \gamma \tag{3.3}$$

Onde:

 μ é a média de vazões da amostra (m³/s); γ é a assimetria da amostra.

Desta forma, foram testados cinco valores de *M*, equidistantes entre si, variando de 1,15 a 2,41. A partir destas informações, os limites inferiores de vazão aplicados, juntamente com seus respectivos tempos de permanência, obtidos a partir da Figura 3.4, são apresentados na Tabela 3.7. Observa-se da Tabela 3.7 que os limites de vazão variaram de um valor próximo a média (544 m³/s) a 1140 m³/s, enquanto os tempos de permanência foram de 35% para vazão média até 8%, para o maior limite inferior de vazão.

TABELA 3.7 – PARÂMETROS MULTIPLICADORES, LIMITES INFERIORES DE VAZÃO E TEMPOS DE PERMANÊNCIA

M	$L (m^3/s)$	Permanência
1,15	544	35
1,48	700	20
1,78	842	12
2,10	993	10
2,41	1140	8

Assim, para cada valor de *L* da Tabela 3.7, foram delimitados os eventos de cheia de forma a identificar suas respectivas datas de início, de fim e de pico, bem como suas vazões iniciais e de pico, a partir do procedimento indicado:

- 1. Todas as vazões médias diárias da série são verificadas;
- Um evento é detectado quando a vazão média diária passa a ser superior à vazão média da série;
- 3. Registra-se a data defasada em um dia àquela correspondente à vazão média;
- 4. O início do evento é definido pela data da Etapa 3, defasada ainda pelo maior *lag* das vazões de entrada do modelo em relação ao local de previsão vazões das estações fluviométricas até União da Vitória;

- 5. Identificado o início de um evento, busca-se por uma vazão superior ao limite inferior de vazão *L*;
- 6. Quando encontrado o valor de L, o evento é considerado válido;
- 7. A cada passo após o início de um evento válido, é registrada a maior vazão e sua respectiva data, sobrepondo-se ao registro anterior quando excedida a vazão;
- 8. A cada sequência de vazões de um evento válido, ao detectar uma vazão inferior à vazão média da série, registra-se a data de fim do evento.

Cada valor de *L* retornou um número de eventos, para os quais foram obtidas as características: a média de intervalos entre dois eventos consecutivos e o número de sobreposições de eventos. A Tabela 3.8 apresenta as características identificadas para os eventos de cada limite inferior de vazão:

TABELA 3.8 – CARACTERÍSTICAS DOS EVENTOS PARA OS DIFERENTES LIMITES MÍNIMOS DE VAZÃO

M	L (m ³ /s)	Número de eventos	Média de intervalos (dias)	Número de sobreposições
1,15	544	328	58	52
1,48	700	194	97	12
1,78	842	145	130	9
2,10	993	109	171	6
2,41	1140	88	212	3

A partir da Tabela 3.8, tem-se que quanto menor o valor de *L*, maior o número de sobreposições entre eventos consecutivos, tornando-se inviável a utilização de um limiar de vazão tão baixo quanto a média (544 m³/s). Da mesma forma, quanto maior o valor de *L*, menor o número de eventos identificados e, neste contexto, a utilização de valores maiores de *L* resultaria em uma diminuição representativa da extensão das séries de eventos por grupo, dificultando a modelagem.

Para definição do valor ideal do limiar inferior de vazão, fez-se necessária a compreensão de outras condições da pesquisa, como a distribuição cronológica dos eventos selecionados, que pode ser interpretada pela média de intervalos entre dois eventos consecutivos da Tabela 3.8: quanto maior o valor de *L*, maior a média. Neste caso, o processo de agrupamento de eventos tende a diminuir significativamente a quantidade de eventos disponíveis nos grupos de eventos, podendo inviabilizar a realização da estratégia de

amostragem para calibração dos grupos quando há um intervalo muito grande entre dois eventos consecutivos.

Ainda, é necessário considerar a característica de escoamento da área de estudo, com base no limiar de extravasamento do canal principal do rio Iguaçu em União da Vitória, igual a 1.387 m³/s (Steffen e Gomes, 2018). Neste contexto, foram estabelecidos intervalos de vazões e identificadas as quantidades de eventos pertencentes a cada intervalo, conforme apresenta a Tabela 3.9:

TABELA 3.9 – FREQUÊNCIA ABSOLUTA DOS PICOS DE EVENTOS PARA DIFERENTES LIMITES MÍNIMOS DE VAZÕES

	Intervalo	ões (m³/s)	Total de	
L (m ³ /s)	< 1.000	1.000 - 1.387	> 1.387	eventos
544	223	48	57	328
700	89	48	57	194
842	40	48	57	145
993	04	48	57	109
1140	00	31	57	88

Na Tabela 3.9, como era esperado, observa-se uma diminuição no número de eventos com picos inferiores a $1.000 \text{ m}^3/\text{s}$ com o aumento do limite inferior de vazão (L), tornando inviável a utilização dos valores em destaque.

Visto que a vazão média de União da Vitória apresenta uma permanência em torno de 35%, enquanto o limiar de extravasamento do canal principal do rio Iguaçu em União da Vitória (1.387 m³/s) apresenta uma permanência em torno de 5%, optou-se pela utilização de todas as cheias cujos picos atingiram valores superiores à 700 m³/s; este limiar corresponde a um tempo de permanência de 20% (Tabela 3.7), enquanto o limite de 842 m³/s equivale a 12%, o que resultaria em uma menor amplitude de vazões, dificultando a identificação de grupos com características nitidamente distintas.

Portanto, a utilização exclusiva de cheias com picos equivalentes a um tempo de permanência inferior a 20% (limite inferior de vazão de 700 m³/s) (i) reduziria significativamente a quantidade de eventos disponíveis, que, quando agrupados, forneceriam classes com número de cheias inviável para a posterior modelagem, além de (ii) espaçar a distribuição temporal dos eventos, o que não é conveniente para a estratégia de validação cruzada aplicada.

Desta forma, a série de médias diárias retornou um total de 194 eventos, com picos superiores a 700 m³/s em União da Vitória. A sobreposição dos eventos consecutivos foi tratada,

de forma que a ocorrência de um segundo evento iniciasse no dia imediatamente posterior ao final do primeiro.

A Tabela 3.10 indica a frequência de distribuição dos picos dos eventos utilizados nesta pesquisa: quase 30% dos eventos apresentaram extravasamento em 52 anos de observação, enquanto 25% atingiram vazão de, pelo menos 1.000 m³/s, que equivale a duas vezes a vazão média em União da Vitória.

TABELA 3.10 – FREQUÊNCIA DOS PICOS DE EVENTOS PARA DIFERENTES INTERVALOS DE VAZÕES

	Intervalos	Intervalos dos picos de vazões (m³/s)						
	700 – 1.000	1.000 - 1.387	> 1.387	eventos				
Frequência absoluta	89	48	57	194				
Frequência relativa	45,9%	24,7%	29,4%	100,0%				

A amostra de eventos obtida foi disposta em ordem cronológica na Figura 3.5, considerando as datas de pico ocorridas em União da Vitória; o maior espaço de tempo entre dois picos consecutivos foi de pouco menos de 2 anos, entre abril de 1985 e dezembro de 1986, enquanto a média de intervalos foi de 97 dias.

FIGURA 3.5 – DISTRIBUIÇÃO TEMPORAL DOS EVENTOS SELECIONADOS



As séries dos 194 eventos delimitados foram extraídas da série completa, para as quatro estações fluviométricas (União da Vitória, Fluviópolis, São Bento e Rio Negro).

3.1.3 Indicadores de cheia

Os indicadores de cheia correspondem às variáveis características dos eventos que compõem uma amostra para agrupamento de dados. Para cada um dos eventos selecionados em União da Vitória e nas estações fluviométricas a montante, foram estabelecidos seis indicadores e seus respectivos valores, calculados:

1. Vazão de pico em União da Vitória (Q_p);

- 2. Tempo de pico em União da Vitória (t_p);
- 3. Vazão inicial em União da Vitória (Qt0);
- 4. Vazão em Fluviópolis defasada em um dia (*lag 1*) do pico em União da Vitória ((**Q**_{tp-1})_{FL});
- 5. Vazão em São Bento defasada em um dia (*lag 1*) do pico em União da Vitória ((**Q**_{tp-1})s_B);
- Vazão em Rio Negro defasada em dois dias (*lag 2*) do pico em União da Vitória ((Q_{tp-2})_{RN}).

Devido às diferenças físicas e de ordem de grandeza, os indicadores de cheia precisaram passar por processo de tratamento, que os reduz a uma escala comum e adimensional. Foi utilizada a padronização linear, definida pela Equação 3.1. Os indicadores normalizados são os dados de entrada para o agrupamento realizado.

3.2 MÉTODO DE AGRUPAMENTO

Utilizou-se o algoritmo *fuzzy c-means* para o agrupamento dos dados. Para uma amostra $\mathbf{X} = \{x_1, ..., x_n\}$, disposta em um espaço com d dimensões, em que d é representado pelo número de indicadores de cheia, e um conjunto de centroides $\mathbf{V} = \{v_1, ..., v_c\}$ (Ren et al., 2016; Yang e Nataliani, 2017), os graus de similaridade dos elementos de \mathbf{X} com os grupos de \mathbf{V} são definidos a partir da Equação (3.4) (Wang et al., 2015; Ren et al., 2016; Yang e Nataliani, 2017; Steffen e Gomes, 2018), sendo que o conjunto dos graus de similaridade de um determinado agrupamento difuso define a matriz de partição difusa (\mathbf{U}).

$$\mu_{ki} = \left[\sum_{t=1}^{c} \left(\frac{d_{ik}}{d_{tk}}\right)^{\frac{2}{m-1}}\right]^{-1} \tag{3.4}$$

$$\sum_{k=1}^{c} \mu_{ki} = 1 \,\forall i \tag{3.5}$$

$$0 \le \mu_{ki} \le 1,0 \tag{3.6}$$

Onde:

 μ_{ki} é o grau de similaridade da *i*-ésima observação com o *k*-ésimo grupo, ou elemento da matriz de partição difusa (U);

 d_{ik} é a distância Euclidiana da *i*-ésima observação ao *k*-ésimo grupo, ou elemento da matriz de distâncias Euclidianas (**D**);

 d_{tk} é a distância Euclidiana da t-ésima observação ao k-ésimo grupo, ou elemento da matriz de distâncias Euclidianas (**D**);

c é o número total de grupos;

m é o parâmetro de incerteza;

i é a i-ésima observação da amostra (i = 1, 2, ..., n);

n é o número total de observações da amostra.

Portanto, a determinação da matriz de partição difusa (U) é função das distâncias Euclidianas (d_{ik}), obtidas pela Equação (3.7):

$$d_{ik} = \|x_i - v_k\|^2 = \sqrt{\sum_{j=1}^d (x_{ij} - v_{kj})^2}$$
(3.7)

Onde:

 x_{ij} é o valor tratado da *i*-ésima observação do *j*-ésimo indicador;

 v_{kj} é a coordenada do k-ésimo centroide, correspondente ao j-ésimo indicador;

d são as dimensões de X, ou número total de indicadores de cheia para cada classe, ou número de características de X.

Uma vez que as combinações que os graus de aderência podem assumir para cada amostra são infinitas (Hall e Minns, 1999), utiliza-se da minimização da função objetivo para determinar a matriz de partição difusa final; a função objetivo (*J*) atua como um critério de parada do algoritmo e é representada pela Equação (3.8) (Wang *et al.*, 2015; Ren *et al.*, 2016; Yang e Nataliani, 2017; Steffen e Gomes, 2018):

$$J = \sum_{k=1}^{c} \sum_{i=1}^{n} (\mu_{ki})^{m} (d_{ik})^{2}$$
(3.8)

A função objetivo atinge seu valor mínimo quando a diferença, em módulo, entre os máximos graus de similaridade das observações dos dois últimos passos do processo iterativo é menor que o erro máximo tolerado, ou seja (Wang *et al.*, 2015; Ren *et al.*, 2016; Yang e Nataliani, 2017; Steffen e Gomes, 2018):

$$\max_{ik}\{|(\mu_{ki})^{p+1} - (\mu_{ki})^p|\} < \varepsilon_t \tag{3.9}$$

Onde:

 ε_t é o erro máximo tolerado, igual a 10^{-5} ;

 $(\mu_{ki})^{p+1}$ é o grau de similaridade da *i*-ésima observação ao *k*-ésimo grupo no último passo (p+1);

 $(\mu_{ki})^p$ é o grau de similaridade da *i*-ésima observação ao *k*-ésimo grupo no penúltimo passo (p).

As coordenadas dos centroides dos grupos são determinadas pela Equação (3.10) (Wang *et al.*, 2015; Ren *et al.*, 2016; Yang e Nataliani, 2017; Steffen e Gomes, 2018):

$$v_{kj} = \frac{\sum_{k=1}^{c} (\mu_{ki})^m x_{ij}}{\sum_{k=1}^{c} (\mu_{ki})^m}$$
(3.10)

Portanto, o algoritmo *fuzzy c-means* pode ser definido pelo processo iterativo a seguir (Bezdek *et al.*, 1984; Steffen e Gomes, 2018):

- 1. Definição, a priori, de *c* e *m*;
- 2. Definição de uma matriz de partição difusa inicial, aleatória (U);
- 3. Cálculo dos centroides dos grupos (V) (Equação 3.10);
- 4. Cálculo das distâncias Euclidianas (Equação 3.7);
- 5. Cálculo da função objetivo (*J*) (Equação 3.8);
- 6. Atualização da matriz de partição difusa (U) (Equação 3.4);
- 7. Verificação do critério de parada do algoritmo (Equação 3.9):
- 8. Retorno à etapa 3, quando não atendido o critério de parada.

Este processo foi aplicado ao indicadores definidos na Seção 3.1.3, através de implementação em Python e aplicação da função *fuzz.cluster.cmeans* da biblioteca *skfuzzy*. No

código, forçou-se que os centroides fossem ordenados a partir do indicador de vazão máxima em União da Vitória, do menor para o maior. Portanto, os grupos foram numerados de 1 a *c*, nos quais as vazões em 1 são as mais baixas, enquanto em *c*, as mais altas.

A definição do número de grupos (c) e do grau de incerteza (m) foi realizada empiricamente. Para cada combinação de $c \in \mathbb{Z}$ no intervalo de 2 a 10, e m, de 1,25 a 2,00 incrementado em 0,25, o algoritmo foi executado e suas saídas, comparadas. Escolheu-se a combinação que apresentou comportamento mais homogêneo da matriz de partição difusa final para as simulações realizadas. Adotou-se m = 1,25.

Para os resultados obtidos para m igual a 1,25, observou-se que quando c era maior ou igual a 4, o grupo das maiores vazões continha poucos eventos, inviabilizando a modelagem do grupo isoladamente; por exemplo, quando c=4, os resultados apresentaram apenas três eventos no grupo das vazões extremas máximas. Desta forma, adotou-se c=3, na tentativa de promover a maior distinção entre os eventos de cada grupo.

A aleatoriedade da matriz de partição difusa inicial, eventualmente, é capaz de alterar os resultados obtidos no agrupamento. Por isso, efetuou-se uma análise de sensibilidade da matriz inicial aleatória, na qual o código foi executado, repetidamente, por 100 vezes e foram extraídas as informações referentes a cada evento analisado. O critério de avaliação se deu a partir das médias das distâncias dos eventos pertencentes a cada grupo até seu respectivo centroide. Por exemplo, para c igual a 3, considerou-se a distância de cada evento que compõe o grupo 1 até o centroide deste grupo e calculou-se a média entre estas distâncias; sucessivamente, o processo se repetiu para os grupos 2 e 3. O valor de referência para avaliação é a média extraída das três médias, obtidas para os grupos 1 a 3.

Os 100 valores obtidos foram avaliados e observou-se uma amplitude de 5.10^{-7} , sendo, portanto, observada a maior diferença entre os valores apenas na sétima casa decimal; o desvio-padrão foi equivalente a 5.10^{-8} , indicando que a aleatoriedade da matriz de partição difusa inicial não exerce grande influência sobre o agrupamento. Portanto, o resultado da última execução foi armazenado para a análise de agrupamento.

3.3 MODELO HIDROLÓGICO PARA PREVISÃO DE VAZÕES

Esta Seção apresenta os métodos aplicados à obtenção dos modelos hidrológicos para previsão de vazões, posteriormente utilizados para obtenção das Configurações de previsão.

3.3.1 Dados de entrada dos modelos

Para a previsão de vazões em União da Vitória com um dia de horizonte de previsão $(\mathbf{Q}_{UV(t+1)})$, os dados de entrada utilizados para estimar os modelos foram as vazões médias diárias das estações: União da Vitória $(\mathbf{Q}_{UV(t)})$, Fluviópolis $(\mathbf{Q}_{FL(t)})$ e São Bento $(\mathbf{Q}_{SB(t)})$, no instante inicial, e Rio Negro, defasada de um dia $(\mathbf{Q}_{RN(t-1)})$. As defasagens utilizadas nas vazões de entrada dos modelos se diferem daquelas utilizadas no agrupamento, visto que o objetivo é realizar a previsão com 1 dia de horizonte.

Cada série de vazões, do período de 1964 a 2015, foi normalizada pelo método de padronização linear, conforme Equação (3.1). O período de 1964 a 2004 foi utilizado para ajuste dos modelos, enquanto o período de 2005 a 2015 foi aplicado para a fase de previsão.

Os modelos foram estimados a partir de conjuntos distintos de dados de vazão: série completa; todos os eventos de cheia – definidos na Seção 3.1.2; e três grupos de eventos de cheia – conforme número de grupos identificado na Seção 3.2. Desta forma, a quantidade de modelos estabelecidos corresponde a cinco.

3.3.2 Critérios de qualidade de ajuste

A avaliação dos ajustes e das previsões foi realizada a partir dos índices de performance, assinaturas hidrológicas baseadas nas Curvas de Permanência, parâmetros estatísticos e medidas de erros relativos.

3.3.2.1 Índices de performance

Os índices de performance considerados foram a Eficiência de Nash-Sutcliffe (NS) (Lawin *et al.*, 2019; Liang et al., 2019; Wijayarathne e Coulibaly, 2020; Althoff et al., 2021; Brêda et al., 2021; Mosavi et al., 2021; Ebtehaj e Bonakdari, 2022; Moura et al., 2022; Singh e Saravanan, 2022; Difi *et al.*, 2023; Zakhrouf *et al.*, 2023), a Eficiência de Kling-Gupta (KG) (Gupta et al., 2009; Wijayarathne e Coulibaly, 2020; Althoff et al., 2021; Moura et al., 2022; Singh e Saravanan, 2022; Lappicy e Lima, 2023), o Erro Relativo Absoluto Médio (MARE) (Staudinger et al., 2011; Liang et al., 2019; Althoff et al., 2021; Ebtehaj e Bonakdari, 2022; Moura et al., 2022), e a Eficiência de Nash-Sutcliffe Modificada (Tucci, 2005). As métricas foram selecionadas com a finalidade de avaliar diferentes fases do hidrograma: o NS é uma medida de associação estatística, que indica uma porcentagem da variância observada,

explicada pelos dados previstos, e que enfatiza eventos extremos em detrimento das vazões médias (Unduche et al., 2018; Moura et al., 2022), o que significa que a métrica é mais sensível aos picos dos hidrogramas, conduzindo a um aumento do coeficiente (Althoff et al., 2021); KG é capaz de melhor diagnosticar a importância relativa da média, variabilidade e dinâmica de vazões (Unduche et al., 2018; Althoff et al., 2021; Moura et al., 2022); MARE calcula o erro como uma porcentagem da vazão observada e foca nas vazões baixas a médias (MARE) (Unduche et al., 2018; Moura et al., 2022), dando a mesma ênfase aos erros relativos independentemente das condições das vazões baixas ou superiores (Althoff et al., 2021). A Eficiência de Nash-Sutcliffe Modificada considera a vazão do dia da previsão, ao invés da média — utilizada no cálculo da Eficência de Nash-Sutcliffe tradicional (Tucci, 2005).

Nash-Sutcliffe (NS):

O coeficiente de Nash-Sutcliffe (NS) é uma estatística normalizada, que se relaciona com o Erro Quadrado Médio (MSE); NS determina a relação entre o erro quadrado médio e a variância dos dados observados e varia de [−∞; 1] − quanto mais próximo de 1, melhor. A Equação (3.11) define a relação entre NS e MSE (Gupta *et al.*, 2009):

$$NS = 1 - \frac{MSE}{\sigma_o^2} \tag{3.11}$$

Onde:

NS é o coeficiente de Nash-Sutcliffe;

 σ_0^2 é a variância das vazões observadas;

MSE é o Erro Quadrado Médio, definido pela Equação (3.12):

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (Q_{s_i} - Q_{o_i})^2$$
(3.12)

Onde:

n é o tamanho da amostra;

 Q_{o_i} são as vazões observadas;

 Q_{s_i} são as vazões simuladas.

O coeficiente de NS pode ser obtido a partir da Equação (3.13) (Gupta *et al.*, 2009; Lawin *et al.*, 2019; Althoff et al., 2021):

$$NS = 1 - \frac{\sum_{i=1}^{n} (Q_{s_i} - Q_{o_i})^2}{\sum_{i=1}^{n} (Q_{o_i} - \overline{Q_o})^2}$$
(3.13)

Onde:

 $\overline{Q_o}$ é a vazão média observada.

Os valores da eficiência de Nash-Sutcliffe e do Erro Quadrado Médio são os dois critérios bastante difundidos para calibração e avaliação de modelos hidrológicos com dados observados (Gupta *et al.*, 2009).

Nash-Sutcliffe Modificado (NSM):

O coeficiente de Nash-Sutcliffe Modificado considera em seu cálculo a vazão do dia da realização da previsão, ao contrário do tradicional NS, que faz uso da média da série observada; segundo Tucci (2005), a predominância da previsão pela média não é adequada. O coeficiente de NSM pode ser obtido a partir da Equação (3.14) (Tucci, 2005):

$$NSM = 1 - \frac{\sum_{i=1}^{n} (Q_{s_i} - Q_{o_i})^2}{\sum_{i=1}^{n} (Q_{o_i} - Q_{o_{i-1}})^2}$$
(3.14)

Onde:

 $Q_{o_{i-1}}$ é a vazão do dia da previsão.

Kling-Gupta (KG):

O coeficiente de Kling-Gupta consiste em uma decomposição do coeficiente de Nash-Sutcliffe, que facilita a análise da importância relativa dos diferentes componentes da modelagem hidrológica (Lawin *et al.*, 2019) por meio do coeficiente de correlação (r), da variabilidade da modelagem (α), definida pela razão entre os desvios-padrão simulado e observado, e tendência (β), definida pela razão entre as médias simulada e observada (Gupta *et al.*, 2009). A Equação (3.15) define o cálculo do coeficiente KG (Gupta *et al.*, 2009):

$$KG = 1 - \sqrt{(r_{KG} - 1)^2 + (\alpha_{KG} - 1)^2 + (\beta_{KG} - 1)^2}$$
(3.15)

Onde:

KG é o coeficiente de Kling-Gupta;

 r_{KG} é o coeficiente de correlação;

 α_{KG} é a variabilidade da modelagem; definida pela Equação (3.16);

 β_{KG} é a tendência, definida pela Equação (3.17).

$$\alpha_{KG} = \frac{\sigma_s}{\sigma_o} \tag{3.16}$$

$$\beta_{KG} = \frac{\mu_s}{\mu_o} \tag{3.17}$$

Onde:

 σ_s é o desvio-padrão simulado;

 σ_o é o desvio-padrão observado;

 μ_s é a média simulada;

 μ_o é a média observada.

Erro Relativo Absoluto Médio (MARE):

MARE mede a diferença média relativa entre as vazões previstas e observadas, e é calculado como a média das diferenças absolutas, divididas pelos valores observados, como indica a Equação (3.18) (Letessier et al., 2023):

$$MARE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{Q_{o_i} - Q_{s_i}}{Q_{o_i}} \right|$$
 (3.18)

MARE fornece a medida do erro percentual médio do modelo; sua vantagem se torna mais evidente quando a magnitude do erro relativo à vazão observada é relevante; baixos valores de MARE indicam um modelo mais acurado, representando um médio desvio percentual baixo, se comparados os valores simulados aos observados (Althoff et al., 2021; Letessier et al., 2023).

3.3.2.2 Assinaturas hidrológicas

Considerando a série completa, a série com todos os eventos e as séries de eventos por grupos, foram construídas as Curvas de Permanência observadas e previstas para cada etapa da modelagem. Para cada condição, foram extraídas as vazões de percentis 70, 35, 20 e 5%.

A delimitação dos percentis de 35 a 5% se deu com base nos critérios de seleção dos eventos de cheia (Seção 3.1.2), em que: 35% equivalem a vazão média em União da Vitória; 20% representam o limiar de corte dos eventos; e 5% correspondem à vazão de extravasamento.

Já a vazão de 70% foi considerada de acordo com Yilmaz *et al.* (2008) e Fatehifar *et al.* (2021), delimitando a transição de uma região de vazões intermediárias no hidrograma para uma região de vazões mais baixas.

3.3.2.3 Medidas de erros relativos

As medidas de erros relativos foram calculadas para os parâmetros estatísticos – máxima, média, desvio-padrão, assimetria e volume escoado – das séries observadas e previstas, bem como, para as vazões com diferentes tempos de permanência, de forma que traduzissem um percentual de erro gerado pela previsão, em relação às séries observadas. A Equação (3.19) define o cálculo das medidas de erros relativos (Fatehifar *et al.*, 2021):

$$\varepsilon_r = \frac{P_e - P_o}{P_o}.100\tag{3.19}$$

Onde:

 ε_r é o erro relativo, expresso em percentual;

 P_e é o parâmetro estimado;

 P_o é o parâmetro observado.

3.3.3 Algoritmo para previsão de vazões

Utilizou-se o algoritmo *Support Vector Machine*, em seu formato de regressão nãolinear (SVR), como método de previsão de vazões. A partir da função de custo ε -indiferente de Vapnik (Equação (2.16)), Schölkopf *et al.* (2000) apresentaram uma modificação do algoritmo, denominada ν -SVR, que automaticamente minimiza ε , tornando o método independente do ajuste deste parâmetro por intermédio de algoritmos de otimização.

No algoritmo ν -SVR, ν é o parâmetro responsável pelo controle do número de vetores de suporte – ou a quantidade de erros da solução (Schölkopf *et al.*, 2000); considerando que $\nu \in (0,1]$, quanto mais próximo de 1, maior é o número de vetores de suporte e, consequentemente, maior é a tolerância aos erros (Chang e Lin, 2022).

O parâmetro C, correspondente à constante de regularização, deve ser definido a priori. A partir do conjunto de parâmetros, o algoritmo ν -SVR objetiva minimizar a Equação (3.20) (Schölkopf *et al.*, 2000; Chang e Lin, 2001; Chang e Lin, 2022):

$$\tau(\mathbf{w}, \xi^*, \varepsilon) = \frac{1}{2} \|\mathbf{w}\|^2 + C \left(\nu \varepsilon + \frac{1}{m} \sum_{i=1}^m (\xi_i + \xi_i^*) \right)$$
(3.20)

Sujeita a:

$$((\mathbf{w} \cdot x_i) + b) - y_i \le \varepsilon + \xi_i \tag{3.21}$$

$$y_i - ((\mathbf{w} \cdot x_i) + b) \le \varepsilon + \xi_i^*, \tag{3.22}$$

$$\xi_i, \xi_i^* \geq 0$$

 $\varepsilon > 0$

$$i = 1, 2, ..., m$$

Onde:

 ξ_i é a variável de folga que especifica o maior erro de treinamento, sujeito à tolerância; ξ_i^* é a variável de folga que especifica o menor erro de treinamento, sujeito à tolerância; ε é a tolerância;

C é a constante que define o grau de penalização de custo quando há erro de treinamento, ou constante de regularização.

Para as restrições, introduz-se multiplicadores de Lagrange e, posteriormente, reescrevendo-as, o problema de otimização a ser solucionado é dado pela maximização da Equação (3.23), para $\nu \ge 0$, C > 0:

$$W(\alpha^*) = \sum_{i=1}^{m} (\alpha_i^* + \alpha_i) y_i - \frac{1}{2} \sum_{i,j=1}^{m} (\alpha_i^* + \alpha_i) (\alpha_j^* + \alpha_j) k(\mathbf{x}_i, \mathbf{x}_j)$$
(3.23)

Sujeita a:

$$\sum_{i=1}^{m} (\alpha_i - \alpha_i^*) = 0 \tag{3.24}$$

$$\alpha_i^{(*)} \in \left[0, \frac{C}{m}\right]$$

$$\sum_{i=1}^{m} (\alpha_i + \alpha_i^*) \le C.v$$

A equação de regressão adquire o formato da Equação (3.25) (Schölkopf *et al.*, 2000; Chang e Lin, 2001; Chang e Lin, 2022):

$$f(x) = \sum_{i=1}^{m} (\alpha_i^* - \alpha_i) k(\mathbf{x}_i, \mathbf{x}) + b$$
(3.25)

Desta forma, com o objetivo de obter a equação de regressão capaz de prever as vazões em União da Vitória, utilizou-se a classe NuSVR do módulo sklearn.svm, cuja implementação se dá a partir da biblioteca LIBSVM (Chang e Lin, 2022), em Python. Os parâmetros a serem calibrados neste trabalho são: a função kernel; o coeficiente de kernel, denominado Gamma (γ); um limite superior para a fração de erros de treinamento e um limite inferior para a fração de vetores de suporte, cujo parâmetro se denomina nu, com valor definido no intervalo (0, 1]; e o parâmetro de regularização C.

Como função *kernel*, *s*elecionou-se a *Radial Basis Function* (RBF), que permite mapear amostras não lineares em um espaço dimensional superior e apresenta menos parâmetros a serem calibrados – quando comparada à função polinomial –, além de apresentar menor dificuldade numérica (Bhavsar e Panchal, 2012; Freitas, 2016; Hsu *et al.*, 2016).

Considerando as sequências de valores definidos nas Equações 3.26 e 3.27, os parâmetros C e γ , na etapa de calibração, foram combinados um a um, para identificação do melhor conjunto dentre as 64 combinações de C e γ ; inicialmente, testou-se também parâmetros inferiores e superiores aos intervalos, cujos resultados foram insatisfatórios.

$$C = [0,01; 0,05; 0,1; 0,5; 1,0; 2,0; 5,0; 10]$$
(3.26)

$$\gamma = [0,001;0,01;0,05;0,1;0,33;0,5;1,0;5]$$
 (3.27)

Aos demais parâmetros da classe NuSVR foram atribuídas suas respectivas especificações padronizadas, inclusive para o parâmetro v (nu), cujo valor padrão é de 0,5. Ou seja, foram calibrados os parâmetros C e γ do algoritmo.

3.3.4 Calibração dos parâmetros do modelo

O período de 1964 a 2004 foi subdividido em dois períodos distintos, denominados de períodos de treinamento e teste. Utilizou-se uma estratégia de amostragem dividida (Hallouin *et al.*, 2020), com base em uma janela de 20 anos. A janela de tempo partiu de uma divisão aproximadamente igual para as fases de treinamento e teste, de forma que os maiores eventos de cheia estivessem sempre presentes em uma das etapas; oito amostras foram geradas, sendo que em quatro delas o período de treinamento abrangeu o evento de 1983 (maior cheia histórica) e as outras quatro, não. Isto pode ser observado na Figura 3.6, que representa a estratégia de validação cruzada aplicada.

Treinamento Teste 1 1964 2 1964 3 1964 4 1964 5 1964 6 1964 7 1964 8 1964

FIGURA 3.6 – ESTRATÉGIA DE AMOSTRAGEM DIVIDIDA

As Tabelas 3.11 e 3.12 apresentam as estatísticas dos dados observados para a série completa, considerando as diferentes amostras de treinamento e teste. Como era esperado, em virtude dos períodos de treinamento e teste considerados (Figura 3.6), são observadas semelhanças entre os parâmetros estatísticos para os pares de amostra de treinamento e teste, respectivamente: 1 e 5, 2 e 6, 3 e 7, 4 e 8, 5 e 1, 6 e 2, 7 e 3, 8 e 4.

TABELA 3.11 – ESTATÍSTICAS DA SÉRIE COMPLETA OBSERVADA E SUBDIVIDIDA CONFORME ESTRATÉGIA DE AMOSTRAGEM – PERÍODOS DE TREINAMENTO

Amostra	1	2	3	4	5	6	7	8
Tamanho da amostra	7662	7496	7682	7485	7187	7673	7586	7529
Mínima (m³/s)	61,9	48,6	48,6	48,6	48,6	63,9	63,9	61,9
Máxima (m³/s)	5156,7	5156,7	5156,7	5156,7	3953,6	3953,6	2751,7	2428,5
Média (m³/s)	516,1	522,3	523,1	608,7	560,1	555,8	548,6	474,3
Desvio-padrão (m³/s)	459,3	473,8	496,4	551,6	478,4	462,3	430,9	355,2
Assimetria	3,2	3,0	3,2	2,5	2,1	2,1	1,6	1,7

TABELA 3.12 – ESTATÍSTICAS DA SÉRIE COMPLETA OBSERVADA E SUBDIVIDIDA CONFORME ESTRATÉGIA DE AMOSTRAGEM – PERÍODOS DE TESTE

Amostra	1	2	3	4	5	6	7	8
Tamanho da amostra	7312	7478	7292	7489	7787	7301	7388	7445
Mínima (m³/s)	48,6	63,9	63,9	61,9	61,9	48,6	48,6	48,6
Máxima (m³/s)	3953,6	3953,6	2751,7	2428,5	5156,7	5156,7	5156,7	5156,7
Média (m³/s)	556,9	549,8	549,6	463,4	513,8	515,2	523,1	598,4
Desvio-padrão (m³/s)	475,6	461,2	435,1	350,5	456,6	472,6	502,5	552,0
Assimetria	2,1	2,1	1,6	1,7	3,2	3,1	3,2	2,5

Da mesma forma, as Tabelas 3.13 e 3.14 apresentam as estatísticas dos dados observados para a série de todos os eventos, para as diferentes amostras de treinamento e teste, enquanto as magnitudes dos eventos que compõem cada amostra podem ser observadas nas Figuras 3.7 a 3.14, que representam os hidrogramas observados dos períodos de treinamento e teste, para cada amostra (de 1 a 8). Nota-se que, nos hidrogramas observados, os eventos foram posicionados sequencialmente, de forma que o final de um evento indica o início do próximo.

TABELA 3.13 – ESTATÍSTICAS DA SÉRIE OBSERVADA DE TODOS OS EVENTOS E SUBDIVIDIDA CONFORME ESTRATÉGIA DE AMOSTRAGEM – PERÍODOS DE TREINAMENTO

Amostra	1	2	3	4	5	6	7	8
Tamanho da amostra	2850	2772	2793	3314	2898	3142	3105	2556
Mínima (m³/s)	114	148,1	144,9	135,4	135,4	114	114	114
Máxima (m³/s)	5156,7	5156,7	5156,7	5156,7	3953,6	3953,6	2751,7	2428,5
Média (m³/s)	920,2	945,5	963,0	1030,3	979,8	952,6	929,7	844,3
Desvio-padrão (m³/s)	525,0	539,0	583,9	588,4	497,5	480,5	427,5	359,1
Assimetria	3,3	3,1	3,0	2,5	1,8	1,9	1,3	1,2

A partir das Tabelas 3.13 e 3.14, tem-se que as vazões médias, para as amostras de treinamento e teste, variaram de 840 a 1040 m³/s, sendo os maiores valores atribuídos às amostras com as maiores vazões máximas. Ademais, a partir das Figuras 3.7 a 3.17, observa-se

que, na metade das amostras, os dois maiores eventos observados estiveram separados dentre os períodos de treinamento e teste, com exceção das Amostras 3, 4, 7 e 8 (Figuras 3.9, 3.10, 3.13 e 3.14, respectivamente). Tem-se ainda que, com exceção destes extremos, as janelas de treinamento e teste de cada amostra possuem eventos de magnitudes semelhantes.

TABELA 3.14 – ESTATÍSTICAS DA SÉRIE OBSERVADA DE TODOS OS EVENTOS E SUBDIVIDIDA CONFORME ESTRATÉGIA DE AMOSTRAGEM – PERÍODOS DE TESTE

Amostra	1	2	3	4	5	6	7	8
Tamanho da amostra	2898	2976	2955	2434	2850	2606	2643	3192
Mínima (m³/s)	135,4	114	114	114	114	148,1	144,9	135,4
Máxima (m³/s)	3953,6	3953,6	2751,7	2428,5	5156,7	5156,7	5156,7	5156,7
Média (m³/s)	979,8	954,7	938,2	841,3	920,2	947,5	974,4	1035,1
Desvio-padrão (m³/s)	497,5	486,0	433,4	357,2	525,0	548,0	595,7	594,2
Assimetria	1,8	1,9	1,3	1,2	3,3	3,1	2,9	2,5

FIGURA 3.7 – HIDROGRAMAS DOS EVENTOS OBSERVADOS E SUBDIVIDIDOS CONFORME ESTRATÉGIA DE AMOSTRAGEM – AMOSTRA 1

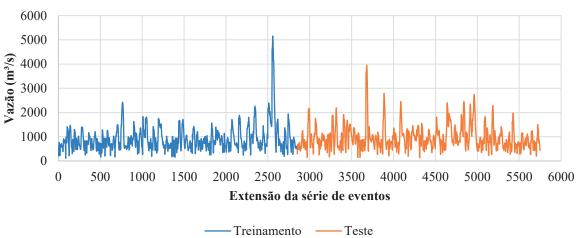
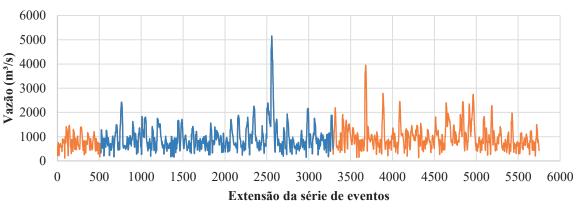


FIGURA 3.8 – HIDROGRAMAS DOS EVENTOS OBSERVADOS E SUBDIVIDIDOS CONFORME ESTRATÉGIA DE AMOSTRAGEM – AMOSTRA 2



— Teste — Treinamento

FIGURA 3.9 – HIDROGRAMAS DOS EVENTOS OBSERVADOS E SUBDIVIDIDOS CONFORME ESTRATÉGIA DE AMOSTRAGEM – AMOSTRA 3

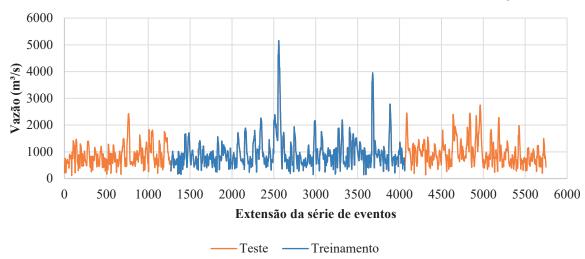


FIGURA 3.10 – HIDROGRAMAS DOS EVENTOS OBSERVADOS E SUBDIVIDIDOS CONFORME ESTRATÉGIA DE AMOSTRAGEM – AMOSTRA 4

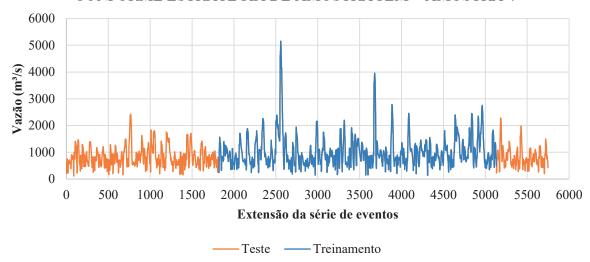
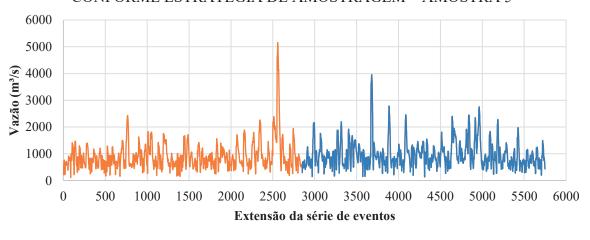


FIGURA 3.11 – HIDROGRAMAS DOS EVENTOS OBSERVADOS E SUBDIVIDIDOS CONFORME ESTRATÉGIA DE AMOSTRAGEM – AMOSTRA 5



—Teste —Treinamento

FIGURA 3.12 – HIDROGRAMAS DOS EVENTOS OBSERVADOS E SUBDIVIDIDOS CONFORME ESTRATÉGIA DE AMOSTRAGEM – AMOSTRA 6

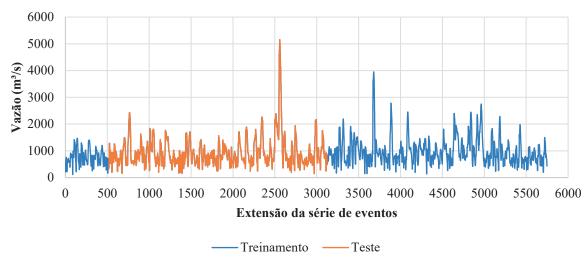


FIGURA 3.13 – HIDROGRAMAS DOS EVENTOS OBSERVADOS E SUBDIVIDIDOS CONFORME ESTRATÉGIA DE AMOSTRAGEM – AMOSTRA 7

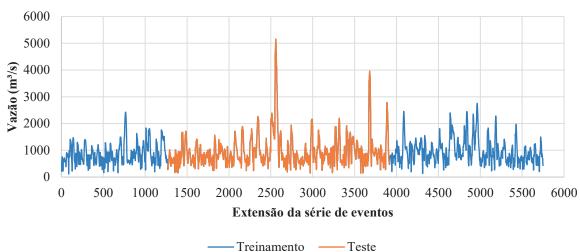
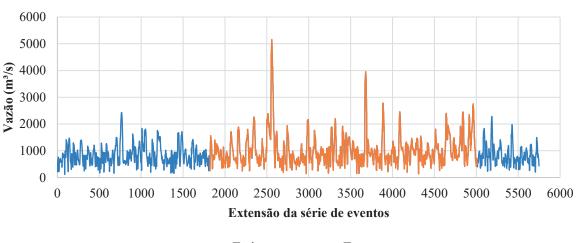


FIGURA 3.14 – HIDROGRAMAS DOS EVENTOS OBSERVADOS E SUBDIVIDIDOS CONFORME ESTRATÉGIA DE AMOSTRAGEM – AMOSTRA 8



---Treinamento ---Teste

Ainda, na etapa de calibração, considerando a estratégia de validação cruzada adotada, para cada conjunto de dados de vazão (Série Completa, Série de Todos os Eventos e Séries de Eventos por Grupo), para as oito amostras e as 64 combinações, o algoritmo *NuSVR* foi executado para o período de treinamento, gerando os modelos, que, posteriormente, foram aplicados ao período de teste correspondente; as métricas para avaliação dos resultados de treinamento e teste foram calculadas (Equações 3.11 a 3.18), bem como, suas principais estatísticas – vazões média, máxima e mínima, desvio-padrão, assimetria e relação de volumes (simulado/observado). Em resumo, os processos de treinamento e validação são apresentados na Figura 3.15.

M conjuntos de parâmetros Treinamento K-amostras da validação cruzada Simulação do período de treinamento (hidrogramas) 64 64 Cálculo da eficiência por métrica de ajuste (eficiência/métrica) 64 64 64 Seleção dos N melhores conjuntos de parâmetros para cada métrica de ajuste 32 32 32 Simulação do período de teste (hidrogramas) 32 32 32 Teste Cálculo da eficiência por métrica de ajuste (eficiência/métrica) 32 32 32 Cálculo da tendência central de previsão (mediana) Estatísticas para as 8 amostras

FIGURA 3.15 – PROCESSOS DE TREINAMENTO E VALIDAÇÃO

3.3.5 Ajuste final dos modelos de previsão

Para o ajuste final dos modelos de previsão, foram observados os índices de performance dos 32 conjuntos de parâmetros de cada amostra do período de teste, para cada categoria de série. Os 32 conjuntos de parâmetros correspondem àqueles com resultados mais próximos à mediana dos índices de performance da fase de treinamento; o número de conjuntos de parâmetros avaliados nesta etapa corresponde àquele do período de teste apresentado na Figura 3.15.

No ajuste final, o mesmo conjunto de parâmetros foi considerado para todos os modelos, sendo selecionado aquele que se repetiu para a maioria dos modelos. Por isso, no intuito de evitar o *overfitting* do modelo, o conjunto de parâmetros selecionado foi aquele que apresentou métricas próximas às medianas dos índices de performance na etapa de teste, capaz de estimar satisfatoriamente as vazões em União da Vitória, considerando as cinco categorias

de modelos. Considerando, ainda, performances próximas entre si para diferentes conjuntos de parâmetros, adotou-se aquele que apresentou valores mais baixos de C e γ , visto que, ao aumentar os valores dos parâmetros, há um aumento no tempo de processamento, mas pouca melhoria de desempenho.

A partir dos valores de *C* e γ estabelecidos – um único conjunto de parâmetros –, foi ajustado um modelo para cada categoria de série – Série Completa, de Todos os Eventos, e por Grupos de Eventos –, para todo o período da estratégia de amostragem da Figura 3.6 (de 1964 a 2004). Após o ajuste dos modelos, efetuou-se a simulação para as todas as séries, por meio de cada modelo gerado; ou seja, foram combinados os modelos ajustados e as séries de dados e, assim, cada categoria de modelo calibrado foi aplicado a cada categoria de série.

Para avaliação dos resultados obtidos na combinação de modelos e séries, foram utilizados os critérios de qualidade de ajuste, definidos na Seção 3.3.2. Para aplicação das métricas, utilizou-se a extensão total de cada série observada e simulada, sem considerar exclusivamente o período simultâneo de dados entre cada categoria de série. Por isso, apenas puderam ser comparadas entre si as simulações de uma mesma série gerada por diferentes modelos, não sendo adequada a comparação das métricas para as diferentes séries simuladas por um mesmo modelo. Por exemplo: a série completa foi simulada pelas cinco categorias de modelos (Série Completa – seu próprio modelo –, Todos os Eventos, Grupo 1, Grupo 2 e Grupo 3), da mesma forma que as demais séries (Todos os Eventos, Grupo 1, Grupo 2 e Grupo 3) foram simuladas pelo modelo de Série Completa. Contudo, visto que cada série possui uma extensão diferente, a comparação entre estas simulações – todas realizadas pelo mesmo modelo – não é adequada, enquanto comparar cinco simulações – por modelos diferentes – da Série Completa tem maior relevância.

3.4 CONFIGURAÇÕES DE MODELAGEM PARA PREVISÃO DE VAZÕES

Partindo do agrupamento, são apresentados os processos metodológicos aplicados ao desenvolvimento da modelagem hidrológica para previsão de vazões a partir da integração da lógica difusa em quatro configurações: (i) por média (ii) por grupo de pertencimento; (iii) ponderada pelas distâncias Euclidianas; e (iv) ponderada pelos graus de pertencimento. A série de todos os eventos foi prevista para o período de 2005 a 2015, com horizonte de previsão de um dia.

3.4.1 Configuração 1 – Previsão Média

Nesta configuração, foram aplicados os modelos por grupo, para previsão da série com todos os eventos. Ao ser identificado um evento, cada modelo prevê as vazões para cada instante de tempo; a média aritmética das previsões é calculada. Para cada evento, são calculados os índices de performance, as vazões para diferentes tempos de permanência, as estatísticas das vazões previstas e os erros relativos dos parâmetros.

Portanto, este processo consiste na previsão por grupos mais simplificada; a integração da lógica difusa se dá pela obtenção dos grupos de eventos, que influenciam diretamente na calibração dos modelos.

3.4.2 Configuração 2 – Previsão por Grupo de Pertencimento

Nesta configuração, o processo de previsão por grupo de pertencimento ocorre de forma que, ao haver a identificação da ocorrência de um evento de cheia, são calculados os parâmetros relacionados ao agrupamento e, na sequência, a previsão é realizada pelo modelo que mais se aproxima da série, no instante de tempo da previsão. O processo é explicitado por:

- 1. Há a identificação da ocorrência de um evento de cheia.
- 2. Com base, exclusivamente, nos valores observados, são calculados os indicadores da cheia em andamento: tempo (tp) e vazão (Qp) de pico em União da Vitória, vazão inicial (Qt0) em União da Vitória, e vazões defasadas do tempo de pico de União da Vitória, para as estações de montante ((Qtp-1)FL, (Qtp-1)SB, (Qtp-2)RN); o cálculo considera a data de início do evento e o instante de tempo atual da cheia.
- 3. Os indicadores obtidos na Etapa 2 são normalizados (Equação 3.1).
- 4. Calcula-se a distância Euclidiana (Equação 3.7) entre as coordenadas (indicadores normalizados) do evento até o centroide normalizado de cada grupo correspondente a cada modelo.
- 5. Identifica-se com qual grupo o evento mais se assemelha, por meio da menor distância Euclidiana.
- 6. A previsão do próximo instante de tempo é efetuada a partir do modelo do grupo mais similar ao evento, conforme Etapa 5.
- 7. No próximo instante de tempo, são atualizados os indicadores de cheia, a partir das novas informações observadas (retorno à Etapa 2 e repetição das Etapas 3 a 6);

- 8. Ao final do evento, são calculados os índices de performance da previsão por grupo de pertencimento do evento e seus parâmetros estatísticos.
- 9. As etapas se repetem até que se identifique o fim do evento, e reiniciam com a identificação de uma nova cheia.

No modelo de previsão por grupo de pertencimento, portanto, a integração da lógica difusa se dá pela obtenção dos grupos de eventos, baseados no algoritmo *fuzzy c-means*; durante o processo de previsão, a essência da lógica difusa se concentra na definição da posição dos eventos em relação aos centroides: desde o início da previsão de determinada cheia, cada evento possui a possibilidade de pertencer a um dos três grupos, de forma que o pertencimento pode se alterar até a vazão de pico, ponto em que é definido o grupo final de pertencimento, sem considerar semelhança ou pertencimento a outro grupo.

Ou seja, o cálculo dos indicadores de cheia (Etapa 2) é dinâmico, e os resultados podem sofrer alterações conforme o evento progride: (i) no instante inicial, quando é identificado o início de uma cheia, os seis indicadores são calculados com base nas observações disponíveis até então e, assim, não se pode ter clareza sobre a qual grupo o evento pertence — ou com qual grupo a cheia mais se assemelha; (ii) com o avanço no tempo, os seis indicadores de cheia são atualizados a cada passo e o evento tende a diminuir suas semelhanças com certos grupos, ou migrar entre as classes; (iii) em um determinado instante, espera-se que as características — identificadas pelos indicadores de cheia — de um único grupo passem a predominar, estabilizando o pertencimento do evento a esta classe, cujo modelo conclui a previsão. O pertencimento, nesta etapa, indica o modelo que efetua a previsão naquele instante de tempo.

Portanto, os valores finais dos seis indicadores de cheia só podem ser obtidos após o término do evento, porém, conforme o evento avança, os indicadores se aproximam das características finais da cheia.

3.4.3 Configuração 3 – Previsão Ponderada por Distâncias Euclidianas

Nesta configuração, o processo de previsão ponderada pelas distâncias Euclidianas dos eventos aos centroides dos grupos ocorre de forma que, ao haver a identificação da ocorrência de um evento de cheia, são calculados os parâmetros referentes ao agrupamento, as distâncias Euclidianas, os pesos de ponderação para a previsão, e, na sequência, a previsão é composta pela média ponderada pelos pesos. O processo é explicitado por:

- 1. Há a identificação da ocorrência de um evento de cheia.
- 2. Com base nos valores observados, são calculados os indicadores da cheia em andamento: tempo (tp) e vazão (Qp) de pico em União da Vitória, vazão inicial (Qt0) em União da Vitória, e vazões defasadas do tempo de pico de União da Vitória, para as estações de montante ((Qtp-1)FL, (Qtp-1)SB, (Qtp-2)RN); o cálculo considera a data de início do evento e o instante de tempo atual da cheia.
- 3. Calcula-se a distância Euclidiana (Equação 3.7) entre as coordenadas (indicadores) do evento até o centroide de cada grupo ou a cada modelo.
- 4. São calculados os pesos de ponderação para a previsão, conforme Equação (3.28), onde P_i é o peso do *i-ésimo* modelo, d_i é a distância do evento até o *i-ésimo* centroide, correspondente ao *i-ésimo* modelo, c é o número de grupos ou modelos:

$$P_i = \frac{d_i^2}{\sum_{i=1}^c d_i^2} \tag{3.28}$$

- Para o próximo instante de tempo são efetuadas as previsões com base nos três modelos por grupo.
- 6. A previsão do evento é composta pela média ponderada das previsões da Etapa 5, sendo os pesos definidos na Etapa 4.
- 7. No próximo instante de tempo, são atualizados os indicadores de cheia, a partir das novas informações observadas (retorno à Etapa 2 e repetição das Etapas 3 a 6);
- 8. Ao final do evento, são calculados os índices de performance do evento e seus parâmetros estatísticos.
- 9. As etapas se repetem até que se identifique o fim do evento, e reiniciam com a identificação de uma nova cheia.

No modelo de previsão ponderada pelas distâncias Euclidianas, a integração da lógica difusa se inicia pela obtenção dos grupos de eventos, baseados no algoritmo *fuzzy c-means*. Durante o processo de previsão, há a definição dos pesos com base nas distâncias Euclidianas dos eventos até os centroides dos grupos; não há a identificação específica do pertencimento do evento a um determinado grupo. Portanto, nesta configuração, a previsão deixa de ser pela definição de um único grupo de pertencimento, visto que a previsão final considera todas as previsões por grupo, com diferentes pesos, que podem ser interpretadas pelo compartilhamento de características do evento a todos os grupos. Porém, o processo de cálculo dos indicadores de

cheia (Etapa 2) também é dinâmico, podendo os resultados sofrerem alterações com o avanço do evento, conforme destacado na Configuração 2 – Previsão por Grupo de Pertencimento.

3.4.4 Configuração 4 – Previsão Ponderada por Graus de Pertencimento

Na Configuração 4, o processo de previsão ponderada pelos graus de pertencimento dos eventos aos grupos, ocorre de forma que, ao haver a identificação da ocorrência de um evento de cheia, são calculados os parâmetros relacionados ao agrupamento, os graus de pertencimento do evento aos grupos, e, na sequência, a previsão é composta pela média ponderada pelos diferentes pertencimentos. O processo é explicitado por:

- 1. Há a identificação da ocorrência de um evento de cheia.
- 2. Com base, exclusivamente, nos valores observados, são calculados os indicadores da cheia em andamento: tempo (tp) e vazão (Qp) de pico em União da Vitória, vazão inicial (Qt0) em União da Vitória, e vazões defasadas do tempo de pico de União da Vitória, para as estações de montante ((Qtp-1)FL, (Qtp-1)SB, (Qtp-2)RN); o cálculo considera a data de início do evento e o instante de tempo atual da cheia.
- 3. Calcula-se a distância Euclidiana (Equação 3.7) entre as coordenadas (indicadores) do evento até o centroide de cada grupo ou a cada modelo.
- 4. São calculados os graus de pertencimento do evento aos três grupos, por meio da Equação 3.4.
- Para o próximo instante de tempo são efetuadas as previsões com base nos três modelos por grupo.
- 6. A previsão do evento é composta pela média ponderada das previsões da Etapa 5, sendo os pesos compostos pelos graus de pertencimento definidos na Etapa 4.
- 7. No próximo instante de tempo, são atualizados os indicadores de cheia, a partir das novas informações observadas (retorno à Etapa 2 e repetição das Etapas 3 a 6).
- 8. Ao final do evento, são calculados os índices de performance da previsão por graus de pertencimento do evento e seus parâmetros estatísticos.
- 9. As etapas se repetem até que se identifique o fim do evento, e reiniciam com a identificação de uma nova cheia.

No modelo de previsão ponderada pelos graus de pertencimento dos eventos aos grupos, a integração da lógica difusa se inicia pela obtenção dos grupos de eventos. Durante o

processo de previsão, a essência da lógica difusa está agregada à definição da posição dos eventos em relação aos centroides e à obtenção dos graus de pertencimento dos eventos aos grupos. Assim como na Configuração 3, nesta configuração, a previsão deixa de ser por um único grupo de pertencimento, visto que a previsão final considera todas as previsões por grupo, em maior ou menor grau.

Durante o processo de previsão, a essência da lógica difusa se concentra na definição da posição do evento em relação aos centroides e, consequentemente, na determinação do grau de pertencimento do evento aos grupos a cada instante de tempo: desde o início da previsão de uma cheia, cada evento possui a possibilidade de apresentar seu máximo grau de pertencimento a um dos três grupos, o qual terá maior peso na definição da previsão naquele instante de tempo; com o avançar do evento, os máximos graus de pertencimento podem se alterar até a vazão de pico ser atingida — ponto em que não há mais mudança nos pesos da previsão.

Ou seja, conforme observado nas Configurações 2 e 3, nesta previsão o cálculo dos indicadores de cheia (Etapa 2) também é dinâmico e os resultados podem sofrer alterações conforme a progressão do evento: (i) no instante inicial, ao ser identificado o início de um evento, os seis indicadores são calculados com base nas observações disponíveis e, assim, são calculados os graus de pertencimento do evento aos grupos, ou pesos da previsão (Etapa 4); (ii) com o avanço da cheia no tempo, os seis indicadores são atualizados a cada passo e o evento tende a diminuir suas semelhanças com certos grupos, ou migrar entre as classes, de forma que os pesos das previsões também possam se alterar; (iii) em um determinado instante de tempo, espera-se que as características do evento – ou indicadores de cheia – de um único grupo passem a predominar, estabilizando o máximo grau de pertencimento do evento a este grupo e fazendo com que a classe tenha maior peso na previsão até o final do processo.

Portanto, na previsão de um evento, é possível que seu máximo grau de pertencimento migre dentre os grupos a cada instante de tempo, até que se identifique o pico do evento e o início do período de recessão, o que significa que os valores finais dos seis indicadores de cheia só podem ser obtidos após o término da cheia, mas, com o avanço do evento, os indicadores se aproximam de suas características observadas.

4 RESULTADOS E DISCUSSÕES

O Capítulo 4 apresenta os resultados e discussões relacionados à análise de agrupamento, à geração dos modelos de previsão – por meio das etapas de treinamento, teste e ajuste dos modelos finais – e à previsão de vazões das Configurações 1 a 4 – previsões Média, por Grupo de Pertencimento, Ponderada pelas Distâncias Euclidianas e Ponderada pelos Graus de Pertencimento.

4.1 ANÁLISE DE AGRUPAMENTO

O agrupamento de cheias foi avaliado quanto à ferramenta considerada – algoritmo fuzzy c-means – e em relação a sua influência na caracterização dos eventos de cada grupo.

4.1.1 Agrupamento de eventos de cheia

O critério de parada do algoritmo foi atingido na iteração de número 116, partindo da função objetivo (*J*) igual a 20,66 e atingindo o valor mínimo de 11,92.

O grupo 1 – majoritariamente composto por vazões mais baixas em União da Vitória – foi contemplado com 70 eventos, enquanto o grupo 2, 102 eventos. O grupo 3 – de vazões mais elevadas em União da Vitória – apresentou 22 eventos. Os eventos que compõem cada grupo estão indicados no Apêndice A por suas respectivas datas de pico.

Na Figura 4.1, os eventos do Apêndice A foram dispostos em ordem cronológica e identificados por suas datas de ocorrência da vazão máxima. Na Figura 4.1, observa-se que os eventos do Grupo 2 se repetem em intervalos de tempo menores, enquanto no Grupo 3, dois eventos consecutivos estão, em geral, alguns anos distantes entre si, como era de se esperar, visto que a nomeação dos grupos se deu pela ordem de grandeza das vazões em União da Vitória.

Uma vez que a posição dos eventos nos grupos ocorre pela identificação do maior grau de similaridade (U) do evento com o grupo, são apresentados os valores máximos deste parâmetro na Figura 4.2 e, para facilitar sua interpretação, a Tabela 4.1. Ressalta-se que 153 eventos (mais de 78%) apresentaram grau de pertencimento ao seu grupo superior a 0,95, enquanto 17 eventos (em torno de 9%) apresentaram valor inferior a 0,70.

FIGURA 4.1 – DISTRIBUIÇÃO TEMPORAL DOS EVENTOS CLASSIFICADOS NOS QUATRO GRUPOS

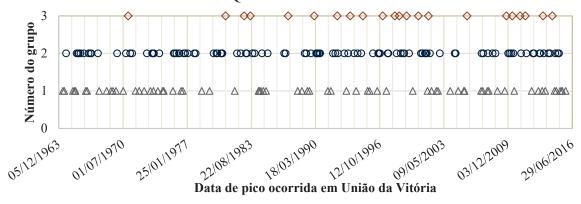


FIGURA 4.2 – MÁXIMOS GRAUS DE SIMILARIDADE DOS EVENTOS COM OS GRUPOS

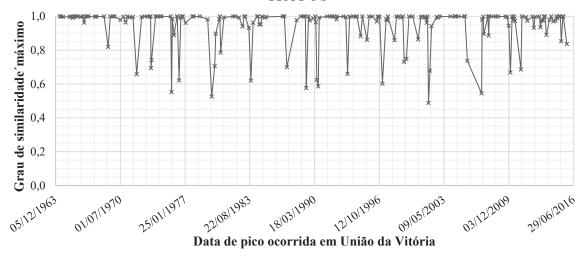


TABELA 4.1 – FREQUÊNCIAS DE INTERVALOS DE MÁXIMOS GRAUS DE SIMILARIDADE, POR SÉRIE DE EVENTOS

]	Intervalo de U	J		
		[0,95; 1,00]	[0,90; 0,95)	[0,80; 0,90)	[0,70; 0,80)	< 0,70	Soma
Todos os eventos	Frequência absoluta	153	6	12	6	17	194
	Frequência relativa	78,87%	3,09%	6,19%	3,09%	8,76%	100%
Cuma 1	Frequência absoluta	59	1	5	0	5	70
Grupo 1	Frequência relativa	84,3%	1,5%	7,1%	0,0%	7,1%	100%
Crupa 2	Frequência absoluta	82	3	5	4	8	102
Grupo 2	Frequência relativa	80,4%	2,9%	4,9%	3,9%	7,9%	100%
Crupo 3	Frequência absoluta	12	2	2	2	4	22
Grupo 3	Frequência relativa	54,5%	9,1%	9,1%	9,1%	18,2%	100%

A Tabela 4.1 apresenta também as frequências de ocorrência de certos intervalos de graus de similaridade para cada grupo de eventos: para os eventos dos Grupos 1 e 2, mais de 80% dos eventos apresentaram valores de U acima de 0,95, enquanto os eventos do Grupo 3

obtiveram resultados mais baixos – em torno de 54% dos eventos com grau de pertencimento acima de 0,95; 18% abaixo de 0,70 e cerca de 9% para cada um dos demais intervalos.

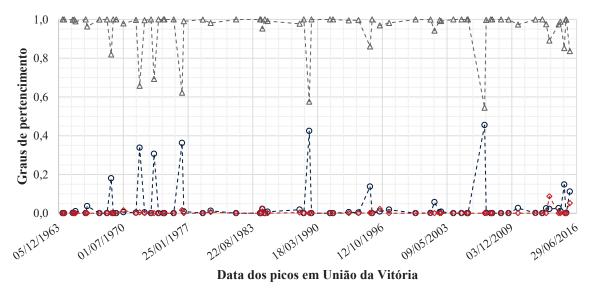
Ainda, a Tabela 4.2 apresenta os parâmetros estatísticos para os máximos graus de similaridade, descrevendo o comportamento dos intervalos de U da Tabela 4.1. As estatísticas indicaram valores próximos entre os Grupos 1 e 2, enquanto para o Grupo 3, os valores mínimo, médio e máximo diminuíram, distanciando-se do valor ideal (igual a 1,0), e a amplitude e o desvio-padrão aumentaram.

TABELA 4.2 – PARÂMETROS ESTATÍSTICOS DOS MÁXIMOS GRAUS DE SIMILARIDADE

Parâmetro	Grupo 1	Grupo 2	Grupo 3	
Total de observações	70	102	22	
Mínimo	0,5445	0,5257	0,4882	
Máximo	1,0000	1,0000	0,9997	
Amplitude	0,4555	0,4743	0,5115	
Médio	0,9562	0,9478	0,8826	
Desvio-padrão	0,1031	0,1131	0,1526	

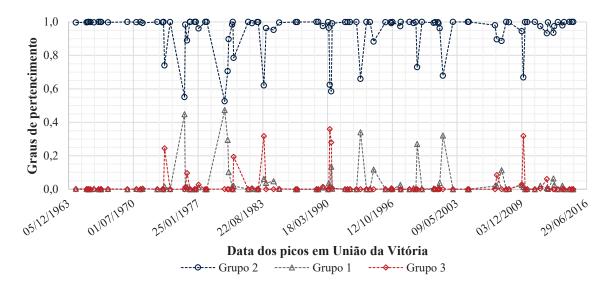
Em virtude do método, os eventos compartilham semelhanças com todos os grupos, em maior ou menor grau, por isso, foram elaboradas as Figuras 4.3 a 4.5. Observa-se da Figura 4.3 que os eventos cujos máximos pertencimentos foram inferiores a 0,90 compartilharam semelhanças com os Grupos 2 e 3 com graus de semelhança variando de 0,10 a 0,45; em grande maioria, o compartilhamento de características se deu com o Grupo 2.

FIGURA 4.3 – PERTENCIMENTO DOS EVENTOS DO GRUPO 1 AOS DEMAIS GRUPOS



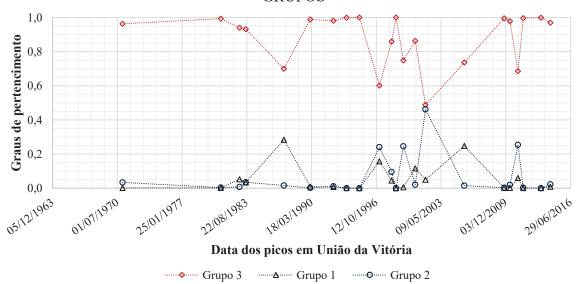
 A partir da Figura 4.4, pode-se afirmar que cerca de 15 eventos do Grupo 2 compartilharam informações relevantes com os demais grupos. Quando o valor máximo de pertencimento foi inferior a 0,90, os graus compartilhados com os Grupos 1 e 3 variaram de 0,10 a 0,45, aproximadamente. No entanto, para este grupo, a quantidade de semelhanças observada entre o Grupo 2 e os Grupos 1 e 3 foi dividida igualmente.

FIGURA 4.4 – PERTENCIMENTO DOS EVENTOS DO GRUPO 2 AOS DEMAIS GRUPOS.



Da mesma forma, os eventos do Grupo 3, cujos pertencimentos são exibidos na Figura 4.5, indicaram os mesmos níveis de compartilhamento de semelhanças com os outros grupos indicados nas Figuras 4.3 e 4.4, distribuídos igualmente dentre os Grupos 1 e 2.

FIGURA 4.5 – PERTENCIMENTO DOS EVENTOS DO GRUPO 3 AOS DEMAIS GRUPOS



Na Tabela 4.3 e na Figura 4.6, são apresentadas as coordenadas brutas e normalizadas dos centroides dos grupos formados, respectivamente; a Figura 4.6 permite a visualização da variação das coordenadas. Nota-se, por meio da Tabela 4.3 e da Figura 4.6, que o valor da coordenada vazão de pico em União da Vitória (Qp) (indicador 2) no Grupo 3, bem como as de vazões das estações a montante ((Qtp-1)FL, (Qtp-1)SB, (Qtp-2)RN) (indicadores 4, 5 e 6, respectivamente) do mesmo grupo foram superiores às respectivas coordenadas dos Grupos 1 e 2; já as coordenadas *tempo de pico* (indicador 1) e *vazão inicial* (indicador 3) *em União da Vitória* para o Grupo 3 se mantiveram, respectivamente, inferior e intermediária às coordenadas dos outros grupos. Estas características evidenciam a maior magnitude dos eventos do último grupo.

Com exceção dos valores das coordenadas de vazão inicial dos Grupos 1 e 2, os demais indicadores mantiveram-se próximos entre os pares para estes grupos. A coordenada de vazão inicial do Grupo 2 apresentou-se superior à mesma coordenada dos demais grupos. Esta foi a principal diferenciação entre os Grupos 1 e 2.

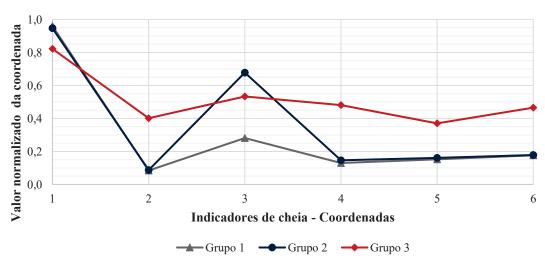


FIGURA 4.6 – VARIAÇÃO DAS COORDENADAS NORMALIZADAS DOS CENTROIDES DOS GRUPOS

TABELA 4.3 – VALORES BRUTOS DAS COORDENADAS DOS CENTROIDES DOS GRUPOS

_	União da Vitória	União da Vitória	União da Vitória	Fluviópolis	São Bento	Rio Negro
Grupo	t _p (dias)	Q _p (m ³ /s)	Q ₀ (m ³ /s)	$(Q_{tp-1})_{FL}$ (m^3/s)	$(Q_{tp-1})_{SB}$ (m^3/s)	$(\mathbf{Q}_{tp-2})_{RN}$ (\mathbf{m}^3/s)
Grupo 1	12	1075,2	246,6	669,7	64,3	148,3
Grupo 2	14	1087,9	434,3	730,4	66,9	150,0
Grupo 3	41	2490,1	365,9	1934,8	130,2	353,2

4.1.2 Caracterização dos grupos de eventos

Considerando os períodos de treinamento, teste e previsão, foram determinados os parâmetros estatísticos para os diferentes conjuntos: Todos os Eventos, Grupo 1, Grupo 2 e Grupo 3. A Tabela 4.4 apresenta as estatísticas para todo o período de dados, enquanto as Tabelas 4.5 a 4.10 subdividem os períodos e as amostras da validação cruzada.

TABELA 4.4 – PARÂMETROS ESTATÍSTICOS DOS EVENTOS DE TODO O PERÍODO DE DADOS – 1964 A 2015

	Todos os eventos	Grupo 1	Grupo 2	Grupo 3
Número de eventos	194	70	102	22
Tamanho da amostra	7356	2078	3508	1770
Mínima (m³/s)	114,00	114,00	314,20	148,10
Máxima (m³/s)	5156,70	1939,50	1915,10	5156,70
Média (m³/s)	961,50	836,61	842,34	1344,31
Desvio-padrão (m³/s)	511,54	349,56	315,55	744,30
Assimetria	2,29	0,59	0,92	1,52

Na Tabela 4.4, nota-se uma diferença entre os parâmetros considerando todos os eventos e os parâmetros dos grupos. O comportamento dos parâmetros, conhecido o processo de nomeação dos grupos, ocorre de forma que, em geral, o conjunto com todos os eventos se mantenha em uma posição intermediária entre os Grupos 1 e 2 – com valores próximos entre si – e o Grupo 3. A partir desta observação, pode-se afirmar que o agrupamento, de fato, colabora com a identificação de padrões e características semelhantes dos eventos de um mesmo grupo, atuando na redução das médias, dos desvios-padrão e das assimetrias dos primeiros grupos, e no aumento destes parâmetros para o Grupo 3, em virtude da maior magnitude dos eventos que o compõem.

Este mesmo comportamento é observado no período de treinamento e teste (Tabela 4.5), sendo que os parâmetros estatísticos dos eventos de 1964 a 2004 se mantiveram próximos àqueles do período completo (1964 a 2015 – Tabela 4.4), para todos os conjuntos.

TABELA 4.5 – PARÂMETROS ESTATÍSTICOS DOS EVENTOS DO PERÍODO DE TREINAMENTO E TESTE – 1964 A 2004

	Todos os eventos	Grupo 1	Grupo 2	Grupo 3
Número de eventos	147	51	81	15
Tamanho da amostra	5748	1578	2900	1270
Mínima (m³/s)	114,00	114,00	327,00	148,10
Máxima (m³/s)	5156,70	1939,50	1915,10	5156,70
Média (m³/s)	950,27	830,27	843,86	1342,32
Desvio-padrão (m³/s)	512,18	337,86	314,75	788,83
Assimetria	2,57	0,57	0,91	1,65

Já para o período de previsão (Tabela 4.6), as estatísticas de todos os eventos e dos Grupos 1 e 2 se mantiveram próximas às do período completo (Tabela 4.4) e às do período de treinamento e teste (Tabela 4.5), porém, para as cheias do Grupo 3, os parâmetros se distanciaram dos resultados dos demais períodos, com exceção da média; ou seja, os valores de mínima, máxima, desvio-padrão e assimetria sofreram alterações no Grupo 3 da Tabela 4.6.

TABELA 4.6 – PARÂMETROS ESTATÍSTICOS DOS EVENTOS DO PERÍODO DE PREVISÃO – 2005 A 2015

	Todos os eventos	Grupo 1	Grupo 2	Grupo 3
Número de eventos	47	19	21	7
Tamanho da amostra	1608	500	608	500
Mínima (m³/s)	132,20	132,20	314,20	224,70
Máxima (m³/s)	3385,50	1901,00	1864,00	3385,50
Média (m³/s)	1001,67	856,62	835,05	1349,34
Desvio-padrão (m³/s)	507,34	383,84	319,46	617,61
Assimetria	1,30	0,59	1,01	0,73

Ainda, visto que o período de treinamento e teste passaram pelo processo de validação cruzada, faz-se necessário avaliar o comportamento dos parâmetros estatísticos deste período considerando a estratégia de amostragem (Figura 3.6). As relações entre as estatísticas das amostras de treinamento e teste e as estatísticas do período total (amostra/período total) foram apresentadas nas Tabelas 4.7 a 4.10; as Tabelas destacam os valores dos parâmetros estatísticos com as maiores alterações observadas quando comparados ao período de 1964 a 2004 (Tabela 4.5). Nas Tabelas, o número de eventos e o tamanho das amostras são apresentados em valores absolutos.

TABELA 4.7 – RELAÇÕES ENTRE OS PARÂMETROS ESTATÍSTICOS DOS PERÍODOS DE TREINAMENTO E TESTE, PARA A ESTRATÉGIA DE AMOSTRAGEM – TODOS OS EVENTOS

	Amostra	1	2	3	4	5	6	7	8
	Número de eventos	77	70	72	74	68	82	79	77
0	Tamanho da amostra	2850	2772	2790	3314	2862	3142	3105	2556
ent	Mínima	1,00	1,30	<u>1,27</u>	<u>1,19</u>	<u>1,19</u>	1,00	1,00	1,00
Treinamento	Máxima	1,00	1,00	1,00	1,00	0,77	0,77	0,53	0,47
rei	Média	0,97	0,99	1,01	1,08	1,04	1,00	0,98	0,89
	Desvio-padrão	1,03	1,05	<u>1,14</u>	<u>1,15</u>	0,97	0,94	0,83	<u>0,70</u>
	Assimetria	<u>1,27</u>	<u>1,19</u>	<u>1,17</u>	0,97	<u>0,70</u>	0,75	0,50	<u>0,48</u>
	Número de eventos	70	77	75	73	79	65	68	70
	Tamanho da amostra	2898	2976	2958	2434	2886	2606	2643	3192
e	Mínima	<u>1,19</u>	1,00	1,00	1,00	1,00	1,30	1,27	<u>1,19</u>
Teste	Máxima	0,77	0,77	0,53	<u>0,47</u>	1,00	1,00	1,00	1,00
L	Média	1,03	1,00	0,99	0,89	0,96	1,00	1,03	1,09
	Desvio-padrão	0,97	0,95	0,85	0,70	1,02	1,07	<u>1,16</u>	<u>1,16</u>
	Assimetria	0,71	0,75	0,49	0,48	1,27	<u>1,19</u>	<u>1,14</u>	0,96

Observa-se na Tabela 4.7 – Todos os eventos – que os valores de mínima e máxima se inverteram entre os períodos de treinamento e teste: as amostras 2 a 5 tiveram alterações nas vazões mínimas do treinamento, enquanto no teste as demais amostras (1, 6, 7 e 8) foram aquelas que se diferenciaram; as amostras 5 a 8 reduziram seus valores máximos no treinamento, enquanto no teste a redução foi observada apenas nas amostras 1 a 4. No treinamento da Tabela 4.7, metade das amostras tiveram desvio-padrão elevado, enquanto a assimetria conservou seus valores apenas na amostra 4. Comportamento similar foi observado no teste da mesma Tabela, porém, a assimetria se manteve equivalente apenas na amostra 8.

TABELA 4.8 – RELAÇÕES ENTRE OS PARÂMETROS ESTATÍSTICOS DOS PERÍODOS DE TREINAMENTO E TESTE, PARA A ESTRATÉGIA DE AMOSTRAGEM – GRUPO 1

	Amostra	1	2	3	4	5	6	7	8
	Número de eventos	31	27	22	19	19	26	30	32
9	Tamanho da amostra	963	863	654	606	596	763	963	972
Treinamento	Mínima	1,00	<u>1,35</u>	1,27	<u>1,19</u>	<u>1,19</u>	1,00	1,00	1,00
lam	Máxima	1,00	1,00	1,00	1,00	0,91	0,80	0,95	0,95
eir	Média	1,01	1,05	0,99	1,01	0,99	0,95	1,00	0,99
Ξ	Desvio-padrão	1,05	1,08	1,00	1,04	0,92	0,87	0,98	0,98
	Assimetria	<u>1,10</u>	0,96	<u>1,11</u>	0,96	0,57	0,54	0,98	0,98
	Número de eventos	20	24	29	32	32	25	21	19
	Tamanho da amostra	615	715	924	972	982	815	615	606
e	Mínima	<u>1,19</u>	1,00	1,00	1,00	1,00	1,35	1,27	<u>1,19</u>
Teste	Máxima	0,91	0,80	0,95	0,95	1,00	1,00	1,00	1,00
I	Média	0,98	0,94	1,01	0,99	1,01	1,05	1,00	1,01
	Desvio-padrão	0,91	0,87	1,00	0,98	1,05	1,09	1,03	1,04
	Assimetria	<u>3,17</u>	<u>3,38</u>	<u>2,19</u>	<u>2,16</u>	<u>5,70</u>	<u>5,35</u>	<u>5,12</u>	<u>4,32</u>

TABELA 4.9 – RELAÇÕES ENTRE OS PARÂMETROS ESTATÍSTICOS DOS PERÍODOS DE TREINAMENTO E TESTE, PARA A ESTRATÉGIA DE AMOSTRAGEM – GRUPO 2

	Amostra	1	2	3	4	5	6	7	8
	Número de eventos	42	38	43	43	38	46	41	41
2	Tamanho da amostra	1508	1455	1608	1610	1375	1563	1400	1370
nen	Mínima	1,00	1,00	1,00	1,00	1,05	1,05	1,06	1,04
Treinamento	Máxima	0,94	0,99	1,00	1,00	1,00	1,00	0,92	0,92
rei	Média	0,99	1,01	1,02	1,04	1,01	1,00	0,98	0,95
	Desvio-padrão	1,02	1,06	1,06	1,03	0,98	0,97	0,91	0,93
	Assimetria	1,05	<u>1,14</u>	1,07	<u>0,87</u>	0,92	<u>0,90</u>	<u>0,79</u>	<u>1,16</u>
	Número de eventos	39	43	38	38	43	35	40	40
	Tamanho da amostra	1392	1445	1292	1290	1561	1337	1500	1530
e	Mínima	1,05	1,05	1,06	1,04	1,00	1,00	1,00	1,00
Teste	Máxima	1,00	1,00	0,92	0,92	0,94	0,94	1,00	1,00
	Média	1,01	0,99	0,98	0,96	0,99	1,00	1,02	1,04
	Desvio-padrão	0,98	0,94	0,92	0,94	1,02	1,04	1,08	1,04
	Assimetria	0,93	<u>0,77</u>	<u>0,79</u>	<u>1,17</u>	1,06	1,10	1,06	<u>0,85</u>

A partir da Tabela 4.8, até a Tabela 4.10, são considerados os eventos por grupos e, devido a isso, são observadas, em geral, maiores discrepâncias (maior distância de 1,0). Em virtude dos períodos de treinamento e teste (Figura 3.6), mais uma vez se observa, por meio das Tabelas 4.7 a 4.10, semelhanças entre as relações de parâmetros estatísticos para os pares de amostra de treinamento e teste, respectivamente: 1 e 5, 2 e 6, 3 e 7, 4 e 8, 5 e 1, 6 e 2, 7 e 3, 8 e 4; maiores divergências foram observadas para as assimetrias. As estatísticas dos conjuntos de eventos do modelo final – após a validação cruzada– correspondem à Tabela 4.5.

TABELA 4.10 – RELAÇÕES ENTRE OS PARÂMETROS ESTATÍSTICOS DOS PERÍODOS DE TREINAMENTO E TESTE, PARA A ESTRATÉGIA DE AMOSTRAGEM – GRUPO 3

	Amostra	1	2	3	4	5	6	7	8
	Número de eventos	4	5	7	12	11	10	8	4
.0	Tamanho da amostra	379	454	528	1098	891	816	742	214
Treinamento	Mínima	<u>1,66</u>	1,00	1,00	1,00	1,00	<u>1,79</u>	1,79	<u>1,79</u>
nan	Máxima	1,00	1,00	1,00	1,00	0,77	0,77	0,53	<u>0,47</u>
rei	Média	1,08	1,04	1,09	1,02	0,97	0,98	0,94	0,89
	Desvio-padrão	<u>1,28</u>	<u>1,22</u>	<u>1,28</u>	1,03	0,85	0,86	0,73	0,76
	Assimetria	1,04	1,09	0,93	1,03	<u>0,66</u>	<u>0,67</u>	0,33	0,27
	Número de eventos	11	10	8	3	4	5	7	11
	Tamanho da amostra	891	816	742	172	379	454	528	1056
ره	Mínima	1,00	<u>1,79</u>	<u>1,79</u>	<u>1,79</u>	<u>1,66</u>	1,00	1,00	1,00
Teste	Máxima	0,77	0,77	0,53	0,47	1,00	1,00	1,00	1,00
	Média	0,97	0,98	0,94	0,90	1,08	1,04	1,09	1,02
	Desvio-padrão	0,85	0,86	0,73	0,79	1,28	1,22	1,28	1,04
	Assimetria	0,66	0,67	0,33	0,27	1,04	1,09	0,93	1,02

4.2 GERAÇÃO DOS MODELOS DE PREVISÃO

Considerando os 64 conjuntos de parâmetros e as 8 amostras da validação cruzada, são apresentados, para a etapa de treinamento dos Modelos de série completa, de todos os eventos e dos Grupos 1, 2 e 3, os mapas de calor do Apêndice B com a variação das métricas de treinamento para cada combinação de parâmetros, e os gráficos dos Apêndices C e D com a distribuição dos resultados dos coeficientes nas etapas de treinamento e teste, respectivamente.

O objetivo é verificar o impacto que determinados eventos exercem sobre os resultados dos modelos; o modelo "ideal" para cada situação é aquele que evidencia menor variabilidade de resultados entre as amostras para um mesmo conjunto de parâmetros, indicando que o modelo é capaz de bem representar a série de dados, independente do período de treinamento.

4.2.1 Treinamento

Para a etapa de treinamento, o Apêndice B traz a evolução de Kling-Gupta, de Nash-Sutcliffe, de Nash-Sutcliffe Modificado e do Erro Relativo Absoluto Médio para os 64 conjuntos de parâmetros em suas 8 amostras da Estratégia de Amostragem; a Figura 4.7 apresenta as médias dos coeficientes entre as amostras — cujos valores são apresentados no Apêndice B — para cada conjunto de parâmetros, durante o treinamento dos modelos.

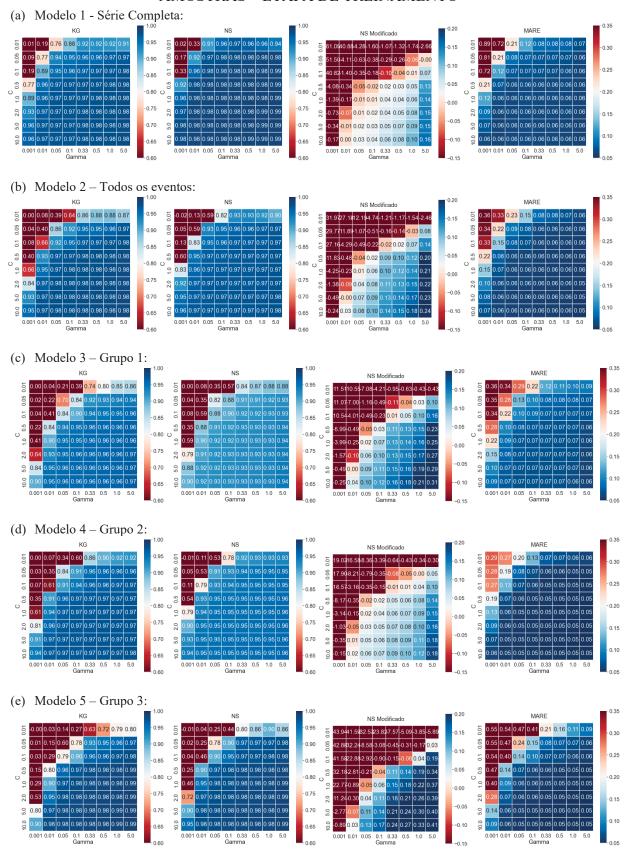
Os resultados da Figura 4.7 indicam que quanto maiores os valores de C e γ , melhores são os desempenhos dos modelos; ou seja, quanto mais próximos do canto inferior direito dos mapas de calor, mais adequadas são as métricas — quanto menores C e γ , piores são os comportamentos. Em geral, o parâmetro C superior a 1,0 apresentou resultados adequados para todos os valores de γ , para KG, NS e MARE. Para o desempenho de NS Modificado, tem-se, a partir da Figura 4.7, que os melhores resultados se concentraram nas combinações de parâmetros com C superior ou igual a 0,5 e γ superior ou igual a 0,05.

Ressalta-se que o aumento de C e γ aumenta a velocidade de processamento, às custas de pouca melhoria de resultado a partir de determinado ponto; a diagonal formada a partir do canto inferior esquerdo (C = 10,0 e γ = 0,001) de um mapa de calor e seu canto superior direito (C = 0,01 e γ = 5,0) representa maior uniformidade de resultados entre amostras (Apêndice B), que, se representadas por meio de linhas contínuas, atingiriam uma relativa sobreposição entre si. Este comportamento indica a redução do desvio-padrão das amostras de determinada combinação de parâmetros e uma maior uniformidade de valores entre combinações.

Conforme aumenta a complexidade dos modelos, a variabilidade entre os conjuntos de parâmetros e entre as amostras se eleva, como no Modelo 1 (Figura 4.7a), onde o intervalo de resultados para diferentes conjuntos de parâmetros foi menor quando comparado àquele definido pelo Modelo 5 (Figura 4.7e), sendo os resultados do Modelo 1, portanto, mais uniformes que os do Modelo 5.

As distribuições dos resultados da Figura 4.7 e das Figuras do Apêndice B estão representadas nas Figuras do Apêndice C (*boxplots*), nos quais estão em evidência os resultados dos diferentes conjuntos de parâmetros dos modelos, para cada amostra da estratégia de validação cruzada e para sua média. Evidencia-se, novamente, a maior variabilidade de resultados entre as amostras de um mesmo modelo, principalmente, no Modelo 5 – Eventos do Grupo 3.

FIGURA 4.7 – RESULTADOS DE KG, NS, NS MODIFICADO E MARE PARA DIFERENTES CONJUNTOS DE PARÂMETROS E MODELOS – MÉDIA ENTRE AS AMOSTRAS – ETAPA DE TREINAMENTO



As médias das amostras dos *boxplots* do Apêndice C deram origem às Tabelas 4.11 a 4.14, nas quais se observam amplitudes interquartis e desvios-padrão mais elevados para os Modelos 3 e 5; apesar de seus resultados adequados, os referidos modelos apresentaram também as piores médias e medianas para praticamente todas as métricas, conforme as Tabelas 4.11 a 4.14. Os Modelos 2 e 4 apresentaram os resultados mais estáveis.

TABELA 4.11 – AVALIAÇÃO MÉDIA DO COMPORTAMENTO DE KG PARA DIFERENTES CONJUNTOS DE PARÂMETROS – FASE DE TREINAMENTO

		Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
So	Quantidade	64	64	64	64	64
ınt	Média	0,900	0,860	0,808	0,852	0,802
conjuntos tros	Desvio-padrão	0,209	0,252	0,280	0,257	0,309
	Mínimo	0,011	0,003	0,001	0,005	0,000
	25%	0,947	0,912	0,839	0,913	0,791
cas (50%	0,972	0,973	0,950	0,967	0,972
ísti de	75%	0,975	0,978	0,959	0,971	0,982
Estatísticas de paı	Máximo	0,980	0,982	0,967	0,975	0,988
	Amplitude interquartis	0,028	0,066	0,120	0,058	0,191

TABELA 4.12 – AVALIAÇÃO MÉDIA DO COMPORTAMENTO DE NS PARA DIFERENTES CONJUNTOS DE PARÂMETROS – FASE DE TREINAMENTO

		Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
SO	Quantidade	64	64	64	64	64
unt	Média	0,929	0,884	0,817	0,864	0,839
conjuntos tros	Desvio-padrão	0,190	0,228	0,243	0,228	0,282
<u> </u>	Mínimo	0,015	-0,018	0,001	-0,009	-0,008
. dos râmo	25%	0,977	0,947	0,877	0,931	0,897
Estatísticas de paı	50%	0,984	0,969	0,920	0,950	0,976
tísti de	75%	0,985	0,972	0,929	0,953	0,981
staı	Máximo	0,986	0,975	0,942	0,958	0,986
	Amplitude interquartis	0,008	0,024	0,052	0,022	0,084

TABELA 4.13 – AVALIAÇÃO MÉDIA DO COMPORTAMENTO DE NS MODIFICADO PARA DIFERENTES CONJUNTOS DE PARÂMETROS – FASE DE TREINAMENTO

		Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
So	Quantidade	64	64	64	64	64
unt	Média	-3,442	-2,707	-1,254	-1,678	-6,071
conjuntos tros	Desvio-padrão	11,973	7,369	3,039	4,526	12,588
	Mínimo	-61,052	-31,974	-11,509	-19,027	-43,936
	25%	-0,443	-0,647	-0,492	-0,351	-3,268
Estatísticas de par	50%	-0,013	0,023	0,038	0,020	-0,042
fisti de	75%	0,045	0,117	0,138	0,074	0,185
stat	Máximo	0,158	0,240	0,305	0,185	0,410
	Amplitude interquartis	0,489	0,764	0,631	0,425	3,453

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
Quantidade	64	64	64	64	64
Média	0,120	0,091	0,111	0,080	0,142
Desvio-padrão	0,177	0,076	0,083	0,062	0,155
Mínimo	0,059	0,054	0,058	0,048	0,045
25%	0,062	0,057	0,068	0,052	0,053
50%	0,064	0,060	0,073	0,054	0,061
75%	0,073	0,076	0,100	0,066	0,144
Máximo	0,885	0,356	0,358	0,290	0,552
Amplitude interquartis	0.011	0.018	0.032	0.014	0.091

0,018

0,032

0,014

0,091

0,011

Amplitude interquartis

TABELA 4.14 – AVALIAÇÃO MÉDIA DO COMPORTAMENTO DE MARE PARA DIFERENTES CONJUNTOS DE PARÂMETROS – FASE DE TREINAMENTO

Portanto, uma vez que os resultados das Tabelas 4.11 a 4.14 e da Figura 4.7 são oriundas da média dos coeficientes das amostras, os eventos que as compõem (Figura 3.6) exercem grande influência sobre os resultados das amostras: a maior variabilidade dos resultados dos modelos – principalmente aqueles treinados por grupo de eventos – entre as amostras se deve à abrangência – ou não – de eventos com características destoantes dos demais, no período de treinamento; isto pode significar uma melhor – ou pior – performance do modelo. Em outras palavras, como esperado, quanto maior a diferença entre os dados de entrada para treinamento do modelo – em virtude das diferentes amostras –, os valores de KG, NS, NS Modificado e MARE mais se distanciam da média das amostras.

Assim, juntamente com o comportamento da diagonal dos mapas de calor que indica maior uniformidade de resultados (Figura 4.7), esta característica indica a relevância da validação cruzada e a necessidade de redução da possibilidade de *under* ou *overfitting* (Hallouin et al., 2020); os conjuntos de parâmetros cujos valores se concentram na região central dos gráficos (Figura 4.7) tendem a reproduzir melhor os dados não utilizados na fase de treinamento. Neste contexto, os piores conjuntos de parâmetros foram observados nos modelos por grupo (Figura 4.7c, d, e). Segundo o NS Modificado, contudo, os Modelos 3 e 4 – dos Grupos 1 e 2, respectivamente – apresentaram melhor comportamento geral.

Desta forma, ao se iniciar a avaliação dos modelos por grupo na etapa de treinamento, a observação dos eventos pertencentes a cada um se torna imprescindível. Para o Modelo 3 -Modelo de Eventos do Grupo 1 – são apresentados, na Tabela 4.15, os eventos com as 10 maiores vazões de pico deste grupo, bem como, aqueles com as 10 menores vazões de pico; no último caso, em primeiro lugar está a menor vazão dentre todos os eventos. Para melhor entendimento das métricas das Amostras, é apresentada a Tabela 4.16.

TABELA 4.15 – RANQUEAMENTO DAS MAIORES E MENORES VAZÕES DE PICO DO GRUPO 1, EM UNIÃO DA VITÓRIA

	10 maiores	vazões de pico		10 menores vazões de pico					
Ordem	Pico do evento	Ordem	Pico do evento	Ordem	Pico do evento	Ordem	Pico do evento		
1°	09/08/1984	6°	27/10/2004	1°	01/09/2001	6°	11/10/1969		
2°	30/08/1972	7°	08/07/1970	2°	21/02/1985	7°	23/04/1973		
3°	25/05/1988	8°	15/07/1965	3°	14/08/1973	8°	07/08/2002		
4°	16/05/1979	9°	14/06/1976	4°	02/11/1968	9°	27/01/1968		
5°	12/07/1996	10°	16/05/1965	5°	28/06/1964	10°	07/05/1964		

TABELA 4.16 – ESTRATÉGIA DE AMOSTRAGEM DOS EVENTOS DO GRUPO 1 – FASE DE TREINAMENTO

			Picos dos even	tos do Grupo 1			
Amostra 1	Amostra 2	Amostra 3	Amostra 4	Amostra 5	Amostra 6	Amostra 7	Amostra 8
07/05/1964	08/04/1969	28/01/1974	16/05/1979	25/05/1988	07/05/1964	07/05/1964	07/05/1964
28/06/1964	22/06/1969	26/07/1974	11/12/1981	27/10/1988	28/06/1964	28/06/1964	28/06/1964
16/05/1965	11/10/1969	05/09/1974	19/05/1984	15/05/1989	16/05/1965	16/05/1965	16/05/1965
15/07/1965	08/07/1970	12/08/1975	18/06/1984	05/08/1989	15/07/1965	15/07/1965	15/07/1965
24/08/1965	15/10/1971	14/06/1976	09/08/1984	24/06/1991	24/08/1965	24/08/1965	24/08/1965
07/09/1966	29/02/1972	20/08/1976	13/11/1984	12/10/1991	07/09/1966	07/09/1966	07/09/1966
30/10/1966	30/08/1972	25/07/1978	21/02/1985	15/05/1993	30/10/1966	30/10/1966	30/10/1966
27/01/1968	23/04/1973	16/05/1979	25/05/1988	13/05/1994	27/01/1968	27/01/1968	27/01/1968
02/11/1968	14/08/1973	11/12/1981	27/10/1988	18/07/1995	02/11/1968	02/11/1968	02/11/1968
08/04/1969	28/01/1974	19/05/1984	15/05/1989	29/09/1995	27/10/1988	08/04/1969	08/04/1969
22/06/1969	26/07/1974	18/06/1984	05/08/1989	12/07/1996	15/05/1989	22/06/1969	22/06/1969
11/10/1969	05/09/1974	09/08/1984	24/06/1991	21/06/1997	05/08/1989	11/10/1969	11/10/1969
08/07/1970	12/08/1975	13/11/1984	12/10/1991	02/03/2000	24/06/1991	08/07/1970	08/07/1970
15/10/1971	14/06/1976	21/02/1985	15/05/1993	01/09/2001	12/10/1991	15/10/1971	15/10/1971
29/02/1972	20/08/1976	25/05/1988	13/05/1994	29/01/2002	15/05/1993	29/02/1972	29/02/1972
30/08/1972	25/07/1978	27/10/1988	18/07/1995	07/08/2002	13/05/1994	30/08/1972	30/08/1972
23/04/1973	16/05/1979	15/05/1989	29/09/1995	02/10/2002	18/07/1995	23/04/1973	23/04/1973
14/08/1973	11/12/1981	05/08/1989	12/07/1996	01/01/2004	29/09/1995	14/08/1973	14/08/1973
28/01/1974	19/05/1984	24/06/1991	21/06/1997	27/10/2004	12/07/1996	13/05/1994	28/01/1974
26/07/1974	18/06/1984	12/10/1991			21/06/1997	18/07/1995	26/07/1974
05/09/1974	09/08/1984	15/05/1993			02/03/2000	29/09/1995	05/09/1974
12/08/1975	13/11/1984	13/05/1994	_		01/09/2001	12/07/1996	12/08/1975
14/06/1976	21/02/1985				29/01/2002	21/06/1997	14/06/1976
20/08/1976	25/05/1988				07/08/2002	02/03/2000	20/08/1976
25/07/1978	27/10/1988				02/10/2002	01/09/2001	25/07/1978
16/05/1979	15/05/1989				01/01/2004	29/01/2002	02/03/2000
11/12/1981	05/08/1989	_			27/10/2004	07/08/2002	01/09/2001
19/05/1984						02/10/2002	29/01/2002
18/06/1984	_					01/01/2004	07/08/2002
09/08/1984						27/10/2004	02/10/2002
13/11/1984	_						01/01/2004
							27/10/2004

A Tabela 4.16 apresenta a estratégia de amostragem dos eventos do Grupo 1, indicando quais eventos deste Grupo foram utilizados na fase de treinamento. As cores destacadas em

ambas as Tabelas indicam o aparecimento do evento da Tabela 4.15 nas Amostras da Tabela 4.16. De maneira geral, cada evento destacado na Tabela 4.15 aparece quatro vezes na Tabela 4.16, em amostras diferentes, exceto o evento com pico ocorrido em 21/02/1985, que está destacado em três amostras na Tabela 4.16. O maior pico observado (09/08/1984) apresentouse nas Amostras 1 a 4, enquanto o menor (01/09/2001), nas Amostras 5 a 8. Ou seja, ambos os eventos não ocorreram simultaneamente em diferentes amostras.

A Tabela 4.17 indica as frequências absolutas (à esquerda) e relativas (à direita) dos eventos da Tabela 4.16 – estratégia de amostragem do Grupo 1 –, presentes na Tabela 4.15 – ranqueamento de vazões do Grupo 1. Com base na Tabela 4.17, tem-se que as Amostras 3, 4 e 5 apresentaram a menor frequência absoluta de eventos na fase de treinamento, bem como, as menores frequências relativas, somadas, de incidência das maiores e menores vazões de pico em União da Vitória: a Amostra 5, por exemplo, indicou 15,8% de seus eventos entre os 10 maiores picos e 10,5% entre os 10 menores picos; logo, suas vazões apresentaram, majoritariamente, os valores intermediários dos eventos do Grupo 1 e menor amplitude.

TABELA 4.17 – FREQUÊNCIAS ABSOLUTAS (À ESQUERDA) E RELATIVAS (À DIREITA) DOS MAIORES E MENORES PICOS DO GRUPO 1 – FASE DE TREINAMENTO

		Amostras								
	1	2	3	4	5	6	7	8		
Total de eventos	31	27	22	19	19	26	30	31		
Maiores	7 22,6%	6 22,2%	4 18,2%	4 21,1%	3 15,8%	4 15,4%	6 20,0%	6 18,8%		
Menores	7 22,6%	4 14,8%	1 04,5%	1 05,3%	2 10,5%	6 23,1%	9 30,0%	9 28,1%		

Por outro lado, a Amostra 1 do Modelo 3 apresentou 22,6% (o maior percentual) de eventos ranqueados entre os maiores picos, e o mesmo percentual, entre os menores. Portanto, quase 50% de seus eventos mantiveram picos nos dois extremos e, consequentemente, a outra metade apontou picos intermediários, fazendo com que os picos dos eventos estivessem bem distribuídos entre as faixas de vazão da amostra. Isso pode ter contribuído com a menor amplitude das métricas na Amostra 1 do referido modelo (Apêndice C – Figura C.3).

As Amostras 7 e 8 do Modelo 3 também apresentaram cerca de 50% de seus picos nos dois extremos, fazendo com que suas vazões se comportassem de modo similar às da Amostra 1. A diferença entre as Amostras 7 e 8 para a Amostra 1 se concentra no fato de a Amostra 1 apresentar maior quantidade de eventos ranqueados nos extremos máximos, enquanto as outras apresentaram maiores quantidades de eventos ranqueados nos extremos mínimos.

Para o Modelo 4 – Modelo de Eventos do Grupo 2 – foram apresentados, na Tabela 4.18, os eventos com as 10 maiores vazões de pico do grupo, bem como, aqueles com as 10

menores vazões de pico; neste caso, na primeira posição está a menor vazão dentre todos os eventos. Para melhor compreensão do comportamento das métricas das Amostras, apresenta-se a Tabela 4.19, com a estratégia de amostragem dos eventos do Grupo 2; esta Tabela indica quais eventos deste Grupo foram utilizados no treinamento. As cores destacadas em ambas as Tabelas indicam o aparecimento do evento da Tabela 4.18 nas Amostras da Tabela 4.19.

TABELA 4.18 – RANQUEAMENTO DAS MAIORES E MENORES VAZÕES DE PICO DO GRUPO 2, EM UNIÃO DA VITÓRIA

	10 maiores vazões de pico				10 menores vazões de pico				
Ordem	Pico do evento	Ordem	Pico do evento	Ordem	Pico do evento	Ordem	Pico do evento		
1°	07/06/1990	6°	06/09/1980	1°	24/12/1983	6°	11/12/1977		
2°	25/09/1989	7°	15/12/1975	2°	31/12/2000	7°	28/02/1982		
3°	20/07/1982	8°	01/08/1990	3°	28/01/1980	8°	14/04/1976		
4°	31/08/1973	9°	13/10/1975	4°	27/08/1997	9°	25/03/1974		
5°	03/10/1983	10°	10/06/1971	5°	12/12/1998	10°	29/05/2001		

Em geral, cada evento destacado na Tabela 4.18 – ranqueamento de vazões do Grupo 2 – se destaca quatro vezes na Tabela 4.19 – estratégia de amostragem do Grupo 2 –, em amostras diferentes, com exceção do evento com pico ocorrido em 25/09/1989, que está presente em cinco das amostras da Tabela 4.19. O maior pico observado (em 07/06/1990) se apresentou nas Amostras 3 a 6, enquanto a menor vazão (em 21/12/1983) se destacou nas Amostras 1 a 4. Ou seja, ambos os eventos ocorreram simultaneamente nas Amostras 3 e 4.

A Tabela 4.20 indica as frequências absolutas e relativas dos eventos da Tabela 4.19 – estratégia de amostragem do Grupo 2 – que estão presentes na Tabela 4.18 – ranqueamento de vazões do Grupo 2. Com base na Tabela 4.20, tem-se que as Amostras 2 e 5 apresentaram a menor quantidade de eventos na fase de treinamento; a Amostra 2, no entanto, ao contrário da 5, apresentou a maior incidência de eventos ranqueados entre as maiores e menores vazões de pico em União da Vitória (Tabela 4.18), indicando quase 40% das vazões nos dois extremos; as Amostras 1 e 3 apresentaram praticamente a mesma incidência de eventos ranqueados que a Amostra 2, porém numa gama de eventos maior, indicando, portanto, em torno de 30% das vazões nos dois extremos. As Amostras 5 a 7 tiveram as menores incidências de eventos ranqueados entre as maiores e menores vazões de pico em União da Vitória, indicando menos de 20% das vazões nos dois extremos.

TABELA 4.19 – ESTRATÉGIA DE AMOSTRAGEM DOS EVENTOS DO GRUPO 2 – FASE DE TREINAMENTO

			Picos dos even	tos do Grupo 2	,		
Amostra 1	Amostra 2	Amostra 3	Amostra 4	Amostra 5	Amostra 6	Amostra 7	Amostra 8
31/08/1964	25/11/1969	25/03/1974	10/10/1979	29/12/1986	31/08/1964	31/08/1964	31/08/1964
03/10/1965	22/10/1970	13/09/1975	28/01/1980	23/02/1987	03/10/1965	03/10/1965	03/10/1965
17/11/1965	28/03/1971	13/10/1975	02/03/1980	04/02/1989	17/11/1965	17/11/1965	17/11/1965
22/12/1965	10/06/1971	15/12/1975	14/07/1980	07/04/1989	22/12/1965	22/12/1965	22/12/1965
23/02/1966	09/12/1972	28/03/1976	07/08/1980	25/09/1989	23/02/1966	23/02/1966	23/02/1966
01/07/1966	28/06/1973	14/04/1976	06/09/1980	09/04/1990	01/07/1966	01/07/1966	01/07/1966
23/12/1966	24/07/1973	17/09/1976	28/02/1982	02/05/1990	23/12/1966	23/12/1966	23/12/1966
06/03/1967	31/08/1973	08/11/1976	20/07/1982	07/06/1990	06/03/1967	06/03/1967	06/03/1967
03/04/1967	25/03/1974	14/02/1977	02/02/1983	01/08/1990	03/04/1967	03/04/1967	03/04/1967
05/12/1967	13/09/1975	19/10/1977	06/03/1983	08/09/1990	05/12/1967	05/12/1967	05/12/1967
25/11/1969	13/10/1975	11/12/1977	03/10/1983	28/12/1991	04/02/1989	25/11/1969	25/11/1969
22/10/1970	15/12/1975	10/10/1979	24/12/1983	31/03/1992	07/04/1989	22/10/1970	22/10/1970
28/03/1971	28/03/1976	28/01/1980	29/09/1984	03/07/1992	25/09/1989	28/03/1971	28/03/1971
10/06/1971	14/04/1976	02/03/1980	21/04/1985	26/02/1993	09/04/1990	10/06/1971	10/06/1971
09/12/1972	17/09/1976	14/07/1980	29/12/1986	22/07/1993	02/05/1990	09/12/1972	09/12/1972
28/06/1973	08/11/1976	07/08/1980	23/02/1987	26/02/1994	07/06/1990	28/06/1973	28/06/1973
24/07/1973	14/02/1977	06/09/1980	04/02/1989	08/07/1994	01/08/1990	24/07/1973	24/07/1973
31/08/1973	19/10/1977	28/02/1982	07/04/1989	21/11/1994	08/09/1990	31/08/1973	31/08/1973
25/03/1974	11/12/1977	20/07/1982	25/09/1989	02/02/1996	28/12/1991	26/02/1994	25/03/1974
13/09/1975	10/10/1979	02/02/1983	09/04/1990	16/08/1996	31/03/1992	08/07/1994	13/09/1975
13/10/1975	28/01/1980	06/03/1983	02/05/1990	10/09/1996	03/07/1992	21/11/1994	13/10/1975
15/12/1975	02/03/1980	03/10/1983	07/06/1990	21/10/1996	26/02/1993	02/02/1996	15/12/1975
28/03/1976	14/07/1980	24/12/1983	01/08/1990	05/08/1997	22/07/1993	16/08/1996	28/03/1976
14/04/1976	07/08/1980	29/09/1984	08/09/1990	27/08/1997	26/02/1994	10/09/1996	14/04/1976
17/09/1976	06/09/1980	21/04/1985	28/12/1991	10/07/1998	08/07/1994	21/10/1996	17/09/1976
08/11/1976	28/02/1982	29/12/1986	31/03/1992	12/12/1998	21/11/1994	05/08/1997	08/11/1976
14/02/1977	20/07/1982	23/02/1987	03/07/1992	03/03/1999	02/02/1996	27/08/1997	14/02/1977
19/10/1977	02/02/1983	04/02/1989	26/02/1993	20/04/1999	16/08/1996	10/07/1998	19/10/1977
11/12/1977	06/03/1983	07/04/1989	22/07/1993	18/10/1999	10/09/1996	12/12/1998	11/12/1977
10/10/1979	03/10/1983	25/09/1989	26/02/1994	01/08/1990	21/10/1996	03/03/1999	03/03/1999
28/01/1980	24/12/1983	09/04/1990	08/07/1994	31/12/2000	05/08/1997	20/04/1999	20/04/1999
02/03/1980	29/09/1984	02/05/1990	21/11/1994	22/02/2001	27/08/1997	18/10/1999	18/10/1999
14/07/1980	21/04/1985	07/06/1990	02/02/1996	29/05/2001	10/07/1998	31/12/2000	31/12/2000
07/08/1980	29/12/1986	01/08/1990	16/08/1996	27/06/2001	12/12/1998	22/02/2001	22/02/2001
06/09/1980	23/02/1987	08/09/1990	10/09/1996	06/08/2001	03/03/1999	29/05/2001	29/05/2001
28/02/1982	04/02/1989	28/12/1991	21/10/1996	01/12/2001	20/04/1999	27/06/2001	27/06/2001
20/07/1982	07/04/1989	31/03/1992	05/08/1997	07/12/2002	31/12/2000	06/08/2001	06/08/2001
02/02/1983	25/09/1989	03/07/1992	27/08/1997	16/06/2004	22/02/2001	01/12/2001	01/12/2001
06/03/1983		26/02/1993	10/07/1998	22/07/2004	29/05/2001	07/12/2002	07/12/2002
03/10/1983	ı	22/07/1993	12/12/1998		27/06/2001	16/06/2004	16/06/2004
24/12/1983		26/02/1994	03/03/1999		06/08/2001	22/07/2004	22/07/2004
29/09/1984		08/07/1994	20/04/1999		01/12/2001		
		21/11/1994	18/10/1999	-	07/12/2002		
					16/06/2004		
					22/07/2004		

TABELA 4.20 – FREQUÊNCIAS ABSOLUTAS (À ESQUERDA) E RELATIVAS (À DIREITA) DOS MAIORES E MENORES PICOS DO GRUPO 2

		Amostras									
	1	2	3	4	5	6	7	8			
Total de eventos	42	38	43	43	38	46	41	41			
Maiores	7 16,7%	8 21,1%	8 18,6%	6 14,0%	3 07,9%	3 06,5%	2 04,9%	4 09,8%			
Menores	6 14,3%	6 15,8%	6 14,0%	5 11,6%	4 10,5%	4 08,7%	4 09,8%	5 12,2%			

Para o Modelo 5 – Modelo de Eventos do Grupo 3 – foi apresentado na Tabela 4.21 o ranqueamento de todos os eventos do Grupo 3 até 2004, em ordem decrescente de vazões de pico; a quantidade de eventos pertencentes a este grupo permitiu a avaliação completa. Para melhor entendimento do comportamento das métricas das Amostras, apresenta-se também a Tabela 4.22 com a estratégia de amostragem dos eventos do Grupo 3 desta etapa; esta Tabela indica quais eventos deste Grupo foram utilizados em cada amostra. As cores destacadas em ambas as Tabelas indicam o aparecimento do evento da Tabela 4.21 nas Amostras da Tabela 4.22.

TABELA 4.21 – RANQUEAMENTO DAS VAZÕES DE PICO DO GRUPO 3, EM UNIÃO DA VITÓRIA

Ord.	Data de pico	Vazão de pico (m³/s)	Ord.	Data de pico	Vazão de pico (m³/s)	Ord.	Data de pico	Vazão de pico (m³/s)
1 ^a	18/07/1983	5156,7	6 ^a	02/05/1998	2447,2	11 ^a	28/05/1987	2167,4
2ª	08/06/1992	3953,6	7 ^a	13/01/1971	2428,5	12 ^a	11/10/2001	1975,6
3ª	06/10/1993	2786,2	8 ^a	25/09/2000	2284	13 ^a	01/01/1981	1887,1
4 ^a	12/10/1998	2751,7	9 ^a	27/11/1982	2263	14 ^a	14/07/1999	1832,9
5 ^a	22/01/1995	2453,3	10 ^a	27/01/1990	2196	15 ^a	04/02/1997	1808,7

TABELA 4.22 – ESTRATÉGIA DE AMOSTRAGEM DOS EVENTOS DO GRUPO 3 – FASE DE TREINAMENTO

	Picos dos eventos do Grupo 3										
Amostra 1	Amostra 2	Amostra 3	Amostra 4	Amostra 5	Amostra 6	Amostra 7	Amostra 8				
13/01/1971	13/01/1971	01/01/1981	01/01/1981	28/05/1987	27/01/1990	13/01/1971	13/01/1971				
01/01/1981	01/01/1981	27/11/1982	27/11/1982	27/01/1990	08/06/1992	22/01/1995	14/07/1999				
27/11/1982	27/11/1982	18/07/1983	18/07/1983	08/06/1992	06/10/1993	04/02/1997	25/09/2000				
18/07/1983	18/07/1983	28/05/1987	28/05/1987	06/10/1993	22/01/1995	02/05/1998	11/10/2001				
	28/05/1987	27/01/1990	27/01/1990	22/01/1995	04/02/1997	12/10/1998					
		08/06/1992	08/06/1992	04/02/1997	02/05/1998	14/07/1999					
		06/10/1993	06/10/1993	02/05/1998	12/10/1998	25/09/2000	-				
			22/01/1995	12/10/1998	14/07/1999	11/10/2001					
			04/02/1997	14/07/1999	25/09/2000		-				
			02/05/1998	25/09/2000	11/10/2001						
			12/10/1998	11/10/2001		•					
			14/07/1999		-						

De maneira geral, cada evento destacado na Tabela 4.21 – ranqueamento de vazões do Grupo 3 – aparece quatro vezes na Tabela 4.22 – estratégia de amostragem do Grupo 3 –, em amostras diferentes, exceto o evento com pico ocorrido em 14/07/1999, que está destacado em cinco amostras na Tabela 4.22. O maior pico observado (ocorrido em 18/07/1983) apresentouse nas Amostras 1 a 4, enquanto o menor (ocorrido em 04/02/1997), nas Amostras 4 a 7. Ou seja, ambos os eventos ocorreram simultaneamente apenas na Amostra 4, o que indica que esta Amostra obteve maior amplitude de vazões. A presença – ou ausência – destes eventos nas referidas amostras justificam o comportamento das métricas apresentado nos *boxplot* (Apêndice C – Figura C.5). Por outro lado, a Amostra 8 apresenta vazões mais baixas e apresentou maior variabilidade, quando comparada às demais.

4.2.2 Teste

Considerando o comportamento dos coeficientes da Figura 4.7, os conjuntos de parâmetros cujos valores se mantiveram entre o primeiro e o terceiro quartis na *etapa de treinamento* foram selecionados, enquanto as médias de seus índices de performance, na *etapa de teste*, foram apresentados na Figura 4.8. No Apêndice D foram apresentados os *boxplots* com os resultados dos 32 conjuntos de parâmetros dos modelos, para cada amostra da estratégia de validação cruzada e para sua média. Observa-se a partir das Figuras do Apêndice D que, além da diminuição da quantidade de *outliers* quando comparados aos *boxplots* da fase de treinamento (Apêndice C), em geral, os modelos por grupo apresentaram desempenho inferior à etapa de treinamento e maior variabilidade de resultados entre as Amostras 4 a 8.

Dentre estes resultados, não necessariamente foram observadas as mesmas 32 combinações de parâmetros para as diferentes amostras de um determinado modelo. Por isso, os conjuntos de parâmetros comuns a todas as amostras de um mesmo modelo foram coletados e seus resultados, por amostra, apresentados no Apêndice E, enquanto as médias das amostras se apresentam na Figura 4.8. As médias das amostras deram origem às Tabelas 4.23 a 4.26, que compõem as estatísticas dos conjuntos de parâmetros.

Nas Tabelas 4.23 a 4.26, observa-se nos Modelos 3 e 5 a maior amplitude interquartis e maior desvio-padrão, para praticamente todas as métricas, bem como, as piores médias e medianas, apesar de seus resultados adequados, conforme as Tabelas 4.23 a 4.26; o Modelo 4 apresentou os resultados mais estáveis.

FIGURA 4.8 – RESULTADOS DE KG, NS, NS MODIFICADO E MARE PARA DIFERENTES CONJUNTOS DE PARÂMETROS E MODELOS – MÉDIA ENTRE AS AMOSTRAS – ETAPA DE TESTE

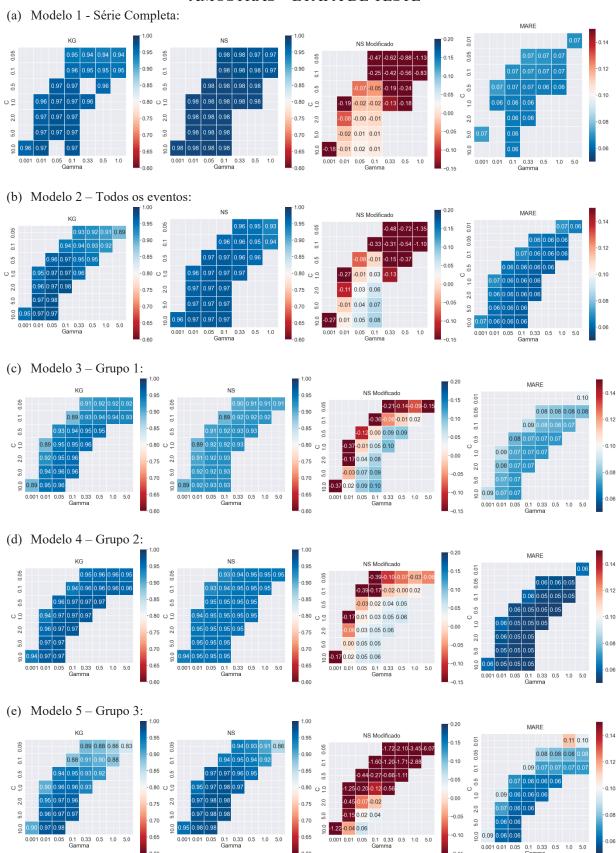


TABELA 4.23 – AVALIAÇÃO MÉDIA DO COMPORTAMENTO DE KG, PARA DIFERENTES CONJUNTOS DE PARÂMETROS – FASE DE TESTE

		Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
so	Quantidade	32	32	32	32	32
conjuntos tros	Média	0,960	0,947	0,924	0,956	0,905
conj	Desvio-padrão	0,012	0,026	0,040	0,019	0,063
	Mínimo	0,927	0,889	0,833	0,909	0,735
dos râm	25%	0,952	0,925	0,916	0,954	0,881
cas	50%	0,962	0,952	0,938	0,964	0,913
ísti de	75%	0,971	0,973	0,954	0,969	0,959
Estatísticas dos de parâmo	Máximo	0,975	0,977	0,958	0,971	0,975
Es	Amplitude interquartis	0,019	0,047	0,037	0,015	0,078

TABELA 4.24 – AVALIAÇÃO MÉDIA DO COMPORTAMENTO DE NS PARA DIFERENTES CONJUNTOS DE PARÂMETROS – FASE DE TESTE

	•	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
soz	Quantidade	32	32	32	32	32
untos	Média	0,979	0,959	0,911	0,946	0,949
conj	Desvio-padrão	0,008	0,015	0,019	0,008	0,033
	Mínimo	0,945	0,896	0,865	0,927	0,855
dos râmo	25%	0,978	0,954	0,906	0,945	0,940
cas pa	50%	0,981	0,963	0,919	0,950	0,960
ísti de	75%	0,984	0,968	0,925	0,952	0,975
Estatísticas de paı	Máximo	0,984	0,971	0,927	0,952	0,979
Es	Amplitude interquartis	0,006	0,015	0,020	0,006	0,034

TABELA 4.25 – AVALIAÇÃO MÉDIA DO COMPORTAMENTO DE NS MODIFICADO PARA DIFERENTES CONJUNTOS DE PARÂMETROS – FASE DE TESTE

	•	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
sos	Quantidade	32	32	32	32	32
untos	Média	-0,327	-0,360	-0,119	-0,061	-1,385
conj	Desvio-padrão	0,526	0,566	0,261	0,154	1,590
	Mínimo	-2,729	-2,806	-0,757	-0,470	-6,068
dos	25%	-0,403	-0,542	-0,185	-0,077	-1,816
cas	50%	-0,183	-0,210	-0,012	0,007	-0,850
ísti de	75%	-0,024	0,000	0,080	0,046	-0,139
Estatísti de	Máximo	0,017	0,077	0,106	0,058	0,061
ES	Amplitude interquartis	0,379	0,542	0,265	0,122	1,678

TABELA 4.26 – AVALIAÇÃO MÉDIA DO COMPORTAMENTO DE MARE PARA DIFERENTES CONJUNTOS DE PARÂMETROS – FASE DE TESTE

		Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
SOS	Quantidade	32	32	32	32	32
Ē	Média	0,066	0,062	0,078	0,056	0,075
conjuntos etros	Desvio-padrão	0,003	0,005	0,010	0,004	0,022
	Mínimo	0,063	0,058	0,069	0,053	0,057
dos râme	25%	0,064	0,059	0,071	0,053	0,060
cas	50%	0,065	0,060	0,075	0,054	0,069
ísti de	75%	0,068	0,064	0,081	0,056	0,080
Estatísticas de paı	Máximo	0,073	0,076	0,102	0,066	0,139
$\mathbf{E}\mathbf{s}$	Amplitude interquartis	0,004	0,005	0,010	0,003	0,019

Nos *testes* dos modelos por grupo também se torna fundamental a identificação dos eventos pertencentes a cada amostra das classes; o ranqueamento de eventos do Grupo 1 foram apresentadas na Tabela 4.15 (treinamento); a Tabela 4.27 é apresentada para facilitar a interpretação do comportamento das métricas das amostras, indicando quais eventos do Grupo 1 foram considerados em cada amostra durante o teste. As cores destacadas nas Tabelas 4.15 – ranqueamento das vazões do Grupo 1 – e 4.27 – estratégia de amostragem do Grupo 1 – indicam o aparecimento do evento em ambas as Tabelas.

TABELA 4.27 – ESTRATÉGIA DE AMOSTRAGEM DOS EVENTOS DO GRUPO 1 – FASE DE TESTE

		Pic	os dos eventos o	do Grupo 1 – T	este		
Amostra 1	Amostra 2	Amostra 3	Amostra 4	Amostra 5	Amostra 6	Amostra 7	Amostra 8
21/02/1985	07/05/1964	07/05/1964	07/05/1964	07/05/1964	08/04/1969	28/01/1974	16/05/1979
25/05/1988	28/06/1964	28/06/1964	28/06/1964	28/06/1964	22/06/1969	26/07/1974	11/12/1981
27/10/1988	16/05/1965	16/05/1965	16/05/1965	16/05/1965	11/10/1969	05/09/1974	19/05/1984
15/05/1989	15/07/1965	15/07/1965	15/07/1965	15/07/1965	08/07/1970	12/08/1975	18/06/1984
05/08/1989	24/08/1965	24/08/1965	24/08/1965	24/08/1965	15/10/1971	14/06/1976	09/08/1984
24/06/1991	07/09/1966	07/09/1966	07/09/1966	07/09/1966	29/02/1972	20/08/1976	13/11/1984
12/10/1991	30/10/1966	30/10/1966	30/10/1966	30/10/1966	30/08/1972	25/07/1978	21/02/1985
15/05/1993	27/01/1968	27/01/1968	27/01/1968	27/01/1968	23/04/1973	16/05/1979	25/05/1988
13/05/1994	02/11/1968	02/11/1968	02/11/1968	02/11/1968	14/08/1973	11/12/1981	27/10/1988
18/07/1995	24/06/1991	08/04/1969	08/04/1969	08/04/1969	28/01/1974	19/05/1984	15/05/1989
29/09/1995	12/10/1991	22/06/1969	22/06/1969	22/06/1969	26/07/1974	18/06/1984	05/08/1989
12/07/1996	15/05/1993	11/10/1969	11/10/1969	11/10/1969	05/09/1974	09/08/1984	24/06/1991
21/06/1997	13/05/1994	08/07/1970	08/07/1970	08/07/1970	12/08/1975	13/11/1984	12/10/1991
02/03/2000	18/07/1995	15/10/1971	15/10/1971	15/10/1971	14/06/1976	21/02/1985	15/05/1993
01/09/2001	29/09/1995	29/02/1972	29/02/1972	29/02/1972	20/08/1976	25/05/1988	13/05/1994
29/01/2002	12/07/1996	30/08/1972	30/08/1972	30/08/1972	25/07/1978	27/10/1988	18/07/1995
07/08/2002	21/06/1997	23/04/1973	23/04/1973	23/04/1973	16/05/1979	15/05/1989	29/09/1995
02/10/2002	02/03/2000	14/08/1973	14/08/1973	14/08/1973	11/12/1981	05/08/1989	12/07/1996
01/01/2004	01/09/2001	18/07/1995	28/01/1974	28/01/1974	19/05/1984	24/06/1991	21/06/1997
27/10/2004	29/01/2002	29/09/1995	26/07/1974	26/07/1974	18/06/1984	12/10/1991	
	07/08/2002	12/07/1996	05/09/1974	05/09/1974	09/08/1984	15/05/1993	
	02/10/2002	21/06/1997	12/08/1975	12/08/1975	13/11/1984		•
	01/01/2004	02/03/2000	14/06/1976	14/06/1976	21/02/1985		
	27/10/2004	01/09/2001	20/08/1976	20/08/1976	25/05/1988		
		29/01/2002	25/07/1978	25/07/1978	27/10/1988		
		07/08/2002	02/03/2000	16/05/1979		•	
		02/10/2002	01/09/2001	11/12/1981			
		01/01/2004	29/01/2002	19/05/1984			
		27/10/2004	07/08/2002	18/06/1984			
			02/10/2002	09/08/1984			
			01/01/2004	13/11/1984	•		
			27/10/2004	21/02/1985			

Em oposição ao observado no Modelo 3 na fase de treinamento (Tabela 4.16), na fase de teste, por meio da Tabela 4.27, o maior pico observado (09/08/1984) se apresentou nas

Amostras 5 a 8, enquanto o menor (01/09/2001), nas Amostras 1 a 4; ou seja, os extremos máximo e mínimo não pertenceram simultaneamente à mesma amostra.

Observando as Tabelas 4.28 (teste) e 4.17 (treinamento), percebe-se que as magnitudes das vazões em cada amostra do teste e a quantidade de eventos que as compõem definiram o comportamento das métricas em cada amostra, mas de maneira antagônica àquela observada no treinamento: as Amostras 3, 4 e 5 apresentaram maior quantidade de eventos na fase de teste e as maiores incidências de eventos ranqueados (Tabela 4.15); logo, para as Amostras 3 a 5, as vazões extremas – máximas e mínimas – e as intermediárias tiveram quantidades equivalentes à, aproximadamente, 50% cada, que se distinguem da divisão de vazões das demais amostras, que majoritariamente foram contempladas com mais de 50% de vazões intermediárias.

TABELA 4.28 – FREQUÊNCIAS ABSOLUTAS (À ESQUERDA) E RELATIVAS (À DIREITA) DOS MAIORES E MENORES PICOS DO GRUPO 1 – FASE DE TESTE

		Amostras							
	1	2	3	4	5	6	7	8	
Total de eventos	20	24	29	32	32	25	21	19	
Maiores	3 15,0%	4 16,7%	6 20,7%	6 18,8%	7 21,9%	6 24,0%	4 19,0%	4 21,1%	
Menores	3 15,0%	6 25,0%	9 31,0%	9 31,0%	8 25,0%	4 16,0%	1 04,8%	1 05,3%	

Para o Grupo 2, as 10 maiores e as 10 menores vazões de pico foram apresentadas na Tabela 4.18 (treinamento); a Tabela 4.29 indica quais eventos deste Grupo foram utilizados em cada amostra na fase de teste. As cores destacadas em ambas as Tabelas indicam o aparecimento do evento da Tabela 4.18 – ranqueamento de vazões do Grupo 2 – nas Amostras da Tabela 4.29. Em oposição ao observado na fase de treinamento do Modelo 4 (Tabela 4.19), na fase de teste o maior pico observado (07/06/1990) apresentou-se nas Amostras 1, 2, 7 e 8, enquanto o menor (24/1/1983), nas Amostras 4 a 8; ou seja, os extremos máximo e mínimo pertenceram simultaneamente às Amostras 7 e 8.

A Tabela 4.30 indica as frequências absolutas e relativas dos eventos da Tabela 4.29 – estratégia de validação cruzada do Grupo 2 – que estão presentes na Tabela 4.18 – ranqueamento de vazões do Grupo 2. Com base na Tabela 4.30, tem-se que as Amostras 2 e 5 apresentaram a maior quantidade de eventos na fase de teste – ao contrário do ocorrido na etapa de treinamento; a Amostra 5, em oposição à 2, apresentou a maior incidência de eventos ranqueados entre as maiores e menores vazões de pico em União da Vitória (Tabela 4.18), indicando quase 30% das vazões nos dois extremos.

TABELA 4.29 – ESTRATÉGIA DE AMOSTRAGEM DOS EVENTOS DO GRUPO 2 – FASE DE TESTE

			Picos dos even	tos do Grupo 2	,		
Amostra 1	Amostra 2	Amostra 3	Amostra 4	Amostra 5	Amostra 6	Amostra 7	Amostra 8
21/04/1985	31/08/1964	31/08/1964	31/08/1964	31/08/1964	25/11/1969	25/03/1974	10/10/1979
29/12/1986	03/10/1965	03/10/1965	03/10/1965	03/10/1965	22/10/1970	13/09/1975	28/01/1980
23/02/1987	17/11/1965	17/11/1965	17/11/1965	17/11/1965	28/03/1971	13/10/1975	02/03/1980
04/02/1989	22/12/1965	22/12/1965	22/12/1965	22/12/1965	10/06/1971	15/12/1975	14/07/1980
07/04/1989	23/02/1966	23/02/1966	23/02/1966	23/02/1966	09/12/1972	28/03/1976	07/08/1980
25/09/1989	01/07/1966	01/07/1966	01/07/1966	01/07/1966	28/06/1973	14/04/1976	06/09/1980
09/04/1990	23/12/1966	23/12/1966	23/12/1966	23/12/1966	24/07/1973	17/09/1976	28/02/1982
02/05/1990	06/03/1967	06/03/1967	06/03/1967	06/03/1967	31/08/1973	08/11/1976	20/07/1982
07/06/1990	03/04/1967	03/04/1967	03/04/1967	03/04/1967	25/03/1974	14/02/1977	02/02/1983
01/08/1990	05/12/1967	05/12/1967	05/12/1967	05/12/1967	13/09/1975	19/10/1977	06/03/1983
08/09/1990	09/04/1990	25/11/1969	25/11/1969	25/11/1969	13/10/1975	11/12/1977	03/10/1983
28/12/1991	02/05/1990	22/10/1970	22/10/1970	22/10/1970	15/12/1975	10/10/1979	24/12/1983
31/03/1992	07/06/1990	28/03/1971	28/03/1971	28/03/1971	28/03/1976	28/01/1980	29/09/1984
03/07/1992	01/08/1990	10/06/1971	10/06/1971	10/06/1971	14/04/1976	02/03/1980	21/04/1985
26/02/1993	08/09/1990	09/12/1972	09/12/1972	09/12/1972	17/09/1976	14/07/1980	29/12/1986
22/07/1993	28/12/1991	28/06/1973	28/06/1973	28/06/1973	08/11/1976	07/08/1980	23/02/1987
26/02/1994	31/03/1992	24/07/1973	24/07/1973	24/07/1973	14/02/1977	06/09/1980	04/02/1989
08/07/1994	03/07/1992	31/08/1973	31/08/1973	31/08/1973	19/10/1977	28/02/1982	07/04/1989
21/11/1994	26/02/1993	02/02/1996	25/03/1974	25/03/1974	11/12/1977	20/07/1982	25/09/1989
02/02/1996	22/07/1993	16/08/1996	13/09/1975	13/09/1975	10/10/1979	02/02/1983	09/04/1990
16/08/1996	26/02/1994	10/09/1996	13/10/1975	13/10/1975	28/01/1980	06/03/1983	02/05/1990
10/09/1996	08/07/1994	21/10/1996	15/12/1975	15/12/1975	02/03/1980	03/10/1983	07/06/1990
21/10/1996	21/11/1994	05/08/1997	28/03/1976	28/03/1976	14/07/1980	24/12/1983	01/08/1990
05/08/1997	02/02/1996	27/08/1997	14/04/1976	14/04/1976	07/08/1980	29/09/1984	08/09/1990
27/08/1997	16/08/1996	10/07/1998	17/09/1976	17/09/1976	06/09/1980	21/04/1985	28/12/1991
10/07/1998	10/09/1996	12/12/1998	08/11/1976	08/11/1976	28/02/1982	29/12/1986	31/03/1992
12/12/1998	21/10/1996	03/03/1999	14/02/1977	14/02/1977	20/07/1982	23/02/1987	03/07/1992
03/03/1999	05/08/1997	20/04/1999	19/10/1977	19/10/1977	02/02/1983	04/02/1989	26/02/1993
20/04/1999	27/08/1997	18/10/1999	11/12/1977	11/12/1977	06/03/1983	07/04/1989	22/07/1993
18/10/1999	10/07/1998	31/12/2000	31/12/2000	10/10/1979	03/10/1983	25/09/1989	26/02/1994
31/12/2000	12/12/1998	22/02/2001	22/02/2001	28/01/1980	24/12/1983	09/04/1990	08/07/1994
22/02/2001	03/03/1999	29/05/2001	29/05/2001	02/03/1980	29/09/1984	02/05/1990	21/11/1994
29/05/2001	20/04/1999	27/06/2001	27/06/2001	14/07/1980	21/04/1985	07/06/1990	02/02/1996
27/06/2001	18/10/1999	06/08/2001	06/08/2001	07/08/1980	29/12/1986	01/08/1990	16/08/1996
06/08/2001	31/12/2000	01/12/2001	01/12/2001	06/09/1980	23/02/1987	08/09/1990	10/09/1996
01/12/2001	22/02/2001	07/12/2002	07/12/2002	28/02/1982		28/12/1991	21/10/1996
07/12/2002	29/05/2001	16/06/2004	16/06/2004	20/07/1982		31/03/1992	05/08/1997
16/06/2004	27/06/2001	22/07/2004	22/07/2004	02/02/1983	•	03/07/1992	27/08/1997
22/07/2004	06/08/2001			06/03/1983		26/02/1993	10/07/1998
	01/12/2001			03/10/1983		22/07/1993	12/12/1998
	07/12/2002			24/12/1983	I		
	16/06/2004			29/09/1984	•		
	22/07/2004			21/04/1985			

TABELA 4.30 – FREQUÊNCIAS ABSOLUTAS (À ESQUERDA) E RELATIVAS (À DIREITA) DOS MAIORES E MENORES PICOS DO GRUPO 2 – FASE DE TESTE

		Amostras								
	1	2	3	4	5	6	7	8		
Total de eventos	39	43	38	38	43	35	40	40		
Maiores	3 07,7 %	2 04,7 %	2 05,3 %	4 10,5 %	7 16,3 %	7 20,0 %	8 20,0 %	6 15,0 %		
Menores	4 10,3 %	4 09,3 %	4 10,5 %	5 13,2%	6 14,0 %	6 17,1 %	6 15,0 %	5 12,5 %		

Já para o Grupo 3, o ranqueamento das vazões de pico do Grupo 3 foi apresentado na Tabela 4.21, na etapa de treinamento; a Tabela 4.31 indica quais eventos deste Grupo foram utilizados em cada amostra da fase de teste. As cores destacadas em ambas as Tabelas indicam o aparecimento do evento da Tabela 4.21 – ranqueamento de vazões do Grupo 3 – nas Amostras da Tabela 4.31.

TABELA 4.31 – ESTRATÉGIA DE AMOSTRAGEM DOS EVENTOS DO GRUPO 3 – FASE DE TESTE

			Picos dos even	tos do Grupo 3	l .		
Amostra 1	Amostra 2	Amostra 3	Amostra 4	Amostra 5	Amostra 6	Amostra 7	Amostra 8
28/05/1987	27/01/1990	13/01/1971	13/01/1971	13/01/1971	13/01/1971	01/01/1981	01/01/1981
27/01/1990	08/06/1992	22/01/1995	25/09/2000	01/01/1981	01/01/1981	27/11/1982	27/11/1982
08/06/1992	06/10/1993	04/02/1997	11/10/2001	27/11/1982	27/11/1982	18/07/1983	18/07/1983
06/10/1993	22/01/1995	02/05/1998		18/07/1983	18/07/1983	28/05/1987	28/05/1987
22/01/1995	04/02/1997	12/10/1998			28/05/1987	27/01/1990	27/01/1990
04/02/1997	02/05/1998	14/07/1999				08/06/1992	08/06/1992
02/05/1998	12/10/1998	25/09/2000				06/10/1993	06/10/1993
12/10/1998	14/07/1999	11/10/2001					22/01/1995
14/07/1999	25/09/2000		•				04/02/1997
25/09/2000	11/10/2001						02/05/1998
11/10/2001		•					12/10/1998

Em oposição ao observado na fase de treinamento do Modelo 5 (Tabela 4.22), na fase de teste o maior pico observado (em 18/07/1983) apresentou-se nas Amostras 5 a 8, enquanto o menor (em 04/02/1997), nas Amostras 1, 2, 3 e 8; ou seja, os extremos máximo e mínimo pertenceram simultaneamente apenas à Amostra 8, e, na fase de treinamento, as amostras que continham algum dos dois extremos se inverteram.

4.2.3 Ajuste final dos modelos

O conjunto de parâmetros com C=0.5 e $\gamma=0.33$ se apresentou no centro dos mapas de calor da Figura 4.8 e pode ser observado nas cinco categorias de modelo; este conjunto foi utilizado para o ajuste final de todos os modelos. As medidas de avaliação dos modelos são

apresentadas na Tabela 4.32; as diagonais principais da Tabela – em negrito – indicam os resultados obtidos na simulação da mesma série utilizada na calibração do modelo; os resultados sublinhados representam os modelos com métricas inferiores a 0,90 para KG e NS, superiores a 0,10 para MARE, e os piores resultados para NS Modificado.

TABELA 4.32 – COEFICIENTES KG, NS, NS MODIFICADO E MARE DAS SIMULAÇÕES DAS SÉRIES – 1964 A 2004

			Se	éries simulada	as	
		Completa	Todos os eventos	Grupo 1	Grupo 2	Grupo 3
	Modelo 1 – Série Completa	0,975	0,984	0,949	0,971	0,982
	Modelo 2 – Todos os Eventos	0,954	0,978	0,934	0,975	0,983
KG	Modelo 3 – Grupo 1	0,585	0,817	0,960	0,930	0,882
	Modelo 4 – Grupo 2	0,885	0,949	0,958	0,972	0,962
	Modelo 5 – Grupo 3	0,948	0,977	0,933	0,973	0,980
	Modelo 1 – Série Completa	0,985	0,974	0,923	0,952	0,982
	Modelo 2 – Todos os Eventos	0,983	0,973	0,919	0,951	0,982
NS	Modelo 3 – Grupo 1	0,766	0,905	0,931	0,953	0,963
	Modelo 4 – Grupo 2	0,968	0,971	0,930	0,953	0,979
	Modelo 5 – Grupo 3	0,982	0,973	0,917	0,950	0,981
	Modelo 1 – Série Completa	0,034	0,092	-0,028	0,003	0,148
NIC	Modelo 2 – Todos os Eventos	-0,088	0,103	-0,087	-0,004	0,181
NS Mod.	Modelo 3 – Grupo 1	<u>-14,095</u>	<u>-2,262</u>	0,117	0,021	<u>-0,731</u>
Mou.	Modelo 4 – Grupo 2	-1,041	0,007	0,053	0,059	0,008
	Modelo 5 – Grupo 3	-0,187	0,071	<u>-0,119</u>	<u>-0,030</u>	0,134
	Modelo 1 – Série Completa	0,063	0,062	0,084	0,058	0,053
	Modelo 2 – Todos os Eventos	0,139	0,058	0,088	0,059	0,055
MARE	Modelo 3 – Grupo 1	0,896	0,205	0,069	0,067	0,144
	Modelo 4 – Grupo 2	0,275	0,075	0,070	0,052	0,064
	Modelo 5 – Grupo 3	0,150	0,066	0,088	0,060	0,053

Esperava-se que cada série, ao ser simulada por seu próprio modelo – treinado com a mesma categoria de série – apresentasse melhores resultados quando comparados àqueles da simulação realizada por outro modelo – valores fora da diagonal. Apesar de, nem sempre, os valores da diagonal principal terem retornado as melhores métricas, os resultados representam um desempenho equivalente.

Observando-se a Tabela 4.32, portanto, tem-se que, para todas as métricas: (i) a Série Completa foi melhor simulada pelos Modelos 1 (seu próprio modelo), 2 e 5; (ii) a série de Todos os Eventos foi bem estimada por todos os modelos, exceto pelo Modelo 3; (iii) todos os modelos bem representaram a série de eventos do Grupo 1, assim como, (iv) a série de eventos do Grupo 2; e (v) a série de eventos do Grupo 3 foi estimada adequadamente por todos os modelos, porém, mais uma vez, o Modelo 3 apresentou desempenho inferior. Portanto, o Modelo 3, para todas

as métricas, apresentou a *pior* performance na simulação de todas as séries, mas com *bons* resultados (KG e NS acima de 0,90) na simulação das séries dos Grupos 1 e 2.

O comportamento dos modelos também pode ser observado estatisticamente, por meio dos erros relativos dos parâmetros, conforme apresenta a Tabela 4.33, na qual os valores em negrito indicam as simulações das séries pelos modelos de mesma categoria.

TABELA 4.33 – ERROS RELATIVOS ENTRE OS PARÂMETROS ESTATÍSTICOS DAS SÉRIES SIMULADAS E OBSERVADAS – 1964 A 2004

			Se	éries simuladas	 	
		Série Completa	Todos os eventos	Grupo 1	Grupo 2	Grupo 3
	Modelo 1 – Série Completa	19,02%	-9,59%	-14,75%	5,44%	6,68%
na	Modelo 2 – Todos os Eventos	115,88%	7,35%	-5,63%	7,40%	36,00%
Mínima	Modelo 3 – Grupo 1	725,67%	203,27%	48,68%	34,29%	220,02%
$\mathbf{\bar{Z}}$	Modelo 4 – Grupo 2	251,93%	59,52%	8,50%	13,58%	77,05%
	Modelo 5 – Grupo 3	131,48%	25,22%	-2,00%	10,28%	34,79%
	Modelo 1 – Série Completa	-0,40%	-0,41%	-5,00%	-6,00%	-0,44%
ma	Modelo 2 – Todos os Eventos	-0,95%	-0,95%	-5,06%	-5,83%	-0,97%
Máxima	Modelo 3 – Grupo 1	-4,58%	-4,50%	-3,45%	-3,79%	-4,47%
Ä	Modelo 4 – Grupo 2	-1,07%	-1,05%	-3,69%	-3,68%	-1,05%
	Modelo 5 – Grupo 3	-1,72%	-1,69%	-4,89%	-5,94%	-1,70%
æ	Modelo 1 – Série Completa	-1,58%	-0,39%	0,40%	-0,52%	-1,00%
	Modelo 2 – Todos os Eventos	2,04%	-1,45%	1,33%	-0,11%	-0,56%
Média	Modelo 3 – Grupo 1	36,24%	11,87%	-2,06%	0,73%	4,27%
Σ	Modelo 4 – Grupo 2	6,61%	0,19%	-2,01%	-1,40%	-1,67%
	Modelo 5 – Grupo 3	2,11%	-0,03%	1,11%	-0,05%	-1,32%
	Modelo 1 – Série Completa	-1,76%	-0,82%	3,59%	-1,54%	-1,19%
-0 ,⊚	Modelo 2 – Todos os Eventos	-4,01%	-0,92%	5,37%	-0,61%	-1,29%
Desvio- padrão	Modelo 3 – Grupo 1	-20,16%	-13,78%	-0,96%	-6,62%	-10,98%
De	Modelo 4 – Grupo 2	-9,36%	-4,88%	1,72%	-0,88%	-3,22%
	Modelo 5 – Grupo 3	-4,65%	-1,82%	5,39%	-0,99%	-1,24%
	Modelo 1 – Série Completa	2,05%	-1,86%	-20,60%	-7,20%	-1,36%
triŝ	Modelo 2 – Todos os Eventos	7,28%	-0,03%	-20,45%	-6,86%	-0,65%
Assimetria	Modelo 3 – Grupo 1	32,21%	19,70%	4,21%	13,89%	17,18%
\SSi	Modelo 4 – Grupo 2	18,55%	10,64%	-5,46%	3,36%	10,11%
N.	Modelo 5 – Grupo 3	8,39%	1,05%	-20,65%	-4,75%	-1,11%
	Modelo 1 – Série Completa	-1,58%	-0,39%	0,40%	-0,52%	-1,00%
ne	Modelo 2 – Todos os Eventos	2,04%	-1,45%	1,33%	-0,11%	-0,56%
Volume	Modelo 3 – Grupo 1	36,24%	11,87%	-2,06%	0,73%	4,27%
$^{0}\Delta$	Modelo 4 – Grupo 2	6,61%	0,19%	-2,01%	-1,40%	-1,67%
	Modelo 5 – Grupo 3	2,11%	-0,03%	1,11%	-0,05%	-1,32%

Tem-se, da Tabela 4.33, que as vazões mínimas foram discrepantes das observadas, inclusive na previsão das séries por seus próprios modelos – em negrito; as vazões máximas simuladas se mantiveram próximas às observadas, com desvios entre -6,00 a -0,40%; as maiores distorções de vazões máximas foram identificadas nas simulações das séries dos Grupos 1 e 2.

Para as demais estatísticas – média, desvio-padrão, assimetria e volume escoado –, algumas distorções foram identificadas na Tabela 4.33, principalmente nas simulações realizadas pelo Modelo 3 para as séries Completa e de Todos os Eventos. As mesmas situações

foram observadas nas simulações de ambas as séries pelo Modelo 4, porém, com menor impacto negativo.

Ao avaliar as vazões previstas, além dos valores das métricas das simulações, é possível perceber com base na Tabela 4.34 – por meio dos erros relativos das vazões de diferentes tempos de permanência – que as dificuldades do Modelo 3 foram, majoritariamente, na simulação da série completa, para tempos de permanência superiores a 20%; neste quesito, os melhores resultados obtidos pelo modelo foram na previsão das vazões mais altas (Q_5).

TABELA 4.34 – ERROS RELATIVOS DOS PERCENTIS DE VAZÕES DAS SÉRIES SIMULADAS – 1964 A 2004

			Sé	éries simuladas		
		Série Completa	Todos os eventos	Grupo 1	Grupo 2	Grupo 3
	Modelo 1 – Série Completa	-0,33%	0,05%	0,53%	-0,48%	-0,61%
	Modelo 2 – Todos os Eventos	11,60%	-0,92%	0,68%	-0,62%	-0,31%
\mathbf{Q}_{70}	Modelo 3 – Grupo 1	108,70%	24,39%	-1,56%	2,02%	14,19%
	Modelo 4 – Grupo 2	30,09%	3,22%	-2,68%	-2,19%	-0,50%
	Modelo 5 – Grupo 3	12,84%	1,00%	0,84%	-0,41%	-0,59%
	Modelo 1 – Série Completa	-2,94%	-0,32%	1,19%	-0,07%	-1,31%
	Modelo 2 – Todos os Eventos	-0,76%	-2,19%	2,03%	0,23%	-1,10%
Q_{35}	Modelo 3 – Grupo 1	29,89%	8,66%	-2,62%	-0,74%	-0,97%
	Modelo 4 – Grupo 2	2,46%	-1,42%	-1,46%	-1,49%	-4,05%
	Modelo 5 – Grupo 3	-1,03%	-0,82%	2,28%	0,23%	-1,78%
	Modelo 1 – Série Completa	-2,52%	-0,84%	2,57%	0,01%	-1,41%
	Modelo 2 – Todos os Eventos	-2,34%	-2,19%	4,26%	0,84%	-1,16%
Q_{20}	Modelo 3 – Grupo 1	11,58%	2,58%	-1,59%	-1,27%	-1,64%
	Modelo 4 – Grupo 2	-2,70%	-3,32%	-0,71%	-1,36%	-3,23%
	Modelo 5 – Grupo 3	-2,77%	-1,91%	4,42%	0,88%	-2,04%
	Modelo 1 – Série Completa	-1,62%	0,00%	1,09%	-0,61%	-1,22%
	Modelo 2 – Todos os Eventos	-1,71%	-0,88%	2,30%	-0,04%	-0,79%
Q_5	Modelo 3 – Grupo 1	-0,22%	-0,19%	-0,49%	-0,81%	-1,80%
	Modelo 4 – Grupo 2	-3,86%	-1,89%	0,00%	-0,38%	-2,63%
	Modelo 5 – Grupo 3	-2,44%	-0,37%	2,16%	-0,08%	-1,43%

Na previsão das demais séries, as dificuldades do Modelo 3 se mantiveram na previsão das vazões baixas (Q_{70}) e, apesar de algumas exceções, os desvios definidos para este modelo apresentaram maior disparidade que os demais ao preverem uma mesma série. Por outro lado, reforça-se que todos os modelos tiveram bom desempenho segundo as vazões altas (Q_5), com desvios não extrapolando o intervalo de \pm 4%.

4.2.4 Validação dos modelos

As cinco séries de dados – Completa, de Todos os Eventos, e dos eventos dos Grupos 1 a 3 – foram previstas pelos modelos para o período de 2005 a 2015. As medidas de avaliação das previsões das séries pelos diferentes modelos são apresentadas na Tabela 4.35; as diagonais principais da Tabela indicam os resultados obtidos na previsão de séries da mesma categoria utilizada no treinamento do modelo; para cada série — cada coluna os resultados sublinhados representam os modelos com métricas inferiores a 0,90 para KG e NS, superiores a 0,10 para MARE, e os piores resultados para NS Modificado.

TABELA 4.35 – COEFICIENTES KG, NS, NS MODIFICADO E MARE DAS PREVISÕES DAS SÉRIES – 2005 A 2015

			Se	éries simulada	as	
		Completa	Todos os eventos	Grupo 1	Grupo 2	Grupo 3
	Modelo 1 – Série Completa	0,973	0,976	0,974	0,968	0,967
	Modelo 2 – Todos os Eventos	0,949	0,969	0,970	0,970	0,964
KG	Modelo 3 – Grupo 1	0,619	0,820	0,958	0,925	0,875
	Modelo 4 – Grupo 2	0,892	0,949	0,972	0,974	0,960
	Modelo 5 – Grupo 3	0,944	0,967	0,969	0,967	0,967
	Modelo 1 – Série Completa	0,984	0,968	0,950	0,948	0,964
	Modelo 2 – Todos os Eventos	0,982	0,967	0,949	0,948	0,964
NS	Modelo 3 – Grupo 1	0,794	0,903	0,955	0,951	0,942
	Modelo 4 – Grupo 2	0,968	0,964	0,954	0,951	0,958
	Modelo 5 – Grupo 3	0,981	0,968	0,948	0,948	0,963
	Modelo 1 – Série Completa	0,104	0,151	0,089	0,081	0,139
NG	Modelo 2 – Todos os Eventos	0,001	0,152	0,075	0,067	0,160
NS Mod.	Modelo 3 – Grupo 1	<u>-10,341</u>	<u>-1,583</u>	0,211	0,128	<u>-0,382</u>
wiou.	Modelo 4 – Grupo 2	-0,739	0,053	0,174	0,161	0,016
	Modelo 5 – Grupo 3	-0,067	0,142	<u>0,061</u>	0,082	0,151
	Modelo 1 – Série Completa	0,068	0,068	0,082	0,067	0,058
	Modelo 2 – Todos os Eventos	0,141	0,064	0,084	0,068	0,060
MARE	Modelo 3 – Grupo 1	0,873	<u>0,201</u>	0,072	0,074	0,133
	Modelo 4 – Grupo 2	0,275	0,082	0,071	0,062	0,072
	Modelo 5 – Grupo 3	0,156	0,073	0,085	0,069	0,057

As mesmas combinações de séries e modelos demarcadas na Tabela 4.32 se destacaram na Tabela 4.35. As medidas das previsões demonstraram performances, em geral, muito próximas às das simulações da etapa anterior, com algumas exceções, principalmente para o coeficiente NS Modificado.

O desempenho das previsões das diagonais principais da Tabela 4.35 foi satisfatório quando os valores são comparados em suas respectivas colunas; esperava-se que cada série, ao ser simulada por outro modelo – não treinado com os dados de mesma categoria – apresentasse piores resultados, quando comparados àqueles da simulação realizada pelo seu próprio modelo. Apesar de, nem sempre, os valores da diagonal principal terem retornado as melhores métricas, os resultados representam um desempenho equivalente.

O Modelo 3 se destacou negativamente, mais uma vez, com os piores desempenhos na previsão das séries Completa e de Todos os Eventos. O comportamento dos modelos também pode ser observado estatisticamente, conforme apresenta a Tabela 4.36.

TABELA 4.36 – ERROS RELATIVOS ENTRE OS PARÂMETROS ESTATÍSTICOS DAS SÉRIES SIMULADAS E OBSERVADAS – 2005 A 2015.

		-	Se	éries simuladas		
		Série Completa	Todos os eventos	Grupo 1	Grupo 2	Grupo 3
	Modelo 1 – Série Completa	23,71%	21,63%	16,27%	11,34%	3,77%
na	Modelo 2 – Todos os Eventos	122,65%	42,61%	28,12%	13,13%	20,78%
Mínima	Modelo 3 – Grupo 1	762,31%	272,56%	90,73%	45,67%	136,36%
Ξ	Modelo 4 – Grupo 2	263,99%	104,04%	44,50%	25,03%	45,41%
	Modelo 5 – Grupo 3	144,12%	60,34%	29,01%	18,38%	18,87%
	Modelo 1 – Série Completa	-3,65%	-3,41%	0,26%	-5,16%	-3,59%
ma	Modelo 2 – Todos os Eventos	-3,93%	-3,89%	-0,12%	-5,34%	-3,87%
Máxima	Modelo 3 – Grupo 1	-1,23%	-0,98%	-3,25%	-2,81%	-1,06%
Ä	Modelo 4 – Grupo 2	1,35%	1,51%	-1,22%	-2,43%	1,37%
	Modelo 5 – Grupo 3	-3,33%	-3,10%	-1,07%	-4,95%	-3,39%
a	Modelo 1 – Série Completa	-1,27%	-0,41%	-0,01%	-0,12%	-1,34%
	Modelo 2 – Todos os Eventos	1,72%	-1,59%	0,81%	0,22%	-1,17%
Média	Modelo 3 – Grupo 1	32,11%	10,38%	-1,96%	0,99%	3,51%
\geq	Modelo 4 – Grupo 2	5,69%	0,03%	-1,59%	-0,73%	-1,89%
	Modelo 5 – Grupo 3	2,32%	0,23%	0,87%	0,82%	-1,78%
	Modelo 1 – Série Completa	-2,30%	-1,69%	0,68%	-1,84%	-2,38%
-0 20 30	Modelo 2 – Todos os Eventos	-4,69%	-2,16%	1,50%	-1,29%	-2,92%
Desvio- padrão	Modelo 3 – Grupo 1	-20,46%	-14,62%	-3,00%	-7,06%	<u>-11,74%</u>
De pa	Modelo 4 – Grupo 2	-9,17%	-4,82%	0,89%	0,01%	-2,86%
	Modelo 5 – Grupo 3	-4,99%	-2,86%	1,55%	-1,88%	-2,20%
	Modelo 1 – Série Completa	0,23%	-4,84%	-26,12%	-8,09%	-6,51%
triŝ	Modelo 2 – Todos os Eventos	4,77%	-1,66%	-28,22%	-8,96%	-1,66%
Assimetria	Modelo 3 – Grupo 1	27,44%	27,26%	-0,15%	9,35%	45,95%
\SSi	Modelo 4 – Grupo 2	<u>16,99%</u>	16,55%	-4,70%	-0,02%	31,19%
•	Modelo 5 – Grupo 3	5,97%	1,27%	<u>-28,74%</u>	-8,29%	1,51%
	Modelo 1 – Série Completa	-1,27%	-0,41%	-0,01%	-0,12%	-1,34%
ne	Modelo 2 – Todos os Eventos	1,72%	-1,59%	0,81%	0,22%	-1,17%
Volume	Modelo 3 – Grupo 1	32,11%	10,38%	-1,96%	0,99%	3,51%
O	Modelo 4 – Grupo 2	5,69%	0,03%	-1,59%	-0,73%	-1,89%
	Modelo 5 – Grupo 3	2,32%	0,23%	0,87%	0,82%	-1,78%

Conforme também observado na Tabela 4.35 (medidas de qualidade), pouca foi a variação entre os resultados da calibração e da validação. Na Tabela 4.36, os erros relativos das vazões mínimas simuladas foram discrepantes, inclusive ao estimar as séries por seus próprios modelos – em negrito; as vazões máximas simuladas se mantiveram próximas às observadas, com desvios entre -5,50 e 1,51%; as maiores distorções para as vazões máximas foram identificadas nas simulações da série do Grupo 2.

Para as demais estatísticas – média, desvio-padrão, assimetria e volume escoado –, algumas distorções foram identificadas na Tabela 4.36, principalmente nas simulações realizadas pelo Modelo 3 para as séries Completa e de Todos os Eventos.

Ao avaliar as vazões previstas por meio dos erros relativos das vazões de diferentes tempos de permanência, para o período de 2005 a 2015, conforme Tabela 4.37, nota-se o mesmo comportamento dos percentis de vazões destacado na etapa de calibração (Tabela 4.34): (i) as dificuldades do Modelo 3 foram observadas, em sua maioria, na simulação da série completa para tempos de permanência superiores a 20%; (ii) os melhores resultados, segundo os erros relativos das vazões, foram identificados para a previsão de vazões mais altas (Q₅), segundo todos os modelos, com desvios não extrapolando o intervalo de ± 4%.

TABELA 4.37 – ERROS RELATIVOS DOS PERCENTIS DE VAZÕES DAS SÉRIES SIMULADAS – 2005 A 2015

		-	Sé	éries simuladas		
		Série Completa	Todos os eventos	Grupo 1	Grupo 2	Grupo 3
	Modelo 1 – Série Completa	-0,20%	0,27%	1,89%	0,15%	-1,47%
	Modelo 2 – Todos os Eventos	10,36%	-0,86%	2,32%	-0,03%	-1,46%
\mathbf{Q}_{70}	Modelo 3 – Grupo 1	99,15%	23,18%	-1,80%	1,46%	7,50%
	Modelo 4 – Grupo 2	26,68%	2,63%	-0,64%	-2,41%	-2,81%
	Modelo 5 – Grupo 3	12,02%	2,09%	3,28%	1,03%	-2,97%
	Modelo 1 – Série Completa	-1,99%	0,34%	1,55%	0,58%	-0,94%
	Modelo 2 – Todos os Eventos	-0,62%	-1,71%	2,42%	1,33%	-1,19%
Q_{35}	Modelo 3 – Grupo 1	28,46%	7,31%	-2,74%	-0,20%	-0,50%
	Modelo 4 – Grupo 2	2,33%	-0,19%	-1,47%	-0,26%	-3,92%
	Modelo 5 – Grupo 3	0,63%	0,54%	2,59%	1,67%	-1,71%
	Modelo 1 – Série Completa	-1,29%	-0,63%	2,02%	0,62%	-2,23%
	Modelo 2 – Todos os Eventos	-1,44%	-1,41%	3,77%	1,23%	-2,34%
Q_{20}	Modelo 3 – Grupo 1	8,38%	1,38%	-0,52%	0,27%	-3,13%
	Modelo 4 – Grupo 2	-2,84%	-2,69%	-0,18%	0,77%	-4,87%
	Modelo 5 – Grupo 3	-1,26%	-1,04%	3,56%	1,50%	-2,83%
	Modelo 1 – Série Completa	-1,70%	0,07%	-1,47%	-1,26%	-0,92%
	Modelo 2 – Todos os Eventos	-1,87%	-0,55%	-0,86%	-0,68%	-0,81%
\mathbf{Q}_{5}	Modelo 3 – Grupo 1	-1,33%	0,22%	-2,85%	-2,15%	-0,86%
	Modelo 4 – Grupo 2	-3,78%	-1,03%	-1,25%	-0,86%	-0,68%
	Modelo 5 – Grupo 3	-1,94%	0,24%	-0,65%	-0,90%	-0,82%

Desta forma, tem-se a partir dos resultados dos critérios de qualidade de ajuste, apresentados nas Tabelas 4.32 e 4.35 para os dois períodos analisados, que os Modelos 1, 2 e 5 foram, de certo modo, equivalentes na simulação de todas as séries; em praticamente todas as simulações, o Modelo 1 foi capaz de melhor reproduzir os parâmetros estatísticos das séries, incluindo os volumes de escoamento.

Por fim, os Modelos 3 e 4, apesar os resultados inferiores dos coeficientes KG, NS, NS Modificado e MARE (Tabelas 4.32 e 4.35), quando comparados aos demais modelos, conseguiram preservar as vazões máximas (Tabelas 4.33 e 4.36) e garantiram pequenos desvios das vazões altas (**Q**₅) (Tabelas 4.34 e 4.37).

4.3 PREVISÃO DE VAZÕES

Nesta Seção, são apresentadas as previsões da série de Todos os Eventos, para o período de 2005 a 2015, obtidas a partir das configurações de previsão apresentados no Capítulo 3 – Seção 3.4: (i) previsão média; (ii) previsão por grupo de pertencimento (iii) previsão ponderada pelas distâncias Euclidianas; e (iv) previsão ponderada pelos graus de pertencimento.

4.3.1 Configuração 1 – Previsão Média

A Tabela 4.38 apresenta as medidas de qualidade da previsão média (Configuração 1), que indica uma boa estimativa de vazões segundo as métricas KG, NS e MARE; NS Modificado apresentou resultado positivo, mas próximo de zero. A Tabela 4.39 apresenta os parâmetros estatísticos das séries observada e prevista pela Configuração 1.

TABELA 4.38 – MEDIDAS DE QUALIDADE DA PREVISÃO MÉDIA – 2005 A 2015

	KG	NS	NS Mod	MARE
Previsão média	0,940	0,965	0,072	0,089

TABELA 4.39 – PARÂMETROS ESTATÍSTICOS DAS SÉRIES OBSERVADA E PREVISTAS PELA MÉDIA – 2005 A 2015

	Mínima (m³/s)	Máxima (m³/s)	Média (m³/s)	Desvio- padrão (m³/s)	Assimetria	Volume (10 ¹³ m³)
Observada	132,2	3385,5	1004,6	503,9	1,3	22,4
Previsão média	268,3	3310,7	1025,2	476,6	1,4	22,9

Neste cenário, segundo a Tabela 4.39, a vazão mínima foi estimada em um valor duas vezes maior que a observada; a vazão máxima, apesar de ligeiramente subestimada, sofreu menor variação em relação ao observado – em torno de 2%. Nota-se ainda uma diminuição no desvio-padrão, e um aumento na média, na assimetria e no volume escoado.

A Figura 4.9 apresenta os Hidrogramas Observado e Previsto pela Configuração 1 – Previsão Média – e evidencia a superestimação das vazões mais baixas, principalmente aquelas entre dois eventos consecutivos, e uma leve redução de alguns picos de cheia, inclusive a maior cheia observada no período – como destacado na Tabela 4.39. Nota-se que, nos hidrogramas, os eventos foram posicionados sequencialmente, de forma que o final de um evento indica o início do próximo.

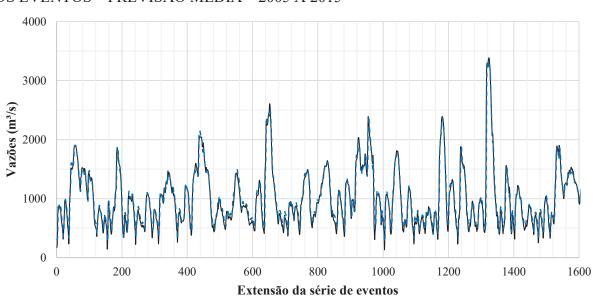


FIGURA 4.9 – HIDROGRAMAS PREVISTO E OBSERVADO DA SÉRIE COM TODOS OS EVENTOS – PREVISÃO MÉDIA – 2005 A 2015

4.3.2 Configuração 2 – Previsão por Grupo de Pertencimento

Observado

A Tabela 4.40 apresenta as medidas de qualidade da previsão grupo de pertencimento (Configuração 2), que indica uma boa estimativa de vazões segundo as métricas KG, NS e MARE; NS Modificado apresentou resultado negativo. A Tabela 4.41 apresenta os parâmetros estatísticos das séries observada e prevista pela Configuração 2.

- - Previsão média

TABELA 4.40 – MEDIDAS DE QUALIDADE DA PREVISÃO POR GRUPO DE PERTENCIMENTO – 2005 A 2015

	KG	NS	NS Mod	MARE
Previsão por Grupo de Pertencimento	0,907	0,944	-0,497	0,127

TABELA 4.41 – PARÂMETROS ESTATÍSTICOS DAS SÉRIES DE EVENTOS OBSERVADA E PREVISTAS POR GRUPO DE PERTENCIMENTO – 2005 A 2015

	Mínima (m³/s)	Máxima (m³/s)	Média (m³/s)	Desvio- padrão (m³/s)	Assimetria	Volume (10 ¹³ m³)
Observada	132,2	3385,5	1004,6	503,9	1,3	22,4
Previsão por Grupo de Pertencimento	443,8	3280,7	1044,9	463,1	1,5	23,8

Neste cenário, segundo a Tabela 4.41, a vazão mínima foi estimada em um valor três vezes maior que a observada; a vazão máxima, apesar de subestimada, foi o parâmetro que

sofreu menor variação em relação ao observado. Nota-se ainda uma diminuição no desviopadrão, e um aumento na média, na assimetria e no volume escoado.

A Figura 4.10 apresenta os Hidrogramas Observado e Previsto pela Configuração 2 – Previsão por Grupo de Pertencimento – e evidencia a superestimação tanto das vazões mais baixas – principalmente aquelas entre dois eventos consecutivos – quanto de alguns picos de cheia, inclusive a maior cheia observada no período – como destacado na Tabela 4.41. Nota-se que, nos hidrogramas, os eventos foram posicionados sequencialmente, de forma que o final de um evento indica o início do próximo.

4000 3000 Vazões (m³/s) 2000 1000 1000 1200 200 400 600 1400 800 1600 Extensão da série de eventos Observado --- Previsão por grupo de pertencimento

FIGURA 4.10 – HIDROGRAMAS PREVISTO E OBSERVADO DA SÉRIE COM TODOS OS EVENTOS – PREVISÃO POR GRUPO DE PERTENCIMENTO – 2005 A 2015

4.3.3 Configuração 3 – Previsão Ponderada por Distâncias Euclidianas

A Tabela 4.42 apresenta as medidas de qualidade da previsão ponderada por distâncias Euclidianas (Configuração 3), que indica uma boa estimativa de vazões segundo as métricas KG, NS e MARE; NS Modificado apresentou resultado negativo, próximo de zero. A Tabela 4.43 apresenta os parâmetros estatísticos das séries observada e prevista pela Configuração 3.

TABELA 4.42 – MEDIDAS DE QUALIDADE DA PREVISÃO PONDERADA POR DISTÂNCIAS EUCLIDIANAS – 2005 A 2015

	KG	NS	NS Mod	MARE
Previsão ponderada por distâncias Euclidianas	0,938	0,959	-0,087	0,097

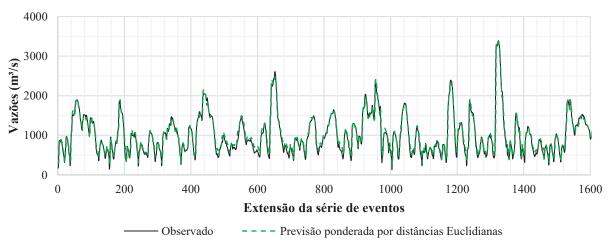
TABELA 4.43 – PARÂMETROS ESTATÍSTICOS DAS SÉRIES DE EVENTOS OBSERVADA E PREVISTAS PELA PONDERAÇÃO POR DISTÂNCIAS EUCLIDIANAS – 2005 A 2015

	Mínima (m³/s)	Máxima (m³/s)	Média (m³/s)	Desvio-padrão (m³/s)	Assimetria	Volume (10 ¹³ m³)
Observada	132,2	3385,5	1004,6	503,9	1,3	22,4
Previsão ponderada por distâncias Euclidianas	255,8	3354,9	1027,9	476,7	1,4	22,9

Neste cenário, segundo a Tabela 4.43, a vazão mínima foi estimada em um valor quase duas vezes maior que a observada; a vazão máxima, apesar de ligeiramente subestimada, mais uma vez foi o parâmetro que sofreu menor variação em relação ao observado. Nota-se ainda uma diminuição no desvio-padrão, e um sutil aumento na média, na assimetria e no volume escoado.

A Figura 4.11 apresenta os Hidrogramas Observado e Previsto pela Configuração 3 – Previsão Ponderada por Distâncias Euclidianas – e evidencia a superestimação tanto das vazões mais baixas e de alguns picos de cheia, inclusive a maior cheia observada no período – como destacado na Tabela 4.43.

FIGURA 4.11 – HIDROGRAMAS PREVISTO E OBSERVADO DA SÉRIE COM TODOS OS EVENTOS – PREVISÃO PONDERADA POR DISTÂNCIAS EUCLIDIANAS – 2005 A 2015



Nota-se que, nos hidrogramas da Figura 4.11, os eventos foram posicionados sequencialmente, de forma que o final de um evento indica o início do próximo.

4.3.4 Configuração 4 – Previsão Ponderada por Graus de Pertencimento

A Tabela 4.44 apresenta as medidas de qualidade da previsão por graus de pertencimento (Configuração 4), que indica uma boa estimativa de vazões segundo as métricas

KG, NS e MARE; NS Modificado apresentou resultado negativo. A Tabela 4.45 apresenta os parâmetros estatísticos das séries observada e prevista pela Configuração 4.

TABELA 4.44 – MEDIDAS DE QUALIDADE DA PREVISÃO PONDERADA POR GRAUS DE PERTENCIMENTO – 2005 A 2015

	KG	NS	NS Mod	MARE
Previsão ponderada por graus de pertencimento	0,905	0,945	-0,456	0,126

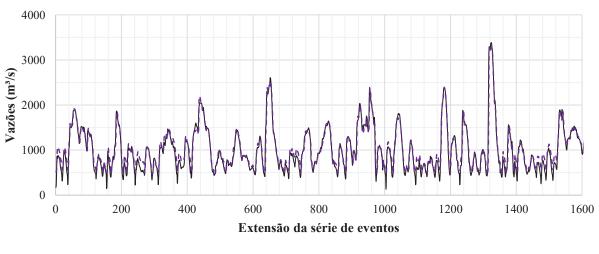
TABELA 4.45 – PARÂMETROS ESTATÍSTICOS DAS SÉRIES OBSERVADA E PREVISTAS PELA PONDERADA POR GRAUS DE PERTENCIMENTO – 2005 A 2015

	Mínima (m³/s)	Máxima (m³/s)	Média (m³/s)	Desvio-padrão (m³/s)	Assimetria	Volume (10 ¹³ m³)
Observada	132,2	3385,5	1004,6	503,9	1,3	22,4
Previsão ponderada por graus de pertencimento	462,2	3283,6	1045,1	462,1	1,5	23,3

Neste cenário, segundo a Tabela 4.45, a vazão mínima foi estimada em um valor mais de três vezes maior que a observada; a vazão máxima, apesar de subestimada, mais uma vez foi o parâmetro que sofreu menor variação em relação ao observado. Nota-se ainda uma diminuição no desvio-padrão, e um aumento na média, na assimetria e no volume escoado.

A Figura 4.12 apresenta os Hidrogramas Observado e Previsto pela Configuração 4 – Previsão Ponderada por Graus de Pertencimento – e evidencia a superestimação das vazões baixas, principalmente aquelas entre dois eventos consecutivos. Nota-se que, na Figura 4.12, os eventos foram posicionados sequencialmente, sendo o final de um evento o início do próximo.

FIGURA 4.12 – HIDROGRAMAS PREVISTO E OBSERVADO DA SÉRIE COM TODOS OS EVENTOS – PREVISÃO PONDERADA POR GRAUS DE PERTENCIMENTO – 2005 A 2015



Observado ---- Previsão ponderada por graus de pertencimento

4.4 COMPARATIVO ENTRE AS CONFIGURAÇÕES DE PREVISÃO

A Figura 4.13 apresenta a variação das medidas de qualidade KG, NS e MARE das previsões obtidas pelos modelos calibrados e pelas Configurações de 1 a 4, enquanto a Figura 4.14, a variação de NS Modificado para as mesmas previsões.

0,00 1,00 0,95 0,05 **S** 0,90 **O** 0,85 0,80 0,10 0,15 0,20 0,80 0,75 0,25 0,70 0,30 Configuração Confightação Previsões NS •MARE KG

FIGURA 4.13 – COMPORTAMENTO DOS COEFICIENTES KG, NS E MARE NAS PREVISÕES – 2005 A 2015

Modelo 1 – Série Completa; Modelo 2 – Todos os Eventos; Modelo 3 – Grupo 1; Modelo 4 – Grupo 2; Modelo 5 – Grupo 3; Configuração 1 – Previsão Média; Configuração 2 – Previsão por Grupo de Pertencimento; Configuração 3 – Previsão Ponderada por Distâncias Euclidianas; Configuração 4 – Previsão Ponderada por Graus de Pertencimento.

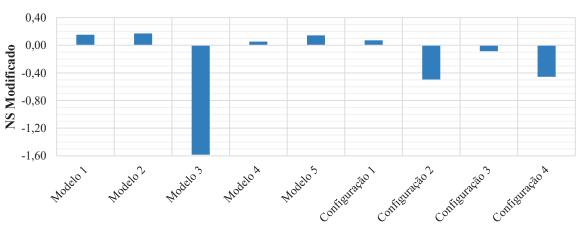


FIGURA 4.14 – COMPORTAMENTO DO COEFICIENTE NS MODIFICADO NAS PREVISÕES – 2005 A 2015

Previsões

Modelo 1 – Série Completa; Modelo 2 – Todos os Eventos; Modelo 3 – Grupo 1; Modelo 4 – Grupo 2; Modelo 5 – Grupo 3; Configuração 1 – Previsão Média; Configuração 2 – Previsão por Grupo de Pertencimento; Configuração 3 – Previsão Ponderada por Distâncias Euclidianas; Configuração 4 – Previsão Ponderada por Graus de Pertencimento.

As performances dos Modelos 1, 2 e 5, e as previsões média e ponderada pelas distâncias Euclidianas se destacaram das demais, segundo KG e MARE (Figura 4.13); segundo NS (Figura 4.13), os resultados entre as previsões dos Modelos e das Configurações foram próximos entre si, com exceção do resultado do Modelo 3, que teve desempenho inferior para todas as métricas. Apesar dos resultados próximos de zero para NS Modificado (Figura 4.14), os Modelos 1, 2 e 5 se destacaram dos demais, enquanto o Modelo 3 e as Configurações de Previsão por Grupo de Pertencimento e Ponderada por Graus de Pertencimento apresentaram os piores comportamentos segundo a métrica.

Na Figura 4.15, pode-se observar os erros relativos entre as estatísticas das séries de vazões observada e previstas; as diferenças entre as vazões mínimas não foram apresentadas graficamente, contudo, reforça-se que estes parâmetros foram superestimados em todos as previsões e os erros relativos variaram de 20 a 270%, aproximadamente, com valores mais baixos atingidos pelos Modelos 1, 2 e 5.

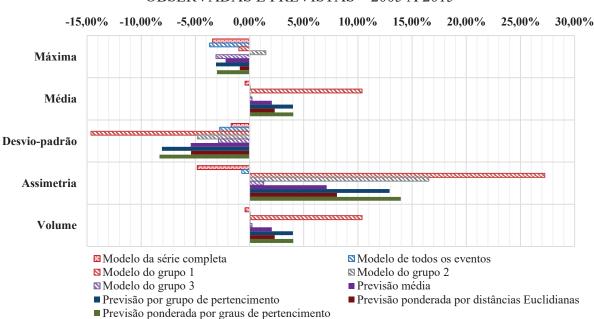


FIGURA 4.15 – ERROS RELATIVOS ENTRE AS ESTATÍSTICAS DAS VAZÕES OBSERVADAS E PREVISTAS – 2005 A 2015

Observa-se na Figura 4.15 que as vazões máximas apresentaram o menor intervalo de erros relativos dentre as previsões, variando de -3,70 a 1,50%; para este parâmetro, o Modelo 3 e a Previsão Ponderada pelas Distâncias Euclidianas apresentaram desvios mais próximos de zero, enquanto os Modelos 1, 2 e 5 e as Previsões Média, por Grupo de Pertencimento e Ponderada por Graus de Pertencimento, os mais distantes; o Modelo 4 foi o único a superestimar a vazão máxima. Para as vazões médias, tem-se, a partir da Figura 4.15, desvios entre -0,40 a

10,40%, sendo que os Modelos 1, 2, 4 e 5 apresentaram valores próximos de zero, enquanto o Modelo 3, o maior desvio; as Configurações 1 a 4 apresentaram desvios intermediários e inferiores a 4%. Reforça-se, ainda, para as vazões médias, que, com exceção daquelas obtidas pelo Modelo 1, todas foram superestimadas.

A Figura 4.16 apresenta as Curvas de Permanência Observada em União da Vitória – no período de 2005 a 2015 – e Previstas; as vazões de permanência 70, 35, 20 e 5% previstas foram extraídas e relacionadas às observadas; os erros relativos das referidas vazões foram explicitados na Figura 4.17.

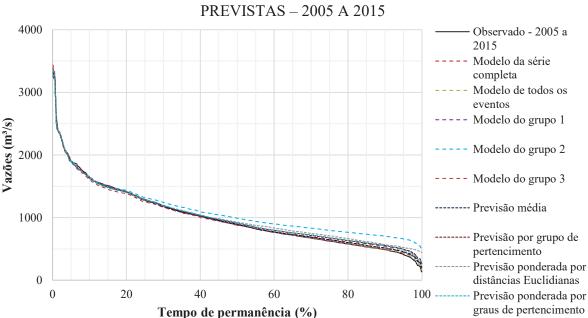


FIGURA 4.16 – CURVAS DE PERMANÊNCIA DAS VAZÕES OBSERVADA E PREVISTAS – 2005 A 2015

De acordo com a Figura 4.17, as vazões com tempo de permanência de 70% (\mathbf{Q}_{70}) e 35% (\mathbf{Q}_{35}) foram superestimadas por todas as previsões, o que justifica a elevação das médias em praticamente todas as simulações (Figura 4.15); para \mathbf{Q}_{35} , no entanto, todos os desvios foram inferiores a 4%, com exceção daquele obtido pelo Modelo 4, que atingiu quase 8% de erro relativo. Os erros relativos das vazões de tempos de permanência de 20 (\mathbf{Q}_{20}) e 5% (\mathbf{Q}_{5}) oscilaram entre -3 e +2%; em sua maioria, as vazões \mathbf{Q}_{20} e \mathbf{Q}_{5} foram subestimadas nas previsões.

Os resultados da Figura 4.17 indicam que os Modelos 1, 2 e 3 e a previsão média conseguiram reproduzir adequadamente as vazões com tempo de permanência inferiores a 35%, com erros relativos de -1,0 a 0,7%. Por outro lado, as demais previsões (Modelos 4 e 5, por Grupo de Pertencimento, Ponderada pelas Distâncias Euclidianas e Ponderada pelos Graus de Pertencimento) demonstraram melhora ao prever vazões mais elevadas (\mathbf{Q}_{20} e \mathbf{Q}_{5}), com erros relativos entre -2,7 a 0,0% para \mathbf{Q}_{20} , e de -1,0 a 0,3% para \mathbf{Q}_{5} . As previsões por grupo de

pertencimento, ponderada pelas distâncias Euclidianas e por graus de pertencimento mantiveram uma decrescente de seus desvios, em módulo, a partir de **Q**70, enquanto as demais apresentaram oscilações, o que indicou uma melhora de desempenho ao preverem vazões mais altas.

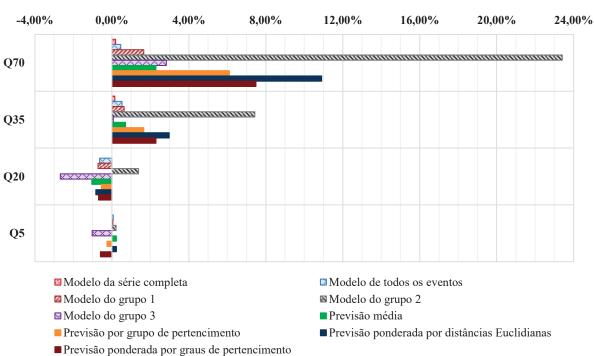


FIGURA 4.17 – ERROS RELATIVOS DE VAZÕES DE DIFERENTES TEMPOS DE PERMANÊNCIA – 2005 A 2015

4.5 AVALIAÇÃO DA PREVISÃO POR EVENTOS

O Apêndice F apresenta os coeficientes obtidos para cada evento nas diferentes configurações de previsão; as Tabelas 4.46 a 4.49 apresentam os intervalos de resultados de KG, NS, NS Modificado e MARE do Apêndice F.

TABELA 4.46 – INTERVALOS DE RESULTADOS DO COEFICIENTE KG PARA AS PREVISÕES DE EVENTOS DE 2005 A 2015

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Previsão média	Previsão por grupo	Previsão distância Euclidiana	Previsão Pertenc.
Mínimo	0,716	0,728	0,482	0,733	0,753	0,727	0,482	0,749	0,483
Máximo	0,985	0,986	0,909	0,976	0,982	0,966	0,979	0,966	0,980
Amplitude	0,269	0,258	0,427	0,243	0,229	0,239	0,498	0,217	0,497
Desvio-padrão	0,063	0,056	0,113	0,064	0,058	0,059	0,141	0,057	0,131
Média	0,914	0,912	0,697	0,875	0,904	0,881	0,791	0,878	0,792

TABELA 4.47 – INTERVALOS DE RESULTADOS DO COEFICIENTE NS PARA AS PREVISÕES DE EVENTOS DE 2005 A 2015

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Previsão média	Previsão por grupo	Previsão distância Euclidiana	Previsão Pertenc.
Mínimo	0,404	0,420	-4,885	0,289	0,454	0,034	-2,499	-0,171	-2,370
Máximo	0,982	0,982	0,950	0,974	0,981	0,978	0,975	0,964	0,975
Amplitude	0,578	0,562	5,835	0,685	0,527	0,944	3,473	1,135	3,345
Desvio-padrão	0,131	0,127	1,363	0,148	0,125	0,198	0,900	0,191	0,842
Média	0,855	0,859	-0,100	0,839	0,855	0,803	0,395	0,810	0,429

TABELA 4.48 – INTERVALOS DE RESULTADOS DO COEFICIENTE NS MODIFICADO PARA AS PREVISÕES DE EVENTOS DE 2005 A 2015

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Previsão média	Previsão por grupo	Previsão distância Euclidiana	Previsão Pertenc.
Mínimo	-0,485	-0,645	-20,112	-0,845	-2,369	-2,028	-13,859	-5,728	-13,800
Máximo	0,491	0,505	0,192	0,430	0,626	0,417	0,469	0,505	0,461
Amplitude	0,976	1,150	20,305	1,275	2,995	2,444	14,328	6,233	14,261
Desvio-padrão	0,162	0,196	4,790	0,288	0,468	0,522	2,758	1,103	2,531
Média	0,164	0,174	-4,534	0,040	0,103	-0,113	-1,561	-0,287	-1,412

TABELA 4.49 – INTERVALOS DE RESULTADOS DO COEFICIENTE MARE PARA AS PREVISÕES DE EVENTOS DE 2005 A 2015

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Previsão média	Previsão por grupo	Previsão distância Euclidiana	Previsão Pertenc.
Mínimo	0,042	0,042	0,084	0,051	0,045	0,047	0,057	0,046	0,058
Máximo	0,150	0,165	0,518	0,217	0,152	0,226	0,518	0,197	0,515
Amplitude	0,108	0,123	0,434	0,166	0,107	0,179	0,462	0,151	0,457
Desvio-padrão	0,025	0,027	0,098	0,034	0,025	0,036	0,132	0,030	0,128
Média	0,078	0,081	0,251	0,094	0,081	0,104	0,172	0,104	0,169

A partir dos coeficientes KG, NS, NS Modificado e MARE do Apêndice F, tem-se que, em geral, os Modelos 1, 2, 4 e 5 apresentaram comportamentos semelhantes na previsão dos eventos, ou seja, quando os Modelos 1 e 2 apresentaram bom desempenho ao prever um evento, os Modelos 4 e 5 também performaram adequadamente, da mesma forma que quando o desempenho dos Modelos 1 e 2 foram inferiores, os Modelos 4 e 5 também apresentaram piora. Isto refletiu nos resultados apresentados nas Tabelas 4.46 a 4.49, em que os parâmetros dos modelos foram próximos entre si; o Modelo 3 apresentou sempre desempenho inferior aos Modelos 1, 2, 4 e 5, refletindo nas médias das Tabelas 4.46 a 4.49; o Modelo 2 foi utilizado como base nas análises posteriores.

As previsões por grupo de pertencimento e ponderada por graus de pertencimento, no entanto, apresentaram os piores desempenhos dentre as simulações, com coeficientes KG, NS,

NS Modificado e MARE (Apêndice F) ainda mais insatisfatórios para os eventos em que o Modelo 2 não performou adequadamente, como, por exemplo, nos eventos com ID 6 e 8, cujos picos observados foram de 957,8 e 722 m³/s, respectivamente. Os desempenhos das previsões por grupo de pertencimento e ponderada por graus de pertencimento se destacaram no evento com ID 24, que atingiram os mesmos patamares de valores dos coeficientes das demais previsões, com exceção daquela gerada pelo Modelo 3. O hidrograma do evento de ID 24 é apresentado na Figura 4.18, a partir da qual se observa que o Modelo 3 indicou maior descolamento do hidrograma observado.

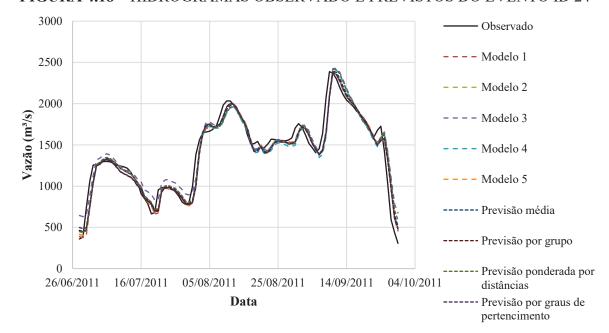


FIGURA 4.18 – HIDROGRAMAS OBSERVADO E PREVISTOS DO EVENTO ID 24

As previsões média e ponderada pelas distâncias Euclidianas, observadas no Apêndice F e nas Tabelas 4.46 a 4.49, apresentaram comportamentos semelhantes ao Modelo 2, porém, naqueles eventos em que o referido modelo não teve desempenho satisfatório, as previsões média e ponderada pelas distâncias Euclidianas apresentaram performances inferiores, o que pode ter sido influenciado pelos resultados menos adequados do Modelo 3.

No Apêndice G, são apresentados os erros relativos referentes aos parâmetros estatísticos de cada evento previsto pelas diferentes configurações, dando origem às Tabelas 4.50 a 4.53, que apresentam os intervalos dos erros para cada previsão; não foram apresentadas as Tabelas resumo dos erros relativos das mínimas vazões previstas e das assimetrias.

Os erros relativos das máximas vazões previstas (Tabela 4.50) indicaram que, isoladamente, os Modelos 1, 2, 4 e 5 tiveram menor oscilação (menor amplitude) de resultados entre os eventos, e as menores média e mediana, enquanto as previsões por grupo de

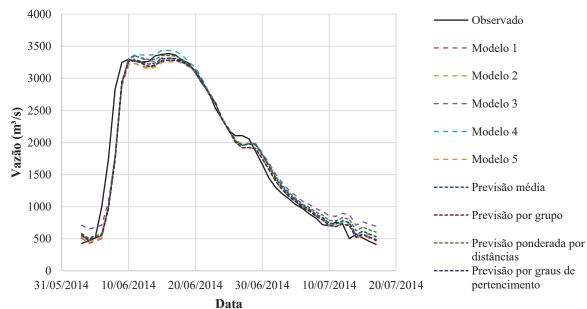
pertencimento, ponderada por graus de pertencimento e do Modelo 3 apresentaram amplitude, média e mediana superiores aos patamares dos referidos modelos isolados. Por outro lado, as previsões média e ponderada pelas distâncias Euclidianas apresentaram amplitudes mais baixas, como os Modelos 1, 2, 4 e 5, mas com médias e medianas superiores. Segundo o Apêndice G (Tabela G.2), o evento de ID 24 apresentou a menor variação entre as previsões, com parâmetros atingindo desvios de -0,3 a 1,7%, o que indica que todas as Configurações foram capazes de estimar adequadamente a vazão máxima deste evento, como também se observou na Figura 4.18 (Hidrograma ID 24).

TABELA 4.50 – DISTRIBUIÇÃO DOS ERROS RELATIVOS DAS VAZÕES MÁXIMAS PREVISTAS – 2005 A 2015

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Previsão média	Previsão por grupo	Previsão distância Euclidiana	Previsão Pertenc.
Mínimo	-3,41%	-3,68%	-5,30%	-5,22%	-6,18%	-3,80%	-5,22%	-4,81%	-5,22%
Máximo	4,29%	5,66%	26,88%	7,80%	8,04%	9,25%	22,70%	9,79%	22,69%
Amplitude	7,70%	9,34%	32,18%	13,02%	14,23%	13,05%	27,92%	14,60%	27,91%
Média	0,14%	0,39%	8,97%	0,19%	0,28%	1,93%	5,03%	1,36%	4,86%
Mediana	0,15%	0,49%	8,81%	-0,01%	0,03%	1,57%	1,17%	1,51%	1,09%

O evento de ID 38, cujo pico observado é o maior do período de previsão (2005 a 2015), apresentou baixa variação entre as previsões, com erros relativos das vazões máximas entre -3,7 a 1,5%, com os melhores resultados obtidos pelo Modelo 3 e pela previsão ponderada pelas distâncias Euclidianas. A Figura 4.19 apresenta o Hidrograma do evento de ID 38, onde se observa certo descolamento do hidrograma observado na região de pico.

FIGURA 4.19 – HIDROGRAMAS OBSERVADO E PREVISTOS DO EVENTO ID 38



O evento de ID 16 apresentou desvios no intervalo de -0,03 a 8,81%, com menor valor absoluto atingido pela previsão ponderada por graus de pertencimento (-0,03%); o evento ID 16 está apresentado na Figura 4.20.

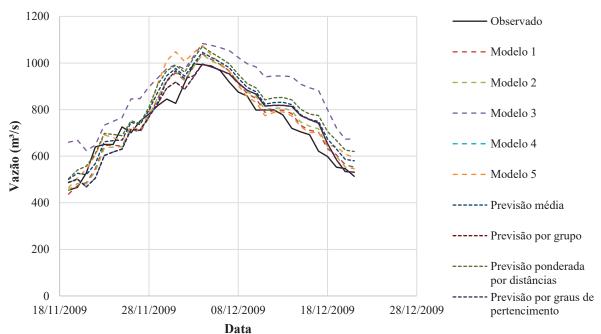


FIGURA 4.20 – HIDROGRAMAS OBSERVADO E PREVISTOS DO EVENTO ID 16

Os hidrogramas previstos apresentados na Figura 4.20 (ID 16) demonstram significativos descolamentos do hidrograma observado, ao longo de todo o evento. Porém, apesar da perceptível diferença no comportamento dos hidrogramas, a máxima variação entre os picos previstos e o observado foi de 87 m³/s, atingida pelo Modelo 3, o que indica que todos os picos foram bem representados, justificando o pequeno intervalo dos desvios de vazão máxima do evento; o comportamento das demais fases dos hidrogramas justificam os coeficientes KG, NS, NS Modificado e MARE obtidos nas previsões do evento (Apêndice F).

O evento de ID 7, por outro lado, apesar de também ter apresentado melhor resultado para a previsão ponderada pelo grau de pertencimento (0,01%), apresentou um intervalo de desvios variando de -0,60 a 3,54%, indicando uma diferença menor entre as previsões, comparado ao evento de ID 16. O evento ID 7 é apresentado na Figura 4.21 e, ao contrário do observado no evento ID 16 (Figura 4.20), os hidrogramas previstos apresentados na Figura 4.21 (ID 7) evidenciam que praticamente não houve divergência entre os resultados de vazões máximas nas previsões do evento ID 7. Reforça-se, ainda, a diferença de magnitude dos picos dos eventos de ID 16 e 7: o evento de menor pico (ID 16) apresentou maior dificuldade para ser estimado pelas Configurações.

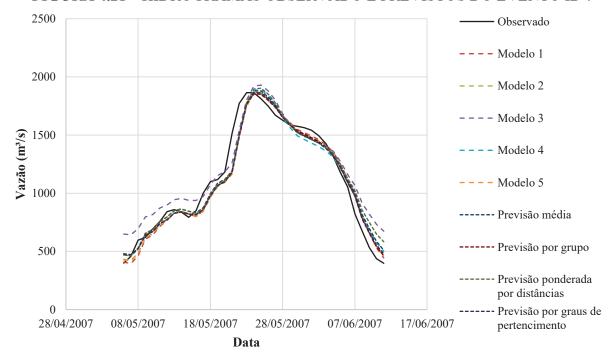


FIGURA 4.21 – HIDROGRAMAS OBSERVADO E PREVISTOS DO EVENTO ID 7

Na Tabela 4.51, os erros relativos das médias vazões previstas indicaram que isoladamente, os Modelos 1, 2, 4 e 5 tiveram menor oscilação (menor amplitude) de resultados entre os eventos e as menores média e mediana, enquanto as previsões por grupo de pertencimento, ponderada pelo grau de pertencimento e do Modelo 3 apresentaram amplitude, média e mediana superiores aos patamares dos referidos modelos isolados. Por outro lado, as previsões média e ponderada pelas distâncias Euclidianas apresentaram amplitudes mais baixas, como os Modelos 1, 2, 4 e 5, mas com médias e medianas superiores.

TABELA 4.51 – DISTRIBUIÇÃO DOS ERROS RELATIVOS DAS VAZÕES MÉDIAS PREVISTAS – 2005 A 2015

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Previsão média	Previsão por grupo	Previsão distância Euclidiana	Previsão Pertenc.
Mínimo	-5,13%	-2,57%	1,05%	-3,17%	-4,57%	-1,40%	-3,17%	-1,53%	-3,11%
Máximo	3,58%	5,20%	37,14%	9,22%	7,13%	11,78%	37,14%	13,40%	37,11%
Amplitude	8,71%	7,77%	36,10%	12,39%	11,70%	13,18%	40,31%	14,93%	40,22%
Média	-0,42%	0,63%	18,28%	1,91%	0,78%	4,24%	10,22%	3,63%	10,06%
Mediana	-0,56%	0,46%	18,35%	1,69%	0,44%	4,10%	3,30%	3,12%	3,67%

Segundo o Apêndice G (Tabela G.3 – erros relativos das vazões médias), o evento de ID 33 apresentou a menor variação entre as previsões, com desvios de -2,44 a 1,16%, indicando que todas as Configurações foram capazes de estimar adequadamente as vazões médias deste evento. O evento ID 33 está apresentado na Figura 4.22, na qual se observa que os maiores

deslocamentos entre os hidrogramas está nos trechos iniciais e finais, mas esta diferença se torna praticamente imperceptível quando o objetivo é avaliar as vazões médias.

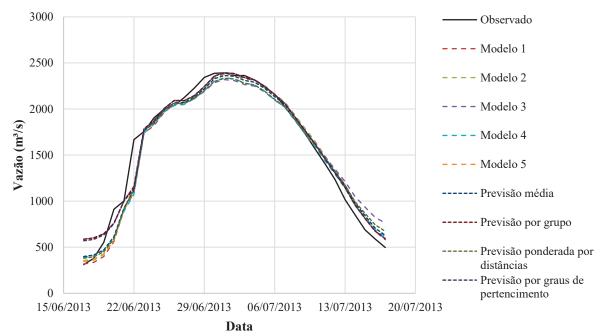


FIGURA 4.22 – HIDROGRAMAS OBSERVADO E PREVISTOS DO EVENTO ID 33

O evento de ID 27, por outro lado, destacou-se por apresentar desvios no intervalo de -0,95 a 5,42% para as previsões, mas com valor de -0,11% para a previsão ponderada por graus de pertencimento, que se destacou como o melhor desempenho neste quesito, seguido da previsão do Modelo 1 e da previsão por grupo de pertencimento; o maior desvio foi observado no Modelo 3, como demonstram os Hidrogramas do evento de ID 27 na Figura 4.23.

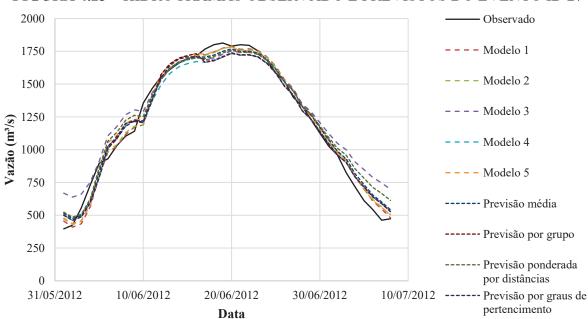


FIGURA 4.23 – HIDROGRAMAS OBSERVADO E PREVISTOS DO EVENTO ID 27

O evento de ID 42 também apresentou melhores resultados para a previsão ponderada por graus de pertencimento (Apêndice G – Tabela G.3), assim como o evento de ID 27. O erro relativo das vazões médias da previsão ponderada por graus de pertencimento se equiparou aos resultados das previsões do Modelo 4 e por grupo de pertencimento, enquanto o Modelo 3 apresentou, novamente, o maior desvio, equivalente a 12,31%, como pode ser observado pelo deslocamento dos hidrogramas na Figura 4.24.

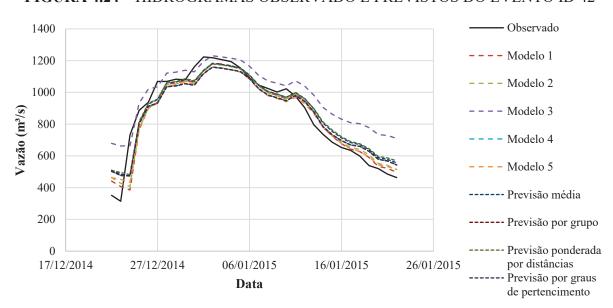


FIGURA 4.24 – HIDROGRAMAS OBSERVADO E PREVISTOS DO EVENTO ID 42

Segundo a Tabela 4.52, os erros relativos dos desvios-padrão das vazões previstas indicaram que, isoladamente, os Modelos 1 e 2 tiveram menor oscilação (menor amplitude) de resultados entre os eventos e as menores média e mediana, enquanto as previsões por grupo de pertencimento, ponderada por graus de pertencimento, e dos Modelos 3 e 4 apresentaram amplitude, média e mediana superiores aos patamares dos referidos modelos isolados. Por outro lado, as previsões média e ponderada pelas distâncias Euclidianas apresentaram amplitudes mais baixas, como os Modelos 1 e 2, mas com médias e medianas superiores.

TABELA 4.52 – DISTRIBUIÇÃO DOS ERROS RELATIVOS DOS DESVIOS-PADRÃO DAS VAZÕES PREVISTAS – 2005 A 2015

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Previsão média	Previsão por grupo	Previsão distância Euclidiana	Previsão Pertenc.
Mínimo	-9,42%	-14,01%	-39,85%	-25,41%	-16,74%	-21,14%	-32,75%	-20,69%	-32,52%
Máximo	16,22%	9,68%	-8,21%	5,17%	10,61%	4,19%	5,17%	4,84%	7,80%
Amplitude	25,64%	23,69%	31,64%	30,58%	27,35%	25,32%	37,92%	25,53%	40,32%
Média	1,49%	-2,01%	-21,86%	-8,55%	-2,91%	-6,86%	-14,44%	-7,25%	-14,31%
Mediana	0,04%	-2,04%	-21,44%	-7,75%	-2,63%	-6,72%	-16,10%	-6,48%	-16,29%

Segundo este critério, a Tabela I.4 do Apêndice G demonstra que o evento de ID 38 apresentou o menor intervalo de erros relativos dos desvios-padrão (de -8,65 a -0,69%), sendo o Modelo 5 o que se manteve mais próximo de zero. O Modelo 3 e a previsão ponderada por distâncias Euclidianas, por outro lado, que tiveram melhores desempenhos segundo os erros relativos das vazões máximas, apresentaram piores performances segundo os erros relativos de desvios-padrão.

Dentre a lista de eventos previstos pela Configuração 4 (previsão ponderada por graus de pertencimento), o evento de ID 15, apresentou um dos melhores resultados quanto aos erros relativos dos desvios-padrão (-0,74%). Porém, este resultado se equiparou àqueles obtidos pelos Modelos 1 e 2, e pela previsão por grupo de pertencimento; as demais previsões apresentaram desvios superiores em módulo. A Figura 4.25 representa os hidrogramas observado e previstos do evento ID 15.

Apesar dos resultados adequados segundo os erros relativos de desvios-padrão, na Figura 4.25, percebe-se maior dificuldade de praticamente todos os modelos e configurações representarem adequadamente o início, o trecho final da recessão e os picos dos hidrogramas, com certo descolamento entre os hidrogramas previstos e o observado.

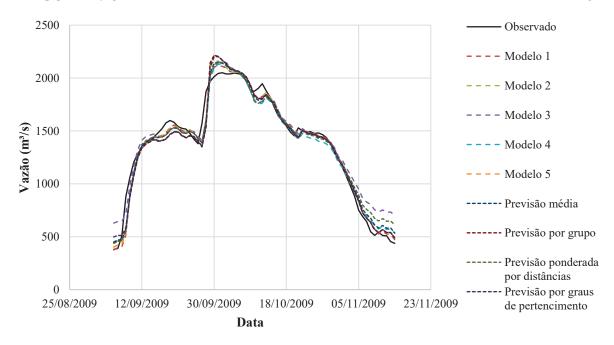


FIGURA 4.25 – HIDROGRAMAS OBSERVADO E PREVISTOS DO EVENTO ID 15

Os eventos de ID 7 e 16, por sua vez, apresentaram maiores dispersões de resultados para todas as previsões e, enquanto a previsão ponderada por graus de pertencimento apresentou melhores desempenhos para os erros relativos de vazões máximas, no quesito de desvios-padrão os valores foram distantes de zero: -6,36% (ID 7) e 7,80% (ID 16); o evento de menor pico (ID

16) apresentou desvio-padrão aumentado, enquanto o evento de maior pico (ID 7) teve seu desvio-padrão reduzido.

Na Tabela 4.53, os erros relativos dos volumes dos hidrogramas previstos indicaram que, isoladamente, os Modelos 1, 2, 4 e 5 tiveram menor oscilação (menor amplitude) de resultados entre os eventos e as menores média e mediana, enquanto as previsões por grupo de pertencimento, ponderada por graus de pertencimento, e do Modelo 3 apresentaram amplitudes, médias e medianas superiores aos patamares dos referidos modelos isolados. Por outro lado, as previsões média e ponderada por distâncias Euclidianas apresentaram amplitudes mais baixas, como os Modelos 4 e 5, mas com médias e medianas superiores.

TABELA 4.53 – DISTRIBUIÇÃO DOS ERROS RELATIVOS DOS VOLUMES ESCOADOS, COM BASE NAS VAZÕES PREVISTAS – 2005 A 2015

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Previsão média	Previsão por grupo	Previsão distância Euclidiana	Previsão Pertenc.
Mínimo	-5,13%	-2,57%	1,05%	-3,17%	-4,57%	-1,40%	-3,17%	-1,53%	-3,11%
Máximo	3,58%	5,20%	37,14%	9,22%	7,13%	11,78%	37,14%	13,40%	37,11%
Amplitude	8,71%	7,77%	36,10%	12,39%	11,70%	13,18%	40,31%	14,93%	40,22%
Média	-0,42%	0,63%	18,28%	1,91%	0,78%	4,24%	10,22%	3,63%	10,06%
Mediana	-0,56%	0,46%	18,35%	1,69%	0,44%	4,10%	3,30%	3,12%	3,67%

Segundo o Apêndice G (Tabela G.6 – erros relativos dos volumes), os eventos de ID 33, 24, 15 e 38, nesta ordem, apresentaram as menores variações entre as previsões, com todos os parâmetros atingindo intervalos de desvios de ± 3%. Este comportamento pode ser observado pelos Hidrogramas dos eventos, que, de maneira geral, preservaram os volumes de escoamento do evento observado: Figura 4.18 (ID 24), Figura 4.19 (ID 38), Figura 4.22 (ID 33) e Figura 4.25 (ID 15). Ainda, os referidos eventos (ID 33, 24, 15 e 38) são os de maiores picos.

Os eventos ID 7 e 16 apresentaram elevados valores de erros relativos dos volumes, em módulo, dentre as previsões. Porém, a previsão ponderada por graus de pertencimento de ambos os eventos se aproximou de \pm 2%, equivalente aos resultados dos Modelos 1 e 2 para os respectivos eventos. Já o evento ID 42 apresentou erro relativo de volume equivalente a -0,72% para a previsão ponderada por graus de pertencimento, valor próximo àqueles retornados pelo Modelo 4 e pela previsão por grupo de pertencimento.

5 CONCLUSÕES E RECOMENDAÇÕES

O presente trabalho tem como objetivo elaborar uma técnica de modelagem hidrológica para previsão de vazões, com base em grupos de eventos previamente definidos e da incorporação da lógica difusa no processo de modelagem, verificada por meio de quatro configurações para previsão: (i) média; (ii) por grupos de eventos; (iii) ponderada pelas distâncias Euclidianas; e (iv) ponderada pelos graus de pertencimento.

Para aplicação dos métodos propostos, inicialmente analisou-se o comportamento das cheias observadas na bacia hidrográfica do rio Iguaçu, a partir da definição das estações fluviométricas existentes a montante de União da Vitória. Dentre os postos a montante previamente definidos, três foram selecionados para definição dos modelos, após a Análise dos Componentes Principais: Fluviópolis (65220000), São Bento (65155000) e Rio Negro (65100000). Nestes postos, resguardando suas defasagens de tempo em relação à União da Vitória, foram delimitados os eventos de cheia para as análises subsequentes.

Para o período de 1964 a 2015, 194 eventos foram utilizados, a partir dos quais foram estabelecidos seis indicadores de cheia para posterior agrupamento: vazão $(\mathbf{Q_p})$ e tempo $(\mathbf{t_p})$ de pico em União da Vitória, vazão inicial $(\mathbf{Q_{t0}})$ em União da Vitória, e as vazões defasadas do tempo de pico de União da Vitória, para as estações de montante $((\mathbf{Q_{tp-1}})_{FL}, (\mathbf{Q_{tp-1}})_{SB}, (\mathbf{Q_{tp-2}})_{RN})$.

O agrupamento de cheias retornou três grupos de eventos: o Grupo 1 foi composto por 70 eventos, com vazões em União da Vitória majoritariamente mais baixas; o Grupo 2 apresentou o maior número de eventos, igual a 102; o Grupo 3 apresentou o menor número de eventos (22), abrangendo as vazões mais elevadas no ponto de interesse. Devido à lógica difusa, os eventos compartilham semelhanças com todos os grupos, e suas quantificações são definidas pelos graus de pertencimento dos eventos às classes; os resultados indicaram que determinados eventos se assemelhavam significativamente a mais de um grupo.

Neste contexto, as condições exibidas nas Figuras 4.3 a 4.5 indicam que os eventos com baixos valores de máximos graus de pertencimento aos seus grupos poderiam, em estudos futuros, ser (i) excluídos da calibração dos modelos por grupo, ou (ii) ponderados antes da execução desta etapa, de forma que os efeitos de eventos com baixos níveis de pertencimento fossem minimizados ao gerar os modelos.

Ainda, por meio do algoritmo ν -SVR – com C = 0.5 e $\gamma = 0.33$ – cinco modelos foram ajustados a partir de diferentes séries de vazões: Série Completa, série de Todos os Eventos, e séries de eventos dos Grupos 1, 2 e 3. Nesta etapa, a primeira hipótese de trabalho não se

verificou: esperava-se que cada série, ao ser simulada por seu próprio modelo, tivesse melhor desempenho comparado àqueles das simulações realizadas por outros modelos. Porém, apesar disso, os resultados das estimativas das séries pelo modelo de mesma categoria indicaram desempenhos equivalentes.

Ao se comparar todas as previsões (Modelos 1 a 5 e Configurações 1 a 4), com horizonte de um dia, para o período de 2005 a 2015, tem-se que: (i) as performances dos Modelos 1, 2 e 5 e as previsões Média e Ponderada pelas Distâncias Euclidianas se destacaram das demais, enquanto o Modelo 3 teve desempenho inferior para todas as métricas; (ii) todas as vazões mínimas foram superestimadas, enquanto as vazões máximas foram levemente subestimadas; (iii) os Modelos 1 e 2 se destacaram na previsão de vazões com diferentes tempos de permanência, enquanto as previsões das Configurações 1 a 4 apresentaram melhoras de desempenho com o aumento da magnitude de vazões. Portanto, segundo este critério, o aumento da complexidade das configurações apresentou melhora de desempenho para vazões elevadas.

Avaliando as previsões por eventos, pode-se afirmar que, em média, as Configurações tiveram desempenhos distintos, seguindo uma ordem de superioridade: previsão média, ponderada pelas distâncias Euclidianas, ponderada pelos graus de pertencimento e por grupo. Porém, visto que a definição da melhor, ou pior, configuração não se deu com base exclusivamente nos índices de performance, mas sim na combinação de índices, percentis de vazões e erros relativos das estatísticas das séries previstas, notou-se uma oscilação entre as melhores e piores configurações dentre os eventos estimados.

Ainda, é válido ressaltar que, apesar da não verificação das hipóteses de trabalho, a pesquisa retornou bons resultados para diversos modelos e configurações de previsão. Ademais, a não verificação das hipóteses pode ter origem nas premissas iniciais do trabalho, cujas alterações podem gerar mudanças nos resultados. Por exemplo: (i) os indicadores de cheia, que, a critério do pesquisador, são definidos a priori para caracterização dos eventos, bem como, (ii) ao contrário do aplicado nesta pesquisa, a não utilização de dados do período de previsão (2005 a 2015) no processo de agrupamento pode alterar a composição dos grupos, modificando todo o processo de calibração de parâmetros e de modelos; (iii) o processo de delimitação dos eventos de cheia altera os dados de entrada do agrupamento e da modelagem em si; (iv) a utilização de diferentes conjuntos de parâmetros entre os modelos por grupo poderia contribuir com a performance dos Modelos 3 e 4, favorecendo a melhoria de desempenho das Configurações.

Por outro lado, tópicos relacionados com as características do rio Iguaçu também podem ser discutidos como limitações que tenham influenciado nos resultados obtidos, como,

por exemplo: (i) as condições topográficas da bacia do rio Iguaçu, particularmente entre Fluviópolis e União da Vitória, onde o rio escoa por um canal principal limitado por margens baixas e planas, mais propensas a transbordar durante altos volumes de escoamento (SEC-CORPRERI, 1999); e (ii) a proximidade de ambas as estações fluviométricas, que requerem dados de vazão medidos em um intervalo menor que um dia.

Além disso, uma alternativa a ser explorada é o horizonte de previsão, visto que no presente estudo foram estimadas vazões com horizonte de um dia; sugere-se que o aumento do horizonte de previsão realizada por grupos, juntamente com o aumento da complexidade da configuração de previsão, seja capaz de apresentar melhor desempenho que os Modelos de Série Completa ou com Todos os Eventos, em virtude da limitação das características dos eventos em cada Grupo ou Modelo. Por outro lado, a utilização de dados de postos fluviométricos mais a montante, ou até mesmo a inclusão de chuva como dado de entrada também poderiam ampliar o horizonte de previsão.

Portanto, além da possibilidade de aplicação da abordagem proposta – que combina análise de agrupamento e modelagem hidrológica – tanto a modelos tradicionais como a modelos inteligentes orientados a dados, tem-se ainda que todas as limitações aqui mencionadas podem ser vistas como oportunidades de trabalhos futuros, tornando possível investigar a aplicabilidade dos processos descritos neste trabalho com variáveis de entrada além das vazões de escoamento, em bacias hidrográficas com menor área de drenagem e menor tempo de concentração, ou em estudos a partir de dados de entrada com escala de tempo inferior a um dia – neste contexto, quanto maior for a bacia hidrográfica em análise, menor deverá ser o intervalo de dados.

REFERÊNCIAS BIBLIOGRÁFICAS

- ADNAN, R. M.; LIANG, Z.; TRAJKOVIC, S.; ZOUNEMAT-KERMANI, M.; LI, B.; KISI, O. *Daily streamflow prediction using optimally pruned extreme learning machine*. **Journal of Hydrology**, v. 577, p. 1-11, 2019. DOI: <u>10.1016/j.jhydrol.2019.123981</u>.
- ALIZAMIR, M.; KISI, O.; ZOUNEMAT-KERMANI, M. Modelling long-term groundwater fluctuations by extreme learning machine using hydro-climatic data. **Hydrological Sciences Journal**, v. 63, n. 1, p. 63,73, 2018. DOI: 10.1080/02626667.2017.1410891.
- ALTHOFF, D.; RODRIGUES, L. N.; BAZAME, H. C. *Uncertainty quantification for hydrological models based on neural networks: the dropout ensemble.* **Stochastic Environmental Research and Risk Assessment**, v. 35, p. 1051-1067, 2021. DOI: 10.1007/s00477-021-01980-8.
- Agência Nacional de Águas ANA. Portal HidroWeb. Brasília: ANA, 2023. Disponível em: https://www.snirh.gov.br/hidroweb/serieshistoricas. Acesso em: 2023.
- ATIQUZZAMAN, M.; KANDASAMY, J. *Prediction of hydrological time-series using extreme learning machine*. **Journal of Hydroinformatics**, v. 18, n. 2, p. 345-353, 2016. DOI: 10.2166/hydro.2015.020.
- AZEVEDO, L. L. A. **Métodos estatísticos em aprendizado de máquinas para problemas de classificação**. 144 f. Dissertação (Mestrado em Estatística), Universidade de Brasília, Brasília, 2018.
- BAI, P.; LIU, X.; XIE, J. Simulating runoff under changing climatic conditions: A comparison of the long short-term memory network with two conceptual hydrologic models. **Journal of Hydrology**, 592, 1-11, 2021. DOI: 10.1016/j.jhydrol.2020.125779.
- BARZEGAR, R.; MOGHADDAM, A. A.; ADAMOWSKI, J.; FIJANI, E. *Comparison of machine learning models for predicting fluoride contamination in groundwater.* **Stochastic Environmental Research and Risk Assessment,** v. 31, n. 10, p. 2705–2718, 2017. DOI: 10.1007/s00477-016-1338-z.
- BETANCUR, S. B.; GASTMANS, D.; VÁSQUEZ, K. V.; SANTAROSA, L. V.; SANTOS, V.; KIRCHHEIM, R. E. *Hydrological responses in equatorial watersheds indicated by Principal Components Analysis (PCA) study case in Atrato River Basin (Colombia)*. **Revista Brasileira de Recursos Hídricos**, v. 25, n. 45, p. 1-12, 2020. DOI: <u>10.1590/2318-0331.252020190165</u>.
- BEZDEK, J. C.; EHRLICH, R.; FULL, W. FCM: the fuzy c-means clustering algorithm. Computers & Geosciences, v. 10, n. 2-3, p. 291–203, 1984.
- BHAVSAR, H.; PANCHAL, M. H. A review on Support Vector Machine for data classification. International Journal of Advanced Research in Computer Engineering & Technology, v. 1, n. 10, 2012.

- BRÊDA, J. P. L. F.; PAIVA, R. C. D.; PEDROLLO, O. C.; PASSAIA, O. A.; COLLISCHONN, W. Modeling coordinated operation of multiple hydropower reservoirs at a continental scale using artificial neural network: the case of Brazilian hydropower system. **Revista Brasileira de Recursos Hídricos**, v. 26, n. 12, p. 1-12, 2021. DOI: <u>10.1590/2318-0331.262120210011</u>.
- CHANG, C.-C.; LIN, C.-J. *Training v-support vector classifiers: Theory and algorithms*. **Neural Computation**, v. 13, n. 9, p. 2119–2147, 2001.
- CHANG, C.-C.; LIN, C.-J. *LIBSVM: a library for Support Vector Machines*. 2022. Disponível em: https://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf.
- CHEN, S.; REN, M. SUN, W. Combining two-stage decomposition based machine learning methods for annual runoff forecasting. **Journal of Hydrology**, v. 603, p. 1-18, 2021. DOI: 10.1016/j.jhydrol.2021.126945.
- CHU, H.; WU, W.; WANG, Q. J.; NATHAN, R.; WEI, J. An ANN-based emulation modelling framework for flood inundation modelling: Application, challenges and future directions. **Environmental Modelling and Software,** v. 124, p. 1-17, 2020. DOI: 10.1016/j.envsoft.2019.104587.
- COBRAPE COMPANHIA BRASILEIRA DE PROJETOS E EMPREENDIMENTOS. **Produto 1.2 Parte D Levantamento e análise de eventos críticos**. Curitiba: 20--. Revisão final.
- CRISTIANINI, N.; SHAWE-TAYLOR, J. An introduction to Support Vector Machines and other Kernel-based learning methods. New York: Cambridge University Press, 2000.
- DAS, J.; MANIKANTA, V.; TEJA, K. N.; UMAMAHESH, N. V. Two decades of ensemble flood forecasting: a state-of-the-art on past developments, present applications and future opportunities. **Hydrological Sciences Journal,** v. 67, n. 3, p. 477-493, 2022. DOI: 10.1080/02626667.2021.2023157.
- DESAI, S.; OUARDA, T. B. M. J. Regional hydrological frequency analysis at ungauged sites with random forest regression. **Journal of Hydrology**, v. 594, p. 1-8, 2021. DOI: 10.1016/j.jhydrol.2020.125861.
- DETZEL, D. H. M. **Modelagem de séries hidrológicas: uma abordagem de múltiplas escalas temporais**. 218 f. Tese (Doutorado em Engenharia de Recursos Hídricos e Ambiental), Universidade Federal do Paraná, Curitiba, 2015.
- DIFI, S.; ELMEDDAHI, Y.; HEBAL, A.; SINGH, V. P.; HEDDAM, S.; KIM, S.; KISI, O. *Monthly streamflow prediction using hybrid extreme learning machine optimized by bat algorithm: a case study of Cheliff watershed, Algeria.* **Hydrological Sciences Journal,** v. 68, n. 2, p. 189-208, 2023. DOI: 10.1080/02626667.2022.2149334.
- DING, Y.; ZHU, Y.; FENG, J.; ZHANG, P.; CHENG, Z. *Interpretable spatio-temporal attention LSTM model for flood forecasting*. **Neurocomputing**, v. 403, p. 348-359, 2020. DOI: 10.1016/j.neucom.2020.04.110.

- EBTEHAJ, I.; BONAKDARI, H. A reliable hybrid outlier robust non-tuned rapid machine learning model for multi-step ahead flood forecasting in Quebec, Canada. **Journal of Hydrology**, v. 614, p. 1-21, 2022. DOI: 10.1016/j.jhydrol.2022.128592.
- EDDINE, Y. T.; NADIR, M.; SABAH, S.; JAAFARI, A. *Integrating Support Vector Machines with different ensemble learners for improving streamflow simulation in an ungauged watershed.* **Water Resources Management**, v. 38, p. 553-567, 2024. DOI: <u>10.1007/s11269-023-03684-w</u>.
- EM-DAT EMERGENCY DISASTERS DATA BASE. Disponível em: http://www.emdat.be. Acesso em: 22 nov. 2015.
- EZUGWU, A. E.; IKOTUN, A. M.; OYELADE, O. O.; ABUALIGAH, L.; AGUSHAKA, J. O.; EKE, C. I.; AKINYELU, A. A. A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. **Engineering Applications of Artificial Intelligence,** v. 110, p. 1-43, 2022. DOI: 10.1016/j.engappai.2022.104743.
- FAN, F. M.; COLLISCHONN, W.; MELLER, A.; BOTELHO, L. C. M. Ensemble streamflow forecasting experiments in a tropical basin: The São Francisco river case study. **Journal of Hydrology**, v. 519, p. 2906-2919, 2014. DOI: 10.1016/j.jhydrol.2014.04.038.
- FAN, F. M.; PAIVA, R. C. D.; COLLISCHONN, W. *Hydrological forecasting practices in Brazil*. In: ADAMS III, T. E.; PAGANO, T. C. (Ed.). **Flood Forecasting: a global perspective.** Elsevier, 2016. p. 249-310.
- FATEHIFAR, A.; GOODARZI, M. R.; HEDESH, S. S. M.; DASTJERDI, P. S. *Assessing watershed hydrological response to climate change based on signature indices.* **Journal of Water and Climate Change**, v. 12, n. 6, p. 2579-2593, 2021. DOI: 10.2166/wcc.2021.293.
- FATHIAN, F.; MEHDIZADEH, S.; SALES, A. K.; SAFARI, M. J. S. *Hybrid models to improve the monthly river flow prediction: Integrating artificial intelligence and non-linear time series models, and future research prospects.* **Journal of Hydrology,** v. 575, p. 1200–1213, 2019. DOI: 10.1016/j.jhydrol.2019.06.025.
- FENG, Z.; SHI, P.; YANG, T.; NIU, W.; ZHOU, J.; CHENG, C. *Parallel cooperation search algorithm and artificial intelligence method for streamflow time series forecasting.* **Journal of Hydrology**, v. 606, p. 1–14, 2022. DOI: <u>10.1016/j.jhydrol.2022.127434</u>.
- FISCHER, S.; SCHUMANN, A. H. *Handling the stochastic uncertainty of flood statistics in regionalization approaches.* **Hydrological Sciences Journal**, v. 67, n. 10, p. 1449-1465, 2022. DOI: 10.1080/02626667.2022.2091410.
- FREITAS, C. Uso de técnicas de classificação de dados na operação de sistemas de reservatórios de usinas hidrelétricas durante cheias. 260 f. Tese (Doutorado em Engenharia de Recursos Hídricos e Ambiental), Universidade Federal do Paraná, Curitiba, 2016.
- GHORBANI, M. A.; KHATIBI, R.; MEHR, A. D.; ASADI, A. *Chaos-based multigene genetic programming: A new hybrid strategy for river flow forecasting*. **Journal of Hydrology**, v. 562, p. 455 467, 2018. DOI: <u>10.1016/j.jhydrol.2018.04.054</u>.

- GIZAW, M. S.; GAN, T. Y. Regional Flood Frequency Analysis using Support Vector Regression under historical and future climate. **Journal of Hydrology**, v. 538, p. 387–398, 2016. DOI: 10.1016/j.jhydrol.2016.04.041.
- GUO, J.; ZHOU, J.; QIN, H.; ZOU, Q.; LI, Q. *Monthly streamflow forecasting based on improved support vector machine model.* **Expert Systems with Applications,** v. 38, p. 13073–13081, 2011. DOI: <u>10.1016/j.eswa.2011.04.114</u>.
- GUPTA, H. V.; KLING, H.; YILMAZ, K. K.; MARTINEZ, G. F. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology, v. 377, p. 80-91, 2009. DOI: 10.1016/j.jhydrol.2009.08.003.
- HAIR JUNIOR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E.; TATHAM, R. L. Análise multivariada de dados. Tradução: Adonai Schlup Sant'anna. 6. Ed. Porto Alegre: Bookman, 2009.
- HALLOUIN, T.; BRUEN, M.; O'LOUGHLIN, F. E. Calibration of hydrological models for ecologically relevant streamflow predictions: a trade-off between fitting well to data and estimating consistent parameter sets?. **Hydrology and Earth System Science**, v. 24, p. 1031-1054, 2020. DOI: 10.5194/hess-24-1031-2020.
- HAMMER, B.; VILLMANN, T. How to process uncertainty in machine learning? *In:* European Symposium on Artificial Neural Networks, 2007, Bélgica.
- HAN, J.; KAMBER, M.; PEI, J. **Data mining: concepts and techniques.** Estados Unidos: Elsevier, 2012.
- HUANG, G-B.; ZHU, Q-Y.; SIEW, C-K. Extreme Learning Machine: a new learning scheme of Feedforward Neural Networks. **IEEE International Joint Conference on Neural Networks**, p. 985–990. DOI: <u>10.1109/IJCNN.2004.1380068</u>.
- HSU, C.-W.; CHANG, C.-C.; LIN, C.-J. *A practical guide to Support Vector Classification*. 2016. Disponível em: http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf. Acessado em: 05 jun. 2021.
- HÜLLERMEIER, E. *Does machine learning need fuzzy logic?*. **Fuzzy Sets and Systems,** v. 281, p. 292-299, 2015. DOI: <u>10.1016/j.fss.2015.09.001</u>.
- IAT Instituto Água e Terra (2020a). **Uso e cobertura da terra no Estado do Paraná**. Disponível em: < http://www.iat.pr.gov.br/>. Acesso em: 31 mai. 2021.
- IAT Instituto Água e Terra (2020b). **Uso cobertura terra**. Disponível em: https://geo.iat.pr.gov.br/portal/home/>. Acesso em: 31 mai. 2021.
- IAT Instituto Água e Terra (2023). Acesso em: 01 fev. 2023. https://geoem.geo.pr.gov.br/geoem/pages/templates/initial-public.jsf?windowid=b41
- IBGE Instituto Brasileiro de Geografia e Estatística (2022). Disponível em: http://www.ibge.gov.br/home>. Acesso em: 31 jan. 2024.
- IBRAHIM, K. S. M. H.; HUANG, Y. F.; AHMED, A. N.; KOO, C. H.; EL-SHAFIE, A. A review of the hybrid artificial intelligence and optimization modelling of hydrological

- streamflow forecasting. Alexandria Engineering Journal, v. 61, n. 1, p. 279-303, 2022. DOI: 10.1016/j.aej.2021.04.100.
- ISLAM, A. R. M. T.; TALUKDAR, S.; MAHATO, S.; KUNDU, S.; EIBEK, K. U.; PHAM, Q. B.; KURIQI, A.; LINH, N. T. T. *Flood susceptibility modelling using advanced ensemble machine learning models.* **Geoscience Frontiers**, v. 12, p. 1-18, 2021. DOI: 10.1016/j.gsf.2020.09.006.
- JOO, H.; LEE, M.; KIM, J.; JUNG, J.; KWAK, J.; KIM, H. S. Stream gauge network grouping analysis using community detection. **Stochastic Environmental Research and Risk Assessment**, 35, 781-795, 2021. DOI: 10.1007/S00477-020-01916-8.
- KAVISKI, E. **Métodos de regionalização de eventos e parâmetros hidrológicos**. Dissertação (Mestrado em Engenharia Hidráulica), Centro de Hidráulica e Hidrologia Professor Parigot de Souza (CEHPAR), Universidade Federal do Paraná, Curitiba, 1992.
- KHOSRAVI, K.; GOLKARIAN, A.; BOOIJ, M. J.; BARZEGAR, R.; SUN, W.; YASEEN, Z. M.; MOSAVI, A. *Improving daily stochastic streamflow prediction: comparison of novel hybrid data-mining algorithms.* **Hydrological Sciences Journal,** v. 66, n. 9, p. 1457-1474, 2021. DOI: 10.1080/02626667.2021.1928673.
- KIM, D.; LEE, J.; KIM, J.; LEE, M.; WANG, W.; KIM, H. S. Comparative analysis of long short-term memory and storage function model for flood water level forecasting of Bokha stream in NamHan River, Korea. **Journal of Hydrology**, v. 606, p. 1-13, 2022. DOI: 10.1016/j.jhydrol.2021.127415.
- KURIAN, C.; SUDHEER, K. P.; VEMA, V. K.; SAHOO, D. *Effective flood forecasting at higher lead times through hybrid modelling framework.* **Journal of Hydrology,** v. 587, p. 1-13, 2020. DOI: 10.1016/j.jhydrol.2020.124945.
- LAPPICY, T.; LIMA, C. H. R. Enhancing monthly streamflow forecasting for Brazilian hydropower plants through climate index integration with stochastic methods. **Revista Brasileira de Recursos Hídricos**, v. 28, n. 48, p. 1-7, 2023. DOI: <u>10.1590/2318-0331.282320230118</u>.
- LATTIN, J.; CARROLL, J. D.; GREEN, P. E. Análise de dados multivariados. São Paulo: Cengage Learning, 2011.
- LAWIN, A. E.; HOUNGUE, R.; PO, Y.; HOUNGUE, N. R.; ATTOGOUINON, A.; AFOUDA, A. A. *Mid-century climate change impacts on Ouémé River discharge at Bonou Outlet (Benin)*. **Hydrology**, v. 6, n. 72, p. 1-20, 2019. DOI: <u>10.3390/hydrology6030072</u>.
- LETESSIER, C.; CARDI, J.; DUSSEL, A.; EBTEHAJ, I.; BONAKDARI, H. Enhancing flood prediction accuracy through integration of meteorological parameters in river flow observations: a case study Ottawa River. **Hydrology**, v. 10, n. 164, p. 1-23, 2023. DOI: 10.3390/hydrology10080164.
- LI, X.; SHA, J.; WANG, Z-L. Comparison of daily streamflow forecasts using extreme learning machines and the random forest method. **Hydrological Sciences Journal,** v. 64, n. 15, p. 1857-1866, 2019. DOI: <u>10.1080/02626667.2019.1680846</u>.

- LIANG, Z.; LI, Y.; HU, Y.; LI, B.; WANG, J. A data-driven SVR model for long-term runoff prediction and uncertainty analysis based on the Bayesian framework. **Theoretical and Applied Climatology,** v. 133, p. 137-149, 2018. DOI: 10.1007/s00704-017-2186-6.
- LIANG, Z.; XIAO, Z.; WANG, J.; SUN, L.; LI, B.; HU, Y.; WU, Y. *An improved chaos similarity model for hydrological forecasting.* **Journal of Hydrology,** v. 577, 2019. DOI: 10.1016/j.jhydrol.2019.123953.
- LIN, J.-Y.; CHENG, C.-T.; CHAU, K.-W. *Using support vector machines for long-term discharge prediciton.* **Hydrological Sciences Journal,** v. 51, n. 4, p. 599-612, 2006. DOI: 10.1623/hysj.51.4.599.
- LIMA, G. R. T.; SCOFIELD, G. B. Feasibility study on operational use of neural networks in a flash flood early warning system. **Revista Brasileira de Recursos Hídricos**, 26(7), 2-11, 2021. DOI: 10.1590/2318-0331.262120200152.
- LIU, J.; ZHANG, Q.; SINGH, V. P.; GU, X.; SHI, P. Nonstationarity and clustering of flood characteristics and relations with the climate indices in the Poyang Lake basin, China. **Hydrological Sciences Journal,** v. 62, n. 11, p. 1809-1824, 2017. DOI: 10.1080/02626667.2017.1349909.
- LOHANI, A. K.; GOEL, N. K.; BHATIA, K. K. *Comparative study of neural network, fuzzy logic and linear transfer function techniques in daily rainfall-runoff modelling under different input domains*. **Hydrological Processes**, v. 25, p. 175–193, 2011. DOI: <u>10.1002/hyp.7831</u>.
- LOHANI A.K., RAKESH KUMAR, SINGH, R.D. *Hydrological time series modeling: a comparison between adaptive neuro fuzzy*. **Journal of Hydrology,** v. 442–443, p. 1–13, 2012. DOI: <u>10.1016/j.jhydrol.2012.03.031</u>.
- LORENA, A. C.; CARVALHO, A. C. P. L. F. Uma introdução às Support Vector Machines. **Revista de Informática Teórica e Aplicada,** v. 14, n. 2, 2007. DOI: <u>10.22456/2175-</u>2745.5690.
- LUPPICHINI, M.; FAVALLI, M.; ISOLA, I.; NANNIPIERI, L.; GIANNECCHINI, R.; BINI, M. Influence of Topographic Resolution and Accuracy on Hydraulic Channel Flow Simulations: Case Study of the Versilia River (Italy). **Remote Sensing**, 11(1630), 1-22, 2019. DOI: 10.3390/rs11131630.
- LUPPICHINI, M.; BARSANTI, M.; GIANNECCHINI, R.; BINI, M. Deep learning models to predict flood events in fast-flowing watersheds. **Science of the Total Environment**, 813, 1-10, 2022. DOI: 10.1016/j.scitotenv.2021.151885.
- MAHDAVI-MEYMAND, A.; SULISZ, W.; ZOUNEMAT-KERMANI, M. Hybrid and integrative evolutionary Machine Learning in hydrology: a systematic review and meta-analysis. **Archives of Computational Methods in Engineering**, 2023. DOI: 10.1007/s11831-023-10017-y.
- MELLER, A. **Previsão de cheias por conjunto em curto prazo**. 224 f. Tese (Doutorado em Recursos Hídricos e Saneamento Ambiental), Universidade Federal do Rio Grande do Sul, Porto Alegre, 2012.

- MIHAILOVIĆ, D. T.; NIKOLIĆ-ĐORIĆ, E.; ARSENIĆ, I.; MALINOVIĆ-MILIĆEVIĆ, S.; SINGH, V. P.; STOŠIĆ, T.; STOŠIĆ, B. *Analysis of daily streamflow complexity by Kolmogorov measures and Lyapunov exponent.* **Physica A**, v. 525, p. 290-303, 2019. DOI: 10.1016/j.physa.2019.03.041.
- MINE, M. R. M.; TUCCI, C. E. M. Gerenciamento da produção de energia e controle de inundação: Foz de Areia no rio Iguaçu. **Revista Brasileira de Recursos Hídricos**, v. 7, n. 3, p. 85-107, 2002. DOI: <u>10.21168/rbrh.v7n3.p85-107</u>.
- MINGOTI, S. A. Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada. Belo Horizonte: Editora UFMG, 2005.
- MOEENI, H.; BONAKDARI, H.; EBTEHAJ, I. *Integrated SARIMA with Neuro-Fuzzy Systems and Neural Networks for Monthly Inflow Prediction.* **Water Resources Management,** v. 31, p. 2141-2156, 2017. DOI: 10.1007/s11269-017-1632-7.
- MOHAMMADI, B.; LINH, N. T. T.; PHAM, Q. B.; AHMED, A. N.; VOJTEKOVÁ, J.; GUAN, Y.; ABBA, S. I.; EL-SHAFIE, A. *Adaptive neuro-fuzzy inference system coupled with shuffled frog leaping algorithm for predicting river streamflow time series.* **Hydrological Sciences Journal,** v. 65, n. 10, p. 1738-1751, 2020. DOI: 10.1080/02626667.2020.1758703.
- MORETTIN, P. A.; SINGER, J. M. Estatística e ciência de dados. Rio de Janeiro: LTC, 2023.
- MOSAVI, A.; GOLSHAN, M.; CHOUBIN, B.; ZIEGLER, A. D.; SIGAROODI, S. K.; ZHANG, F.; DINEVA, A. A. *Fuzzy clustering and distributed model for streamflow estimation in ungauged watersheds.* **Scientific Reports**, v. 11, n. 8243, 2021. DOI: <u>10.1038/s41598-021-87691-0</u>.
- MOURA, M. C.; LINS, I. D.; VELEDA, D.; DROGUETT, E. L.; ARAUJO, M. Sea level prediction by Support Vector Machines combined with Particle Swarm Optimization. *In:* 10th International Probabilistic Safety Assessment & Management Conference, 2010, Seatle.
- MOURA, C. N.; JAN, S.; DETZEL, D. H. M. Evaluating the long short-term memory (LSTM) network for discharge prediction under changing climate conditions. **Hydrology Research**, v. 53, n. 5, p. 657-667, 2022. DOI: 10.2166/nh.2022.044.
- NIU, W.; FENG, Z. Evaluating the performances of several artificial intelligence methods in forecasting daily streamflow time series for sustainable water resources management. **Sustainable Cities and Society,** v. 64, p. 1-12, 2021 DOI: 10.1016/j.scs.2020.102562.
- PINTO, N. L. S.; HOLTZ, A. C. T.; MARTINS, J. A.; GOMIDE, F. L. S. Hidrologia Básica. São Paulo: Blucher, 1976.
- REN, M.; LIU, P.; WANG, Z.; YI, J. A Self-Adaptive Fuzzy c-Means Algorithm for Determining the Optimal Number of Clusters. Computational Intelligence and Neuroscience, p. 1-12, 2016. DOI: 10.1155/2016/2647389.
- RIBEIRO, V. H. A.; REYNOSO-MEZA, G.; SIQUEIRA, H. V. Multi-objective ensembles of echo state networks and extreme learning machines for streamflow series forecasting. **Engineering Applications of Artificial Intelligence**, v. 95, p. 1-19, 2020. DOI: 10.1016/j.engappai.2020.103910.

- ROCHA, P. S. M. **Gestão em áreas de risco de enchentes: estudo de caso para União da Vitória Paraná**. 134 f. Dissertação (Mestrado em Gestão Ambiental). Universidade Positivo, Curitiba, 2012.
- ROSS, T. J. Fuzzy logic with engineering applications. England: John Wiley & Sons Ltd, 2004.
- ROY, A.; CHAKRABORTY, S. Support Vector Machine in structural reliability analysis: A review. **Reliability Engineering and System Safety**, 233, 2023. DOI: 10.1016/j.ress.2023.109126.
- SAADI, M.; OUDIN, L.; RIBSTEIN, P. Random Forest ability in regionalizing hourly hydrological model parameters. **Water**, v. 11, n. 1540, p. 1-22. DOI: <u>10.3390/w11081540</u>.
- SAMANTARAY, S.; SAHOO, P.; SAHOO, A.; SATAPATHY, D. P. (2023). Flood discharge prediction using improved ANFIS model combined with hybrid particle swarm optimisation and slime mould algorithm. **Environmental Science and Pollution Research**, v. 30, p. 83845-83872, 2023a. DOI: 10.1007/s11356-023-27844-y.
- SAMANTARAY, S.; SAHOO, A.; AGNIHOTRI, A. *Prediction of flood discharge using hybrid PSO-SVM algorithm in Barak River Basin.* **MethodsX**, v. 10, p. 1-21, 2023b. DOI: 10.1016/j.mex.2023.102060.
- SATO-ILIC, M.; JAIN, L. C. Innovations in fuzzy clustering: theory and applications. Springer, 2006.
- SCHOLKOPF, B.; SMOLA, A. J. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. London: The MIT Press, 2002.
- SCHOPPA, L.; DISSE, M.; BACHMAIR, S. *Evaluating the performance of random forest for large-scale flood discharge simulation.* **Journal of Hydrology**, v. 590, p. 1-13, 2020. DOI: 10.1016/j.jhydrol.2020.125531.
- SEMA SECRETARIA DE ESTADO DO MEIO AMBIENTE E RECURSOS HÍDRICOS. **Bacias hidrográficas do Paraná.** PEREIRA, M. C. B., SCROCCARO, J. L. (Orgs.), Curitiba, 2010.
- SHALEV-SHWARTZ, S.; BEN-DAVID, S. Understanding Machine Learning: from theory to algorithms. New York: Cambridge University Press, 2014.
- SHARMA, R. K.; KUMAR, S.; PADMALAL, D.; ROY, A. Streamflow prediction using machine learning models in selected rivers of Southern India. International Journal of River Basin Management, 2023. DOI: 10.1080/15715124.2023.2196635.
- SHRESTHA, M. S.; ARTAN, G. A.; BAJRACHARYA, S. R.; SHARMA, R. R. *Using satellite-based rainfall estimates for streamflow modelling: Bagmati Basin.* **Journal of Flood Risk Management,** v. 1, p. 89-99, 2008. DOI: 10.1111/j.1753-318x.2008.00011.x.
- SHUKLA, R.; KUMAR, P.; VISHWAKARMA, D. K.; ALI, R.; KUMAR, R.; KURIQI, A. *Modeling of stage-discharge using back propagation ANN-, ANFIS-, and WANN-based computing techniques.* **Theoretical and Applied Climatology**, v. 147, p. 867-889, 2022. DOI: 10.1007/s00704-021-03863-y.

- SILVA FILHO, T. M.; PIMENTEL, B. A.; SOUZA, R. M. C. R.; OLIVEIRA, A. L. I. Hybrid methods for fuzzy clustering based on fuzzy c-means and improved particle swarm optimization. **Expert Systems with Applications**, v. 42, p. 6315-6328, 2015. DOI: 10.1016/j.eswa.2015.04.032.
- SINGH, V. P. *Hydrologic modeling: progress and future directions*. **Geoscience Letters**, v. 5, n. 15, p. 1-18, 2018. DOI: <u>10.1186/s40562-018-0113-z</u>.
- SINGH, L.; SARAVANAN, S. Assessing streamflow modeling using single and multi-site calibration approach on Bharathpuzha catchment, India: a case study. **Modeling Earth Systems and Environment,** v. 8, p. 4135-4148, 2022. DOI: <u>10.1007/s40808-022-01353-2</u>.
- SIVAKUMAR, B. Chaos theory in hydrology: important issues and interpretations. **Journal of Hydrology**, v. 227, n. 1-4, p. 1-20, 2000. DOI: 10.1016/S0022-1694(99)00186-9.
- SIVAKUMAR, B.; SINGH, V. P. Hydrologic system complexity and nonlinear dynamic concepts for a catchment classification framework. **Hydrology and Earth System Sciences**, v. 16, p. 4119 4131, 2012. DOI: 10.5194/hess-16-4119-2012.
- SIVAKUMAR, B. Chaos in Hydrology: bridging determinism and stochasticity. 1 ed. Australia: Springer, 2017.
- SMOLA, A. J.; SCHOLKOP, B. A tutorial on support vector regression. **Statistics and Computing**, v. 14, p. 199-222, 2004. DOI: <u>10.1023/b:stco.0000035301.49549.88</u>.
- SNIEDER, E.; SHAKIR, R.; KHAN, U. T. *A comprehensive comparison of four input variable selection methods for artificial neural network flow forecasting models*. **Journal of Hydrology**, v. 583, 2020. DOI: <u>10.1016/j.jhydrol.2019.124299</u>.
- SOLOMATINE, D. P.; SEE, L. M.; ABRAHART, R. J. *Data-driven modelling: concepts, approaches and experiences. In*: ABRAHART, R. J.; SEE, L. M.; SOLOMATINE, D. P. (Ed.). **Practical hydroinformatics: computational intelligence and technological developments in water applications.** Springer, 2008. p. 17-30.
- STAUDINGER, M.; STAHL, K.; SEIBERT, J.; CLARK, M. P.; TALLAKSEN, L. M. *Comparison of hydrological model structures based on recession and low flow simulations.* **Hydrology and Earth System Sciences**, v. 15, n. 11, p. 3447-3459, 2011. DOI: <u>10.5194/hess-15-3447-2011</u>.
- STEFFEN, P. C.; GOMES, J. Clustering of historical floods observed on Iguaçu River, in União da Vitória, Paraná. Revista Brasileira de Recursos Hídricos, v. 23, n. 38, p. 1–12, 2018. DOI: 10.1590/2318-0331.231820170107.
- SUN, Y.; BABOVIC, V.; CHAN, E. S. *Multi-step-ahead model error prediction using time-delay neural networks combined with chaos theory.* **Journal of Hydrology,** v. 395, p. 109-116, 2010. DOI: <u>10.1016/j.jhydrol.2010.10.020</u>.
- TAO, C. W. *Unsupervised fuzzy clustering with multi-center clusters*. **Fuzzy Sets and Systems**, v. 128, p. 305-322, 2002.
- TARASOVA, L.; MERZ, R.; KISS, A.; BASSO, S.; BLÖSCHL, G.; MERZ, B.; VIGLIONE, A.; PLÖTNER, S.; GUSE, B.; SCHUMANN, A.; FISCHER, S.; AHRENS, B.; ANWAR, F.;

- BÁRDOSSY, A.; BÜHLER, P.; HABERLANDT, U.; KREIBICH, H.; KRUG, A.; LUN, D.; MÜLLER-THOMY, H.; PIDOTO, R.; PRIMO, C.; SEIDEL, J.; VOROGUSHYN, S.; WIETZKE, L. *Causative classification of river flood events*. **Water**, v. 6, p. 1-23, 2019. DOI: 10.1002/wat2.1353.
- TOMAR, D.; AGARWAL, S. Twin Support Vector Machine: A review from 2007 to 2014. Egyptian Informatics Journal, v. 16, p. 55-69, 2015. DOI: 10.1016/j.eij.2014.12.003.
- TONGAL, H. Nonlinear forecasting of stream flow using a chaotic approach and artificial neural networks. Earth Sciences Research Journal, v. 17, n. 2, p. 119 126, 2013.
- TOZZI, B. K. M.; FILL, H. D. O. A. Verification of the stationarity of flow series in the Iguaçu River basin. Revista Brasileira de Recursos Hídricos, v. 25, n. 10, p. 1-9, 2020. DOI: 10.1590/2318-0331.252020180171.
- TUCCI, C. E. M. Modelos Hidrológicos. 2. ed. Porto Alegre: Editora da UFRGS/ABRH, 2005.
- TUCCI, C. E. M. **Hidrologia: ciência e aplicação**. 4. ed. Porto Alegre: Editora da UFRGS/ABRH, 2012.
- TURAN, M. E.; YURDUSEV, M. L. River flow estimation from upstream flow records by artificial intelligence methods. **Journal of Hydrology,** v. 369, p. 71-77, 2009. DOI: 10.1016/j.jhydrol.2009.02.004.
- UNDUCHE, F.; HABTAMU, T.; SENBETA, D.; ZHU, E. Evaluation of four hydrological models for operational flood forecasting in a Canadian Prairie watershed. **Hydrological Sciences Journal**, v. 63, n. 8, p. 1133-1149, 2018. DOI: 10.1080/02626667.2018.1474219.
- WANG, L. N.; CHEN, X. H.; SHAO, Q. X.; LI, Y. *Flood indicators and their clustering features in Wujiang River, South China*. **Ecological Engineering**, v. 76, p. 66-74, 2015. DOI: 10.1016/j.ecoleng.2014.03.018.
- WANG, X.; WANG, Y.; YUAN, P.; WANG, L.; CHENG, D. *An adaptive daily runoff forecast model using VMD-LSTM-PSO hybrid approach.* **Hydrological Sciences Journal**, v. 66, n. 9, p. 1488-1502, 2021. DOI: <u>10.1080/02626667.2021.1937631</u>.
- WANG, J.; LU, S.; WANG, S-H.; ZHANG, Y-D. *A review on extreme learning machine.* **Multimedia Tools and Applications,** v. 81, p. 41611-41660, 2022. DOI: <u>10.1007/s11042-021-11007-7</u>.
- WIJAYARATHNE, D. B.; COULIBALY, P. *Identification of hydrological models for operational flood forecasting in St. John's, Newfoundland, Canada.* **Journal of Hydrology: Regional Studies,** v. 27, p. 1-16, 2020. DOI: <u>10.1016/j.ejrh.2019.100646</u>.
- WILLIAMS, G. P. Chaos Theory Tamed. Washington: Joseph Henry Press, 1997.
- WU, J.; LIU, H.; WEI, G.; SONG, T.; ZHANG, C.; ZHOU, H. Flash flood forecasting using Support Vector Regression model in a small mountainous catchment. **Water**, v. 11, n. 1327, p. 1-16, 2019. DOI: 10.3390/w11071327.

- YANG, M-S.; NATALIANI, Y. Robust-learning fuzzy c-means clustering algorithm with unknown number of clusters. **Pattern Recognition**, v. 71, p. 45-59, 2017. DOI: 10.1016/j.patcog.2017.05.017.
- YASEEN, Z. M.; JAAFAR, O.; DEO, R. C.; KISI, O.; ADAMOWSKI, J.; QUILTY, J.; EL-SHAFIE, A. *Stream-flow forecasting using extreme learning machines: A case study in a semi-arid region in Iraq.* **Journal of Hydrology**, v. 542, p. 603-614, 2016. DOI: 10.1016/j.jhydrol.2016.09.035.
- YASEEN, Z. M.; AWADH, S. M.; SHARAFATI, A.; SHAHID, S. *Complementary data-intelligence model for river flow simulation*. **Journal of Hydrology**, v. 567, p. 180-190, 2018a. DOI: 10.1016/j.jhydrol.2018.10.020.
- YASEEN, Z. M.; ALLAWI, M. F.; YOUSIF, A. A.; JAAFAR, O.; HAMZAH, F. M.; EL-SHAFIE, A. *Non-tuned machine learning approach for hydrological time series forecasting.* **Neural Computing and Applications**, v. 30, p. 1479-1491, 2018b. DOI: <u>10.1007/s00521-016-</u>2763-0.
- YASEEN, Z. M.; FARIS, H.; AL-ANSARI, N. *Hybridized Extreme Learning Machine Model with Salp Swarm Algorithm: a novel predictive model for hydrological application.* Complexity, p. 1-14, 2020. DOI: 10.1155/2020/8206245.
- YASMIN, N.; SIVAKUMAR, B. *Temporal streamflow analysis: Coupling nonlinear dynamics with complex networks.* **Journal of Hydrology,** v. 564, p. 59 67, 2018. DOI: 10.1016/j.jhydrol.2018.06.072.
- YILMAZ, K. K.; GUPTA, H. V.; WAGENER, T. A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. Water Resources Research, v. 44, 2008. DOI: 10.1029/2007WR006716.
- YU, X.; LIONG, S. Y.; BABOVIC, V. *EC-SVM approach for real-time hydrologic forecasting*. **Journal of Hydroinformatics**, v. 6, n. 3, 2004. DOI: 10.2166/hydro.2004.0016.
- YU, P.-S.; CHEN, S.-T.; CHANG, I.-F. Real-time flood stage forecasting using Support Vector Regression. *In*: ABRAHART, R. J.; SEE, L. M.; SOLOMATINE, D. P. (Ed.). **Practical hydroinformatics: computational intelligence and technological developments in water applications.** Springer, 2008. p. 359-373.
- ZAKHROUF, M.; HAMID, B.; KIM, S.; MADANI, S. *Novel insights for streamflow forecasting based on deep learning models combined the evolutionary optimization algorithm.* **Physical Geography,** v. 44, n. 1, p. 31-54, 2023. DOI: <u>10.1080/02723646.2021.1943126</u>.
- ZHANG, X.; PENG, Y.; ZHANG, C.; WANG, B. *Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting? Some experiment evidences.* **Journal of Hydrology,** v. 530, p. 137-152, 2015. DOI: 10.1016/j.jhydrol.2015.09.047.
- ZHANG, Z.; ZHANG, Q.; SINGH, V. P. *Univariate streamflow forecasting using commonly used data-driven models: literature review and case study.* **Hydrological Sciences Journal,** v. 63, n. 7, p. 1091-1111, 2018. DOI: <u>10.1080/02626667.2018.1469756</u>.

ZHU, S.; ZHOU, J.; YE, L.; MENG, C. Streamflow estimation by support vector machine coupled with different methods of time series decomposition in the upper reaches of Yangtze River. China. **Environmental Earth Sciences,** v. 75, n. 531, p. 1-12, 2016. DOI: 10.1007/s12665-016-5337-7.

APÊNDICE A – EVENTOS AGRUPADOS

TABELA A.1 – AGRUPAMENTO DE EVENTOS

Grupo 1			Grupo 2	Grupo 3		
07/05/1964	26/05/2005	31/08/1964	01/08/1990	13/01/1971		
28/06/1964	20/06/2005	03/10/1965	08/09/1990	01/01/1981		
16/05/1965	26/02/2007	17/11/1965	28/12/1991	27/11/1982		
15/07/1965	28/04/2007	22/12/1965	31/03/1992	18/07/1983		
24/08/1965	18/10/2007	23/02/1966	03/07/1992	28/05/1987		
07/09/1966	12/12/2007	01/07/1966	26/02/1993	27/01/1990		
30/10/1966	31/10/2008	23/12/1966	22/07/1993	08/06/1992		
27/01/1968	04/08/2009	06/03/1967	26/02/1994	06/10/1993		
02/11/1968	21/07/2010	03/04/1967	08/07/1994	22/01/1995		
08/04/1969	01/05/2012	05/12/1967	21/11/1994	04/02/1997		
22/06/1969	10/01/2013	25/11/1969	02/02/1996	02/05/1998		
11/10/1969	03/06/2013	22/10/1970	16/08/1996	12/10/1998		
08/07/1970	24/09/2013	28/03/1971	10/09/1996	14/07/1999		
15/10/1971	04/09/2014	10/06/1971	21/10/1996	25/09/2000		
29/02/1972	09/11/2014	09/12/1972	05/08/1997	11/10/2001		
30/08/1972	04/04/2015	28/06/1973	27/08/1997	20/09/2005		
23/04/1973	31/05/2015	24/07/1973	10/07/1998	03/10/2009		
14/08/1973	20/06/2015	31/08/1973	12/12/1998	08/05/2010		
28/01/1974	24/10/2015	25/03/1974	03/03/1999	25/02/2011		
26/07/1974		13/09/1975	20/04/1999	10/09/2011		
05/09/1974		13/10/1975	18/10/1999	02/07/2013		
12/08/1975		15/12/1975	31/12/2000	17/06/2014		
14/06/1976		28/03/1976	22/02/2001			
20/08/1976		14/04/1976	29/05/2001			
25/07/1978		17/09/1976	27/06/2001			
16/05/1979		08/11/1976	06/08/2001			
11/12/1981		14/02/1977	01/12/2001			
19/05/1984		19/10/1977	07/12/2002			
18/06/1984		11/12/1977	16/06/2004			
09/08/1984		10/10/1979	22/07/2004			
13/11/1984		28/01/1980	19/03/2007			
21/02/1985		02/03/1980	24/05/2007			
25/05/1988		14/07/1980	15/11/2007			
27/10/1988		07/08/1980	07/05/2008			
15/05/1989		06/09/1980	18/08/2008			
05/08/1989		28/02/1982	04/12/2009			
24/06/1991		20/07/1982	29/01/2010			
12/10/1991		02/02/1983	07/04/2010			
15/05/1993		06/03/1983	23/06/2010			
13/05/1994		03/10/1983	03/04/2011			
18/07/1995		24/12/1983	20/10/2011			
29/09/1995		29/09/1984	20/06/2012			
12/07/1996		21/04/1985	02/08/2012			
21/06/1997		29/12/1986	22/02/2013			
02/03/2000		23/02/1987	22/03/2013			
01/09/2001		04/02/1989	02/08/2013			
29/01/2002		07/04/1989	17/01/2014			
07/08/2002		25/09/1989	22/03/2014			
02/10/2002		09/04/1990	03/10/2014			
01/01/2004		02/05/1990	02/01/2015			
27/10/2004		07/06/1990	23/02/2015			
70	0 eventos	1(02 eventos	22 eventos		

APÊNDICE B – VARIAÇÃO DE MÉTRICAS NA FASE DE TREINAMENTO

FIGURA B.1 – EVOLUÇÃO DO COEFICIENTE KG PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DE SÉRIE COMPLETA – FASE DE TREINAMENTO

SC - Kling-Gupta - Treinamento

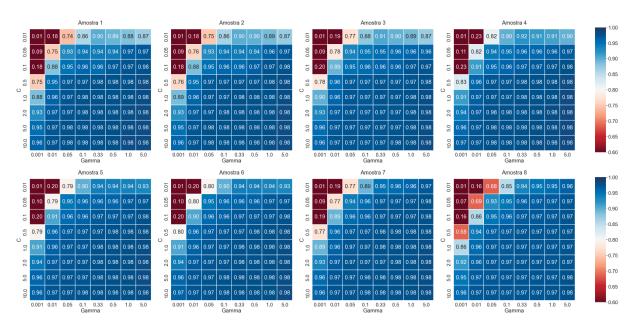


FIGURA B.2 – EVOLUÇÃO DO COEFICIENTE NS PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DE SÉRIE COMPLETA – FASE DE TREINAMENTO

SC - Nash-Sutcliffe - Treinamento

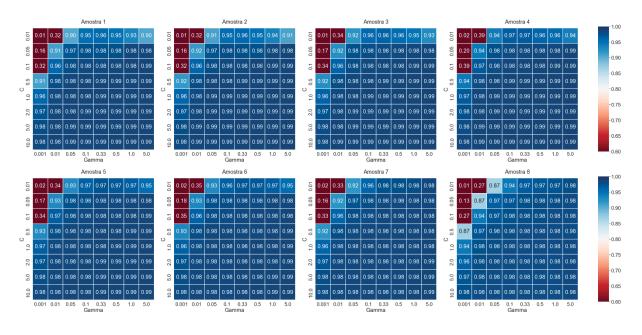


FIGURA B.3 – EVOLUÇÃO DO COEFICIENTE NS MODIFICADO PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DE SÉRIE COMPLETA – FASE DE TREINAMENTO

SC - Nash-Sutcliffe Modificado - Treinamento

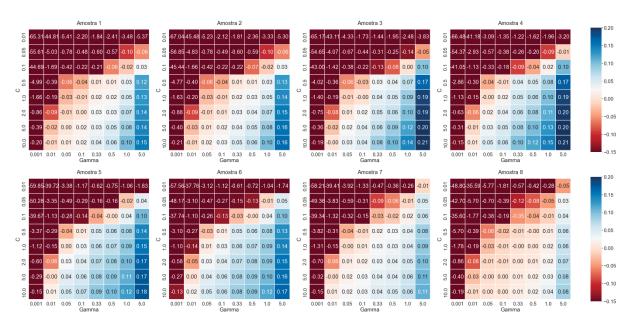


FIGURA B.4 – EVOLUÇÃO DO COEFICIENTE MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DE SÉRIE COMPLETA – FASE DE TREINAMENTO.

SC - Mean Absolute Relative Error - Treinamento

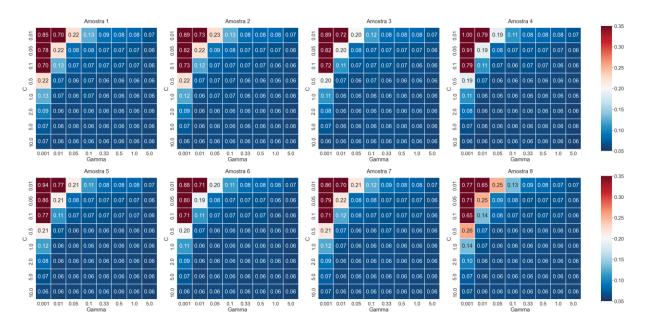


FIGURA B.5 – EVOLUÇÃO DO COEFICIENTE KG PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DE TODOS OS EVENTOS – FASE DE TREINAMENTO

EV - Kling-Gupta - Treinamento

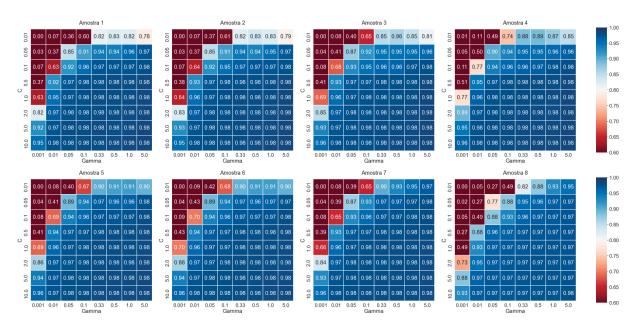


FIGURA B.6 – EVOLUÇÃO DO COEFICIENTE NS PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DE TODOS OS EVENTOS – FASE DE TREINAMENTO

EV - Nash-Sutcliffe - Treinamento

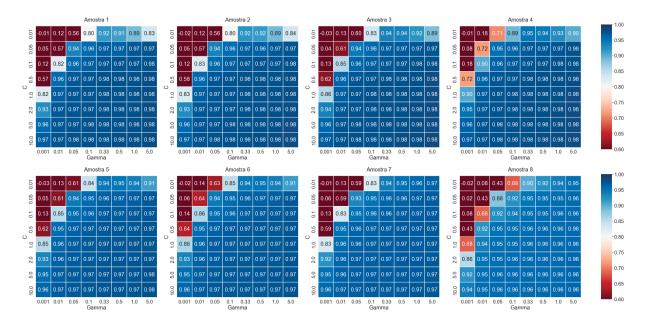


FIGURA B.7 – EVOLUÇÃO DO COEFICIENTE NS MODIFICADO PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DE TODOS OS EVENTOS – FASE DE TREINAMENTO

EV - Nash-Sutcliffe Modificado - Treinamento

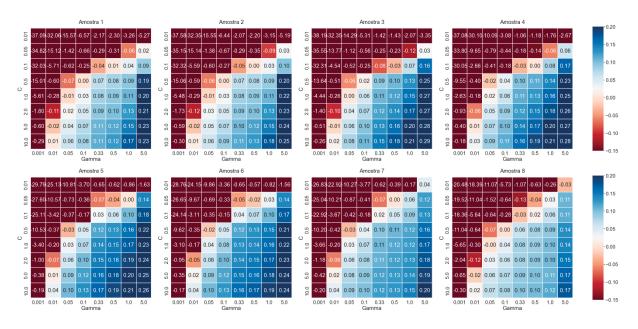


FIGURA B.8 – EVOLUÇÃO DO COEFICIENTE MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DE TODOS OS EVENTOS – FASE DE TREINAMENTO

EV - Mean Absolute Relative Error - Treinamento

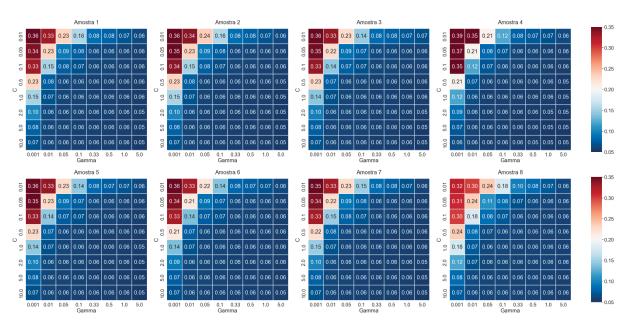


FIGURA B.9 – EVOLUÇÃO DO COEFICIENTE KG PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 1 – FASE DE TREINAMENTO

G1 - Kling-Gupta - Treinamento

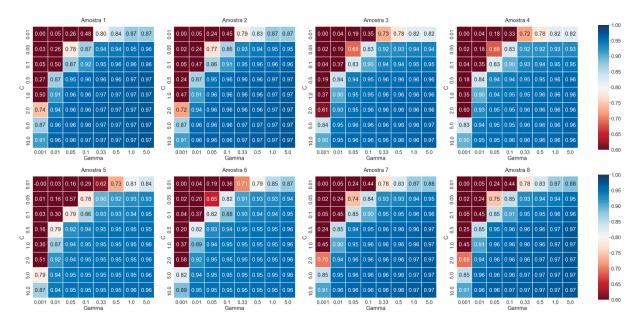


FIGURA B.10 – EVOLUÇÃO DO COEFICIENTE NS PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 1 – FASE DE TREINAMENTO

G1 - Nash-Sutcliffe - Treinamento

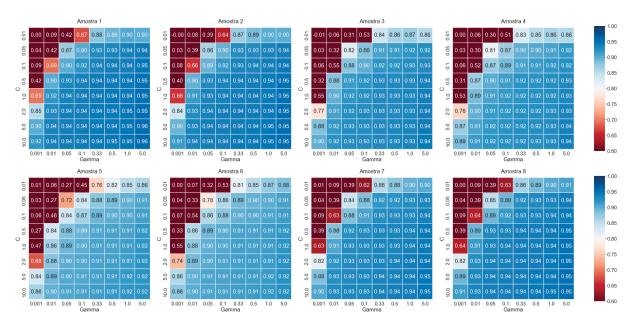


FIGURA B.11 – EVOLUÇÃO DO COEFICIENTE NS MODIFICADO PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 1 – FASE DE TREINAMENTO

G1 - Nash-Sutcliffe Modificado - Treinamento

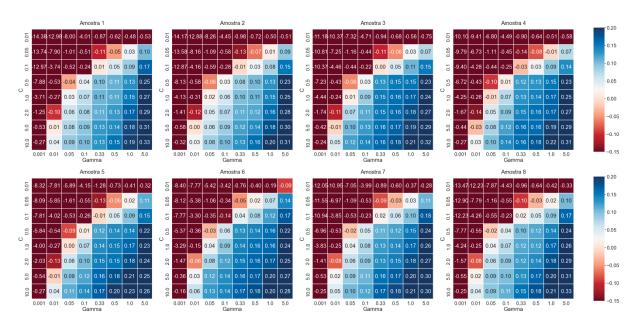


FIGURA B.12 – EVOLUÇÃO DO COEFICIENTE MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 1 – FASE DE TREINAMENTO

G1 - Mean Absolute Relative Error - Treinamento

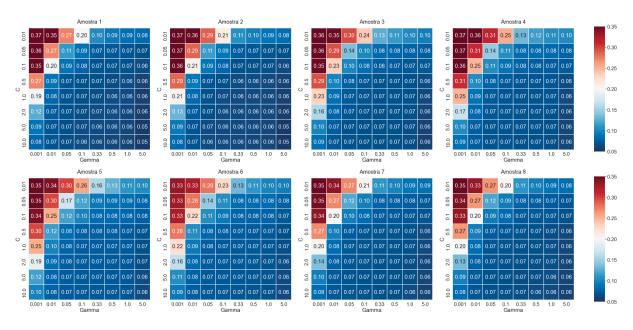


FIGURA B.13 – EVOLUÇÃO DO COEFICIENTE KG PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 2 – FASE DE TREINAMENTO

G2 - Kling-Gupta - Treinamento

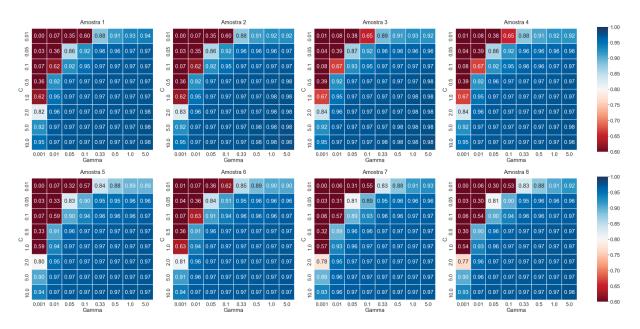


FIGURA B.14 – EVOLUÇÃO DO COEFICIENTE NS PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 2 – FASE DE TREINAMENTO

G2 - Nash-Sutcliffe - Treinamento

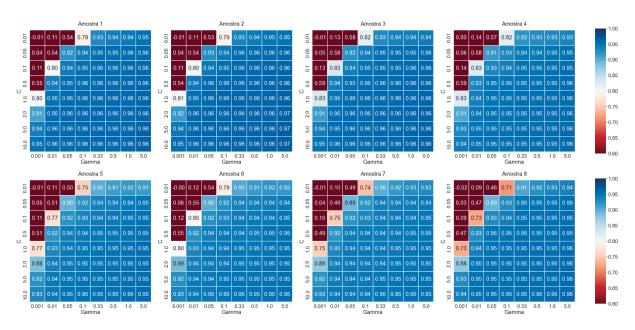


FIGURA B.15 – EVOLUÇÃO DO COEFICIENTE NS MODIFICADO PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 2 – FASE DE TREINAMENTO

G2 - Nash-Sutcliffe Modificado - Treinamento

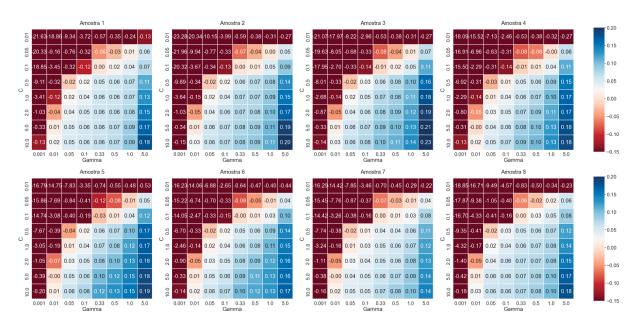


FIGURA B.16 – EVOLUÇÃO DO COEFICIENTE MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 2 – FASE DE TREINAMENTO

G2 - Mean Absolute Relative Error - Treinamento

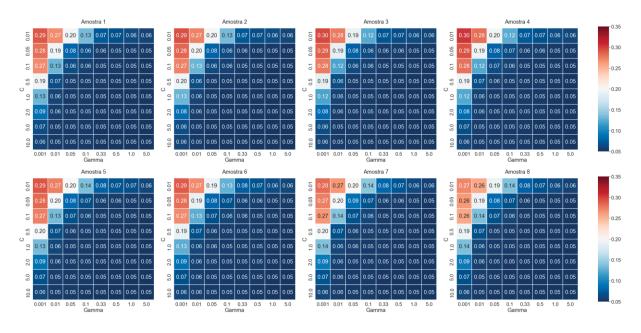


FIGURA B.17 – EVOLUÇÃO DO COEFICIENTE KG PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 3 – FASE DE TREINAMENTO

G3 - Kling-Gupta - Treinamento

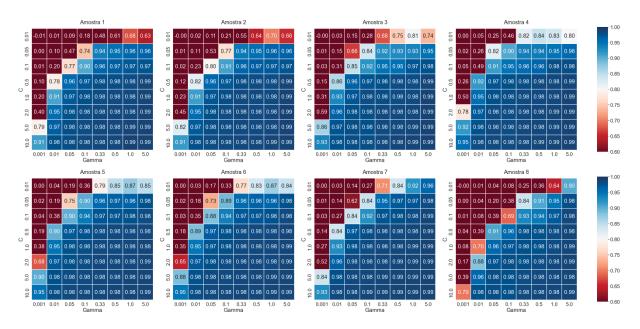


FIGURA B.18 – EVOLUÇÃO DO COEFICIENTE NS PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 3 – FASE DE TREINAMENTO

G3 - Nash-Sutcliffe - Treinamento

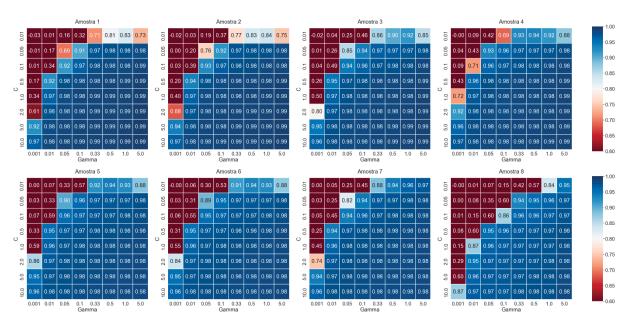


FIGURA B.19 – EVOLUÇÃO DO COEFICIENTE NS MODIFICADO PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 3 – FASE DE TREINAMENTO

G3 - Nash-Sutcliffe Modificado - Treinamento

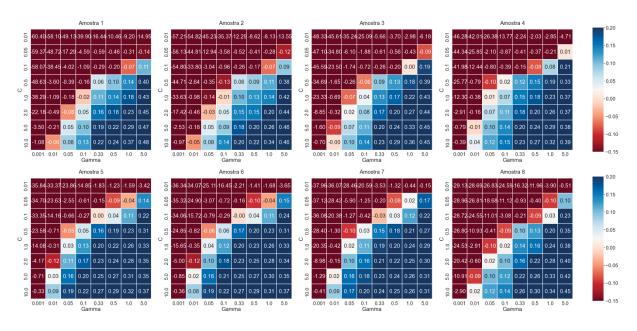
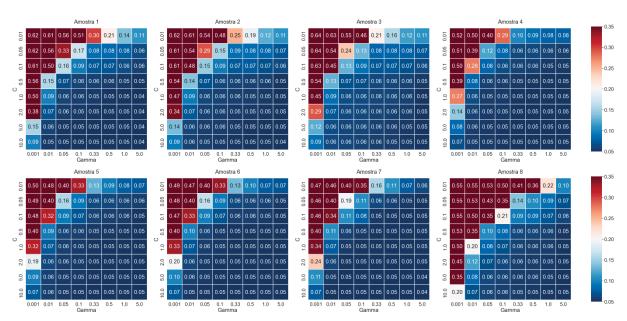


FIGURA B.20 – EVOLUÇÃO DO COEFICIENTE MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 3 – FASE DE TREINAMENTO

G3 - Mean Absolute Relative Error - Treinamento



APÊNDICE C – DISTRIBUIÇÃO DAS MÉTRICAS NA FASE DE TREINAMENTO

FIGURA C.1 – DISTRIBUIÇÃO DOS COEFICIENTES KG, NS, NS MODIFICADO E MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DE SÉRIE COMPLETA – FASE DE TREINAMENTO

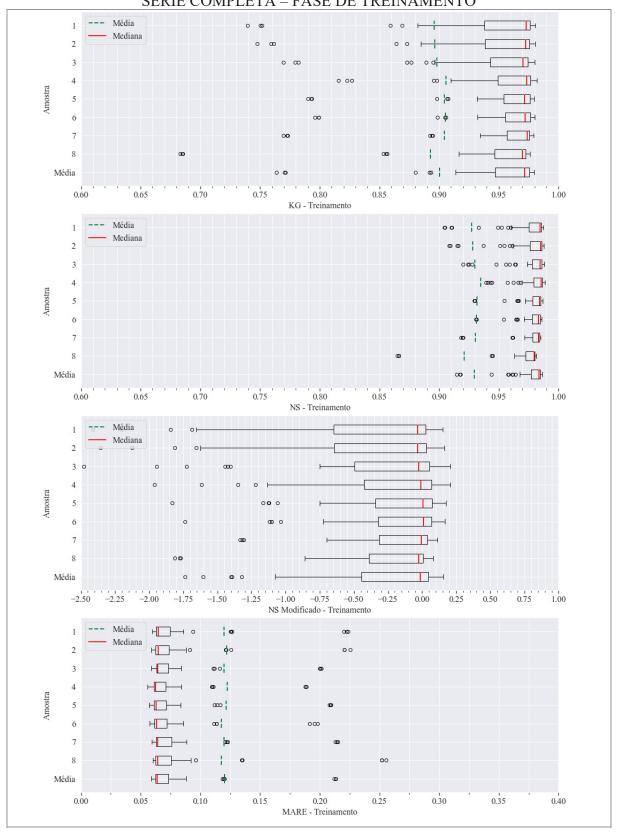


FIGURA C.2 – DISTRIBUIÇÃO DOS COEFICIENTES KG, NS, NS MODIFICADO E MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DE TODOS OS EVENTOS – FASE DE TREINAMENTO

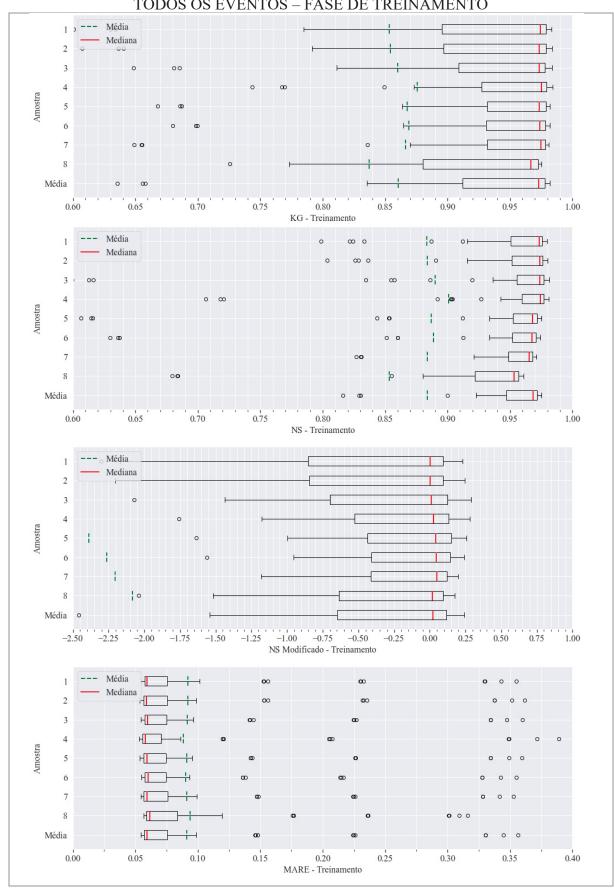


FIGURA C.3 – DISTRIBUIÇÃO DOS COEFICIENTES KG, NS, NS MODIFICADO E MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 1 – FASE DE TREINAMENTO

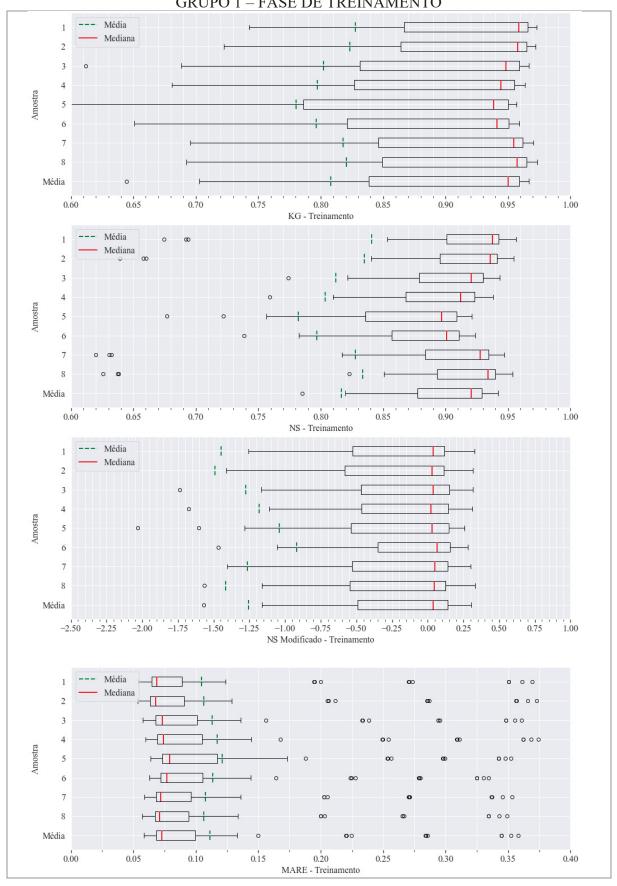


FIGURA C.4 – DISTRIBUIÇÃO DOS COEFICIENTES KG, NS, NS MODIFICADO E MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 2 – FASE DE TREINAMENTO

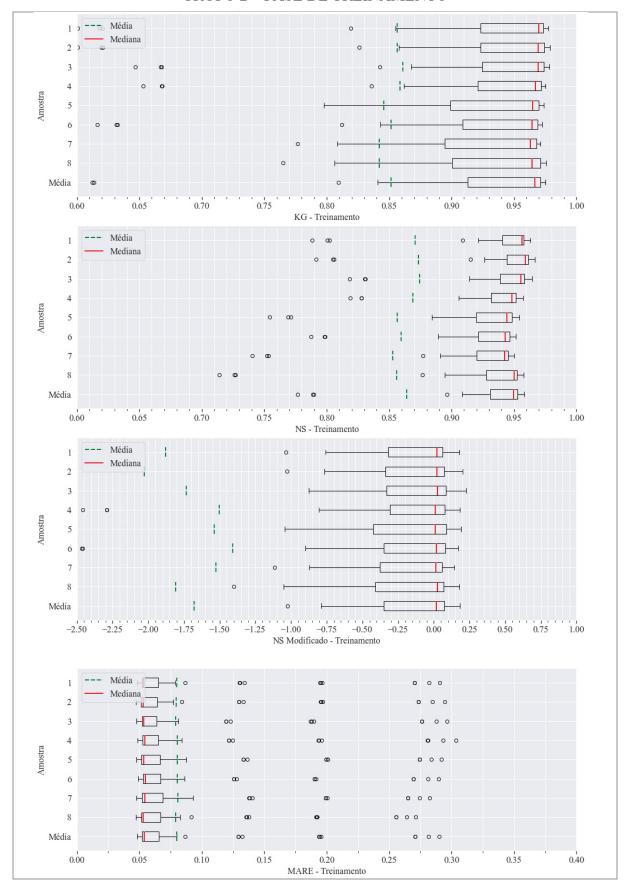
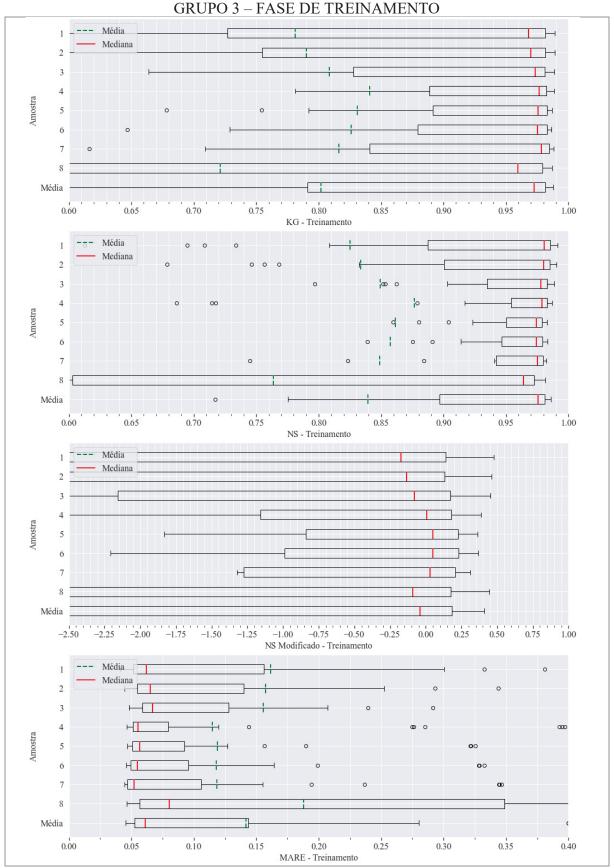


FIGURA C.5 – DISTRIBUIÇÃO DOS COEFICIENTES KG, NS, NS MODIFICADO E MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 3 FASE DE TREINAMENTO



APÊNDICE D – DISTRIBUIÇÃO DAS MÉTRICAS NA FASE DE TESTE

FIGURA D.1 – DISTRIBUIÇÃO DOS COEFICIENTES KG, NS, NS MODIFICADO E MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DE SÉRIE COMPLETA – FASE DE TESTE

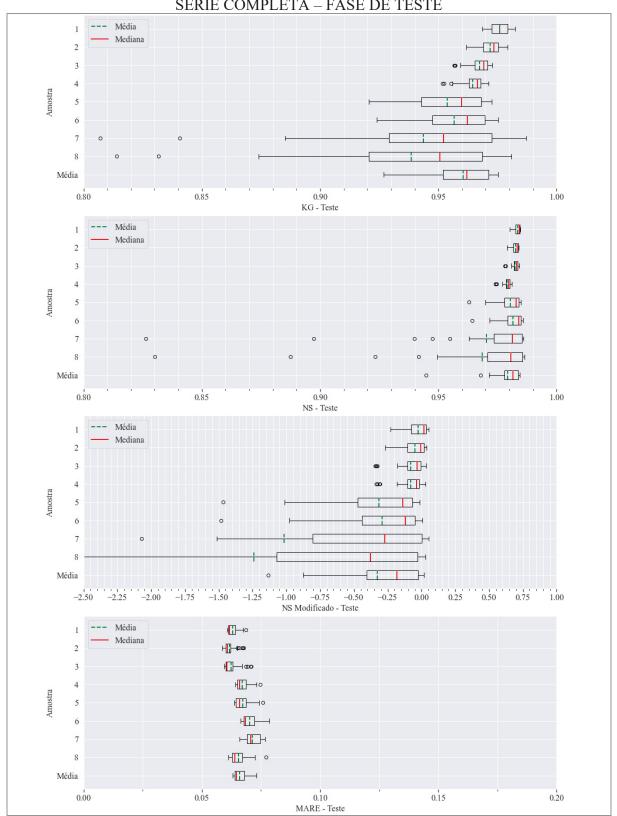


FIGURA D.2 – DISTRIBUIÇÃO DOS COEFICIENTES KG, NS, NS MODIFICADO E MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DE TODOS OS EVENTOS – FASE DE TESTE

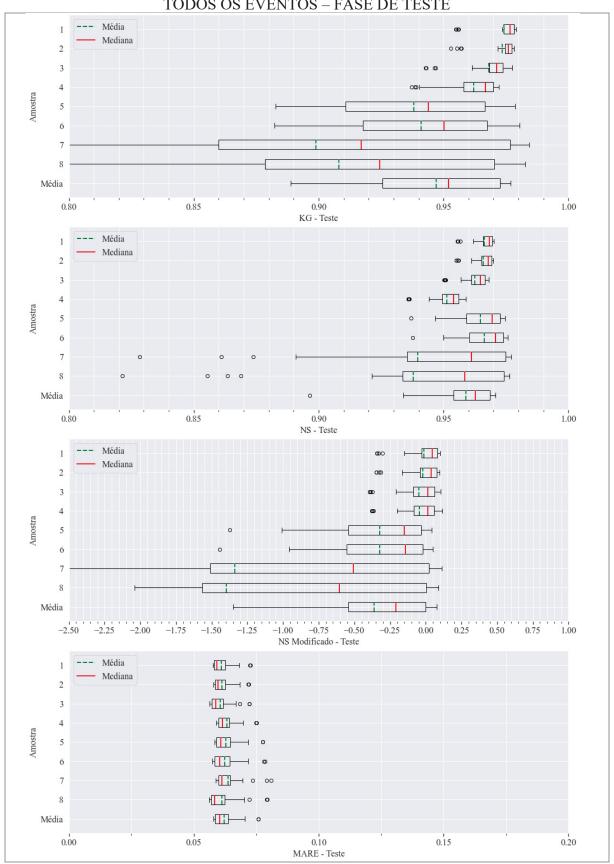


FIGURA D.3 – DISTRIBUIÇÃO DOS COEFICIENTES KG, NS, NS MODIFICADO E MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 1 FASE DE TESTE

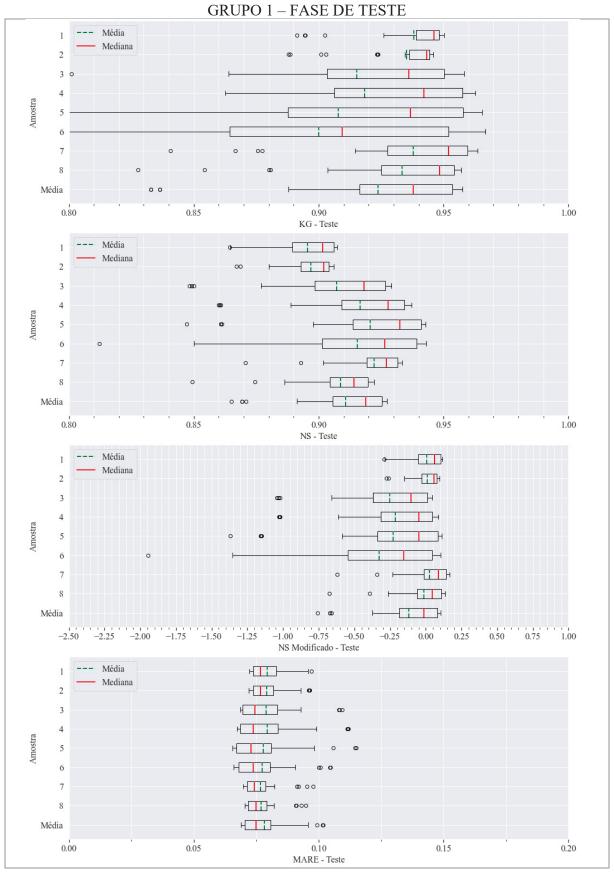


FIGURA D.4 – DISTRIBUIÇÃO DOS COEFICIENTES KG, NS, NS MODIFICADO E MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 2 – FASE DE TESTE

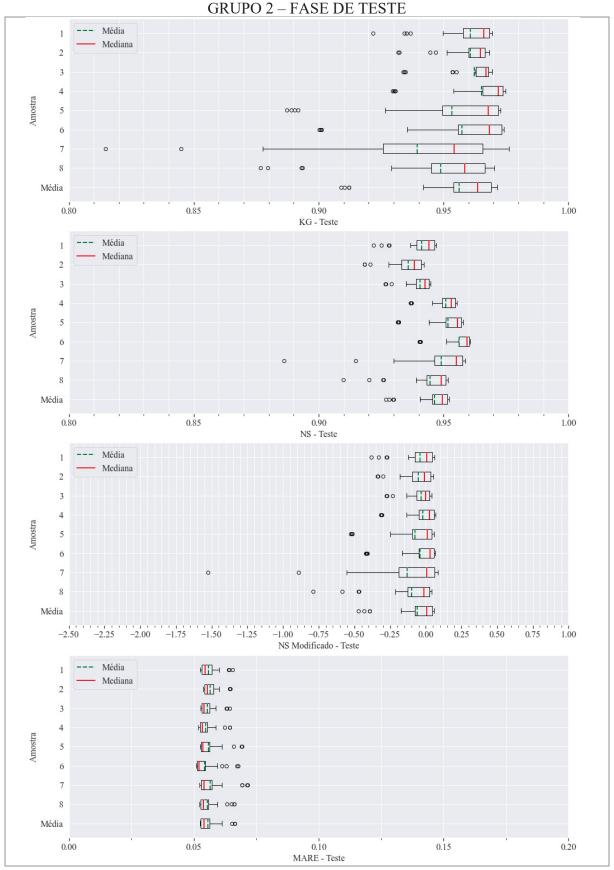
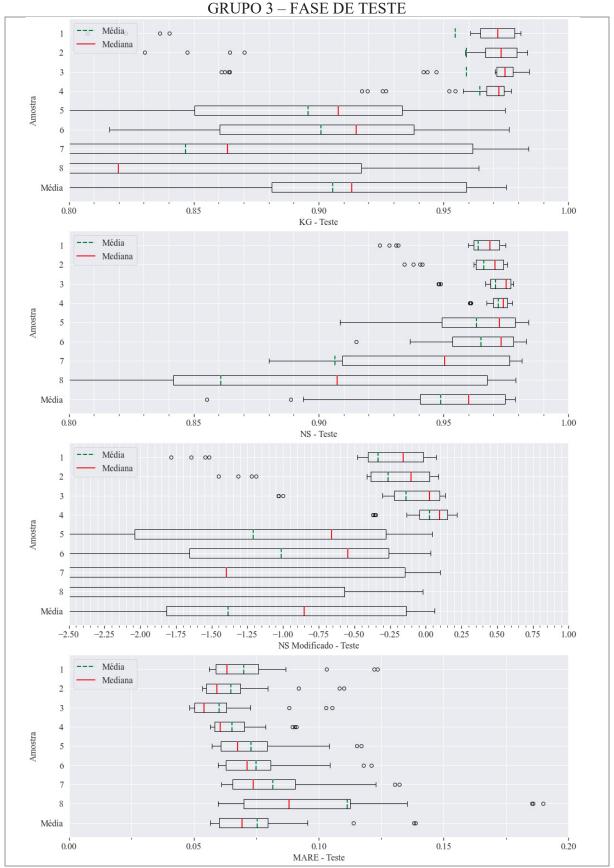


FIGURA D.5 – DISTRIBUIÇÃO DOS COEFICIENTES KG, NS, NS MODIFICADO E MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO GRUPO 3 FASE DE TESTE



APÊNDICE E - VARIAÇÃO DAS MÉTRICAS NA FASE DE TESTE

FIGURA E.1 – EVOLUÇÃO DO COEFICIENTE KG PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DE SÉRIE COMPLETA – FASE DE TESTE

SC - Kling-Gupta - Teste

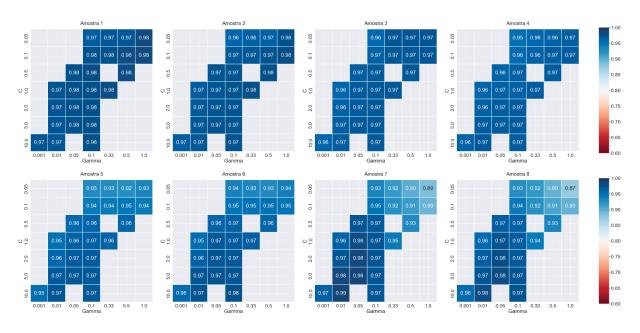


FIGURA E.2 – EVOLUÇÃO DO COEFICIENTE NS PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DE SÉRIE COMPLETA – FASE DE TESTE

SC - Nash-Sutcliffe - Teste

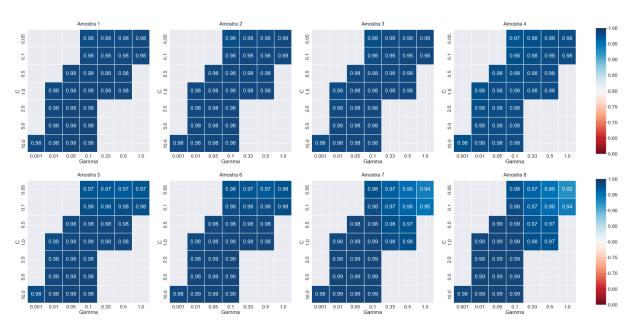


FIGURA E.3 – EVOLUÇÃO DO COEFICIENTE NS MODIFICADO PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DE SÉRIE COMPLETA – FASE DE TESTE

SC - Nash-Sutcliffe Modificado - Teste



FIGURA E.4 – EVOLUÇÃO DO COEFICIENTE MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DE SÉRIE COMPLETA – FASE DE TESTE

SC - Mean Absolute Relative Error - Teste

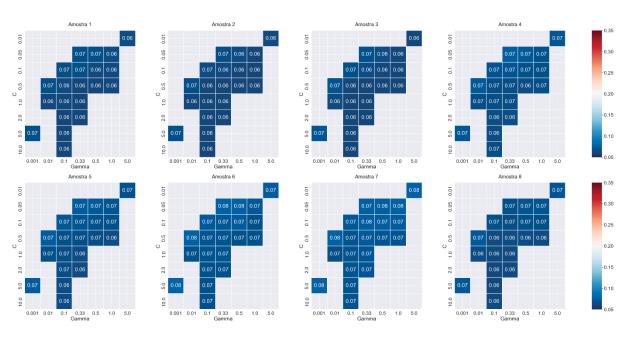


FIGURA E.5 – EVOLUÇÃO DO COEFICIENTE KG PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DE TODOS OS EVENTOS – FASE DE TESTE

EV - Kling-Gupta - Teste

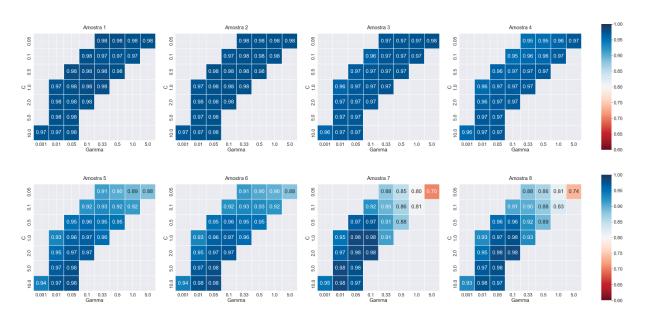


FIGURA E.6 – EVOLUÇÃO DO COEFICIENTE NS PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DE TODOS OS EVENTOS – FASE DE TESTE

EV - Nash-Sutcliffe - Teste

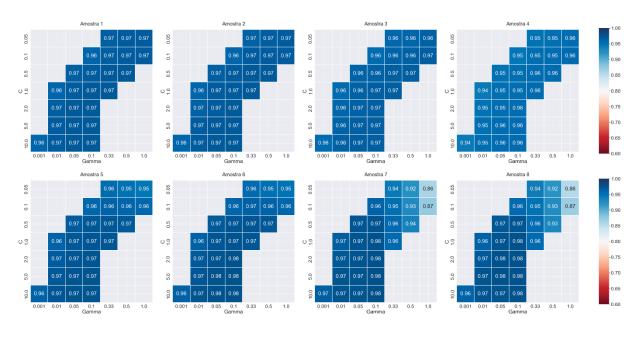


FIGURA E.7 – EVOLUÇÃO DO COEFICIENTE NS MODIFICADO PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DE TODOS OS EVENTOS – FASE DE TESTE

EV - Nash-Sutcliffe Modificado - Teste

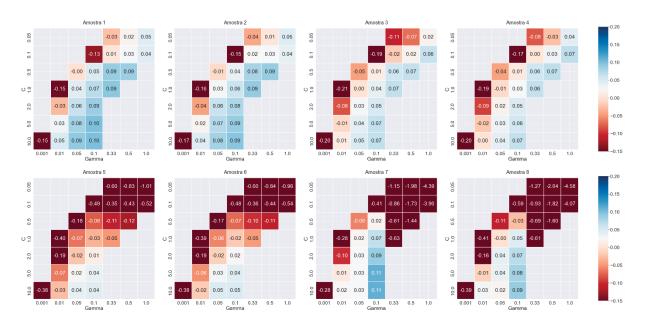


FIGURA E.8 – EVOLUÇÃO DO COEFICIENTE MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DE TODOS OS EVENTOS – FASE DE TESTE

EV - Mean Absolute Relative Error - Teste

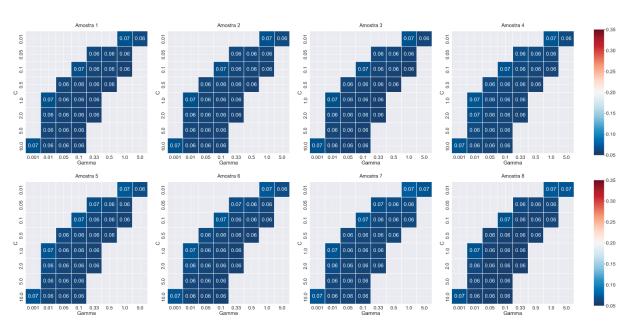


FIGURA E.9 – EVOLUÇÃO DO COEFICIENTE KG PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 1 – FASE DE TESTE

G1 - Kling-Gupta - Teste

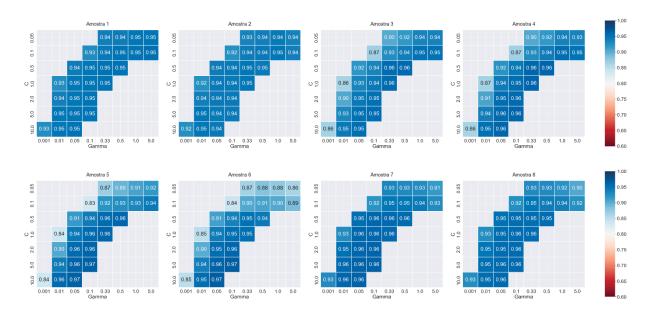


FIGURA E.10 – EVOLUÇÃO DO COEFICIENTE NS PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 1 – FASE DE TESTE

G1 - Nash-Sutcliffe - Teste

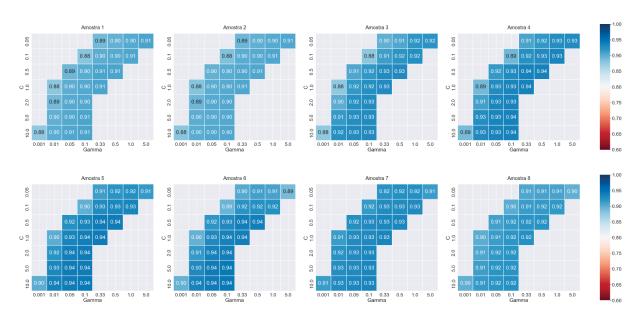


FIGURA E.11 – EVOLUÇÃO DO COEFICIENTE NS MODIFICADO PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 1 – FASE DE TESTE

G1 - Nash-Sutcliffe Modificado - Teste

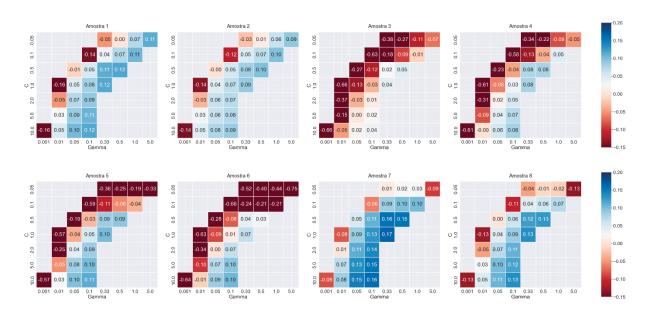


FIGURA E.12 – EVOLUÇÃO DO COEFICIENTE MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 1 – FASE DE TESTE

G1 - Mean Absolute Relative Error - Teste

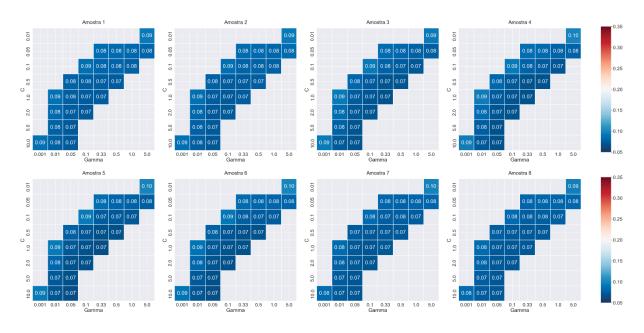


FIGURA E.13 – EVOLUÇÃO DO COEFICIENTE KG PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 2 – FASE DE TESTE

G2 - Kling-Gupta - Teste

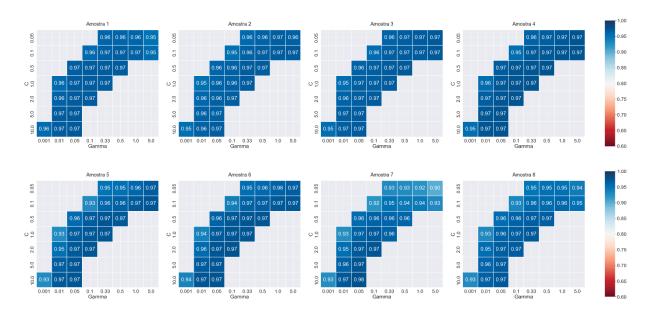


FIGURA E.14 – EVOLUÇÃO DO COEFICIENTE NS PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 2 – FASE DE TESTE

G2 - Nash-Sutcliffe - Teste

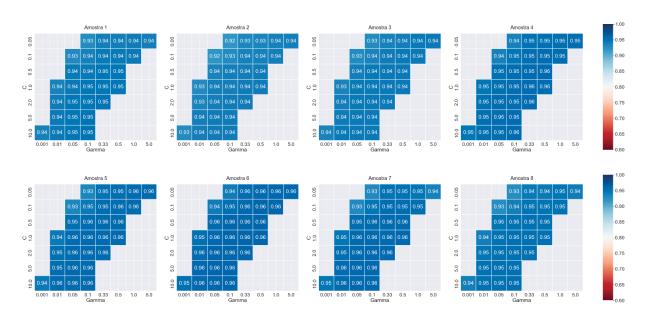


FIGURA E.15 – EVOLUÇÃO DO COEFICIENTE NS MODIFICADO PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 2 – FASE DE TESTE

G2 - Nash-Sutcliffe Modificado - Teste

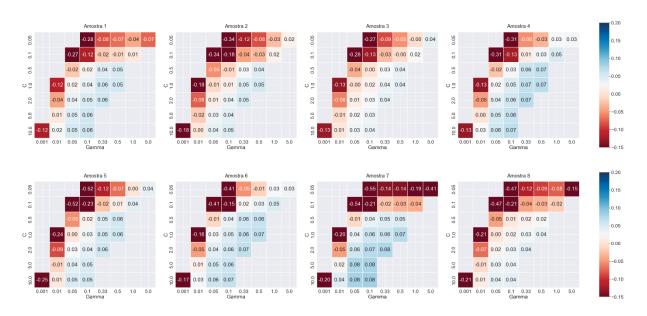


FIGURA E.16 – EVOLUÇÃO DO COEFICIENTE MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 2 – FASE DE TESTE

G2 - Mean Absolute Relative Error - Teste

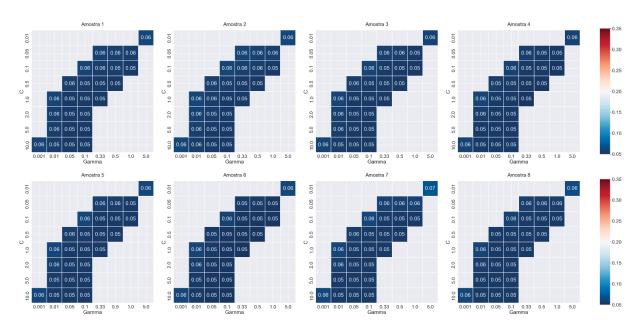


FIGURA E.17 – EVOLUÇÃO DO COEFICIENTE KG PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 3 – FASE DE TESTE

G3 - Kling-Gupta - Teste

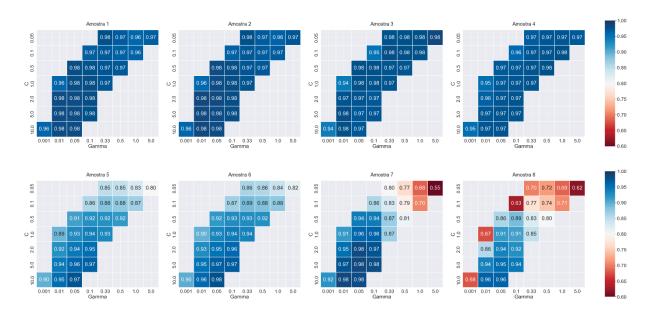


FIGURA E.18 – EVOLUÇÃO DO COEFICIENTE NS PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 3 – FASE DE TESTE

G3 - Nash-Sutcliffe - Teste

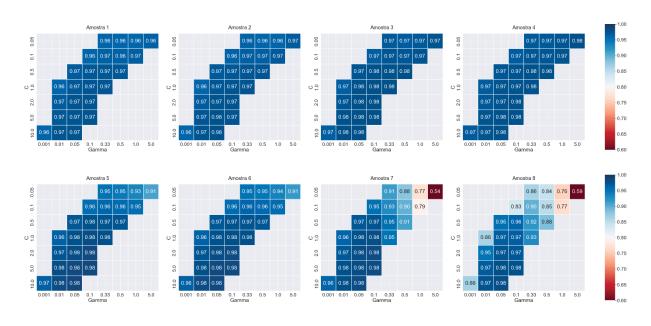


FIGURA E.19 – EVOLUÇÃO DO COEFICIENTE NS MODIFICADO PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 3 – FASE DE TESTE

G3 - Nash-Sutcliffe Modificado - Teste

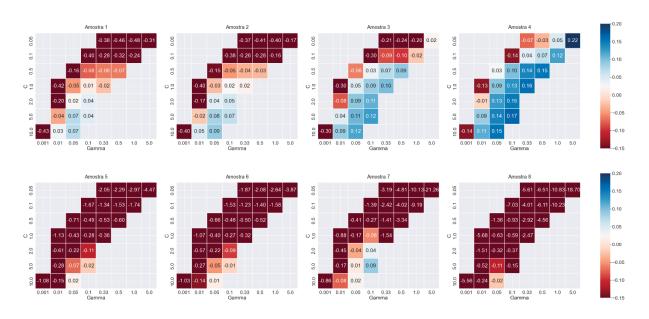
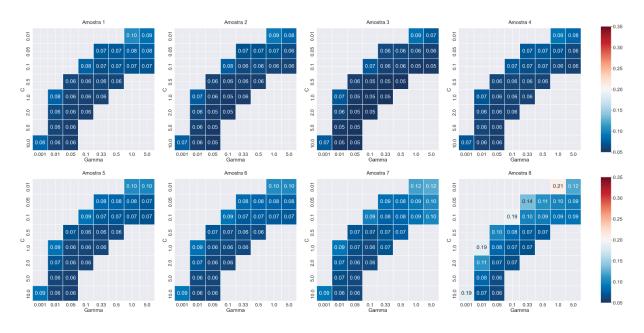


FIGURA E.20 – EVOLUÇÃO DO COEFICIENTE MARE PARA AS AMOSTRAS E CONJUNTOS DE PARÂMETROS DO MODELO DO GRUPO 3 – FASE DE TESTE

G3 - Mean Absolute Relative Error - Teste



APÊNDICE F – COEFICIENTES KG, NS, NS MODIFICADO E MARE DOS EVENTOS PREVISTOS

TABELA F.1 – VALORES DE KG PARA OS EVENTOS PREVISTOS – 2005 A 2015

Data do pico	ID	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Previsão média	Previsão por grupo	Previsão Euclidiana	Previsão Pertenc
26/05/2005	1	0,869	0,893	0,648	0,878	0,877	0,872	0,648	0,876	0,651
20/06/2005	2	0,925	0,931	0,652	0,867	0,925	0,889	0,652	0,902	0,658
20/09/2005	3	0,973	0,977	0,843	0,942	0,978	0,957	0,945	0,946	0,900
26/02/2007	4	0,960	0,937	0,633	0,844	0,919	0,870	0,633	0,884	0,728
19/03/2007	5	0,894	0,876	0,534	0,804	0,819	0,804	0,804	0,777	0,768
28/04/2007	6	0,839	0,854	0,600	0,835	0,823	0,832	0,600	0,823	0,603
24/05/2007	7	0,964	0,952	0,819	0,928	0,944	0,927	0,936	0,919	0,931
8/10/2007	8	0,748	0,788	0,549	0,794	0,800	0,790	0,549	0,810	0,554
5/11/2007	9	0,915	0,929	0,741	0,903	0,924	0,911	0,903	0,922	0,863
2/12/2007	10	0,847	0,904	0,626	0,919	0,879	0,892	0,626	0,916	0,629
07/05/2008	11	0,952	0,921	0,678	0,835	0,938	0,871	0,835	0,874	0,833
8/08/2008	12	0,929	0,899	0,607	0,818	0,900	0,846	0,818	0,858	0,800
31/10/2008	13	0,965	0,961	0,786	0,918	0,948	0,927	0,786	0,933	0,787
4/08/2009	14	0,962	0,970	0,734	0,915	0,969	0,928	0,734	0,942	0,735
3/10/2009	15	0,985	0,986	0,859	0,956	0,979	0,966	0,979	0,938	0,980
4/12/2009	16	0,898	0,893	0,786	0,921	0,871	0,916	0,921	0,897	0,898
29/01/2010	17	0,938	0,936	0,794	0,959	0,897	0,926	0,950	0,881	0,943
7/04/2010	18	0,967	0,952	0,745	0,908	0,950	0,908	0,908	0,879	0,908
8/05/2010	19	0,976	0,976	0,874	0,971	0,975	0,957	0,973	0,927	0,973
3/06/2010	20	0,796	0,809	0,590	0,810	0,797	0,788	0,810	0,768	0,810
1/07/2010	21	0,909	0,920	0,692	0,888	0,921	0,892	0,692	0,912	0,711
5/02/2011	22	0,980	0,971	0,788	0,916	0,965	0,927	0,949	0,895	0,934
3/04/2011	23	0,944	0,925	0,689	0,855	0,915	0,873	0,855	0,865	0,855
0/09/2011	24	0,959	0,956	0,875	0,944	0,958	0,944	0,954	0,938	0,949
0/10/2011	25	0,926	0,926	0,721	0,894	0,926	0,897	0,894	0,891	0,894
01/05/2012	26	0,893	0,920	0,721	0,900	0,921	0,915	0,727	0,918	0,730
20/06/2012	27	0,893	0,922	0,727	0,900	0,962	0,913	0,727	0,886	0,730
2/08/2012	28	0,979	0,937	0,794	0,908	0,902	0,895	0,927	0,880	0,924
0/01/2013	29	0,843	0,825	0,710	0,759	0,920	0,766	0,482	0,880	0,483
				· ·			· ·		· ·	
2/02/2013	30	0,885	0,896	0,564	0,828	0,881	0,846	0,828	0,851	0,786
2/03/2013	31	0,882	0,842	0,536	0,735	0,820	0,771	0,735	0,778	0,715
03/06/2013	32	0,878	0,895	0,605	0,852	0,891	0,863	0,605	0,883	0,623
2/07/2013	33	0,975	0,978	0,860	0,947	0,982	0,957	0,928	0,942	0,932
2/08/2013	34	0,971	0,953	0,749	0,903	0,967	0,915	0,903	0,890	0,904
4/09/2013	35	0,963	0,959	0,829	0,938	0,959	0,940	0,877	0,941	0,870
7/01/2014	36	0,918	0,886	0,613	0,819	0,854	0,832	0,819	0,827	0,799
2/03/2014	37	0,942	0,905	0,676	0,861	0,915	0,867	0,861	0,846	0,861
7/06/2014	38	0,947	0,945	0,909	0,976	0,952	0,951	0,950	0,950	0,951
4/09/2014	39	0,790	0,786	0,548	0,745	0,800	0,764	0,548	0,785	0,583
3/10/2014	40	0,935	0,925	0,778	0,887	0,908	0,897	0,887	0,888	0,886
9/11/2014	41	0,716	0,728	0,516	0,733	0,753	0,727	0,516	0,749	0,535
02/01/2015	42	0,911	0,881	0,656	0,809	0,855	0,829	0,809	0,815	0,807
23/02/2015	43	0,917	0,883	0,575	0,793	0,849	0,808	0,793	0,774	0,793
04/04/2015	44	0,859	0,891	0,595	0,871	0,892	0,868	0,595	0,896	0,669
31/05/2015	45	0,851	0,902	0,610	0,895	0,895	0,888	0,610	0,911	0,613
20/06/2015	46	0,944	0,948	0,688	0,891	0,940	0,904	0,688	0,920	0,705
24/10/2015	47	0,976	0,970	0,860	0,958	0,955	0,952	0,836	0,966	0

TABELA F.2 – VALORES DE NS PARA OS EVENTOS PREVISTOS – 2005 A 2015

Data do	ID	Modelo	Modelo	Modelo	Modelo	Modelo	Previsão	Previsão	Previsão	Previsão
pico		1	2	3	4	5	média	por grupo	Euclidiana	Pertenc.
26/05/2005	1	0,754	0,785	0,146	0,791	0,763	0,763	0,146	0,786	0,151
20/06/2005	2	0,860	0,874	0,365	0,876	0,873	0,864	0,365	0,884	0,373
20/09/2005	3	0,955	0,956	0,917	0,944	0,958	0,956	0,951	0,949	0,946
26/02/2007	4	0,925	0,927	-0,012	0,916	0,945	0,895	-0,012	0,927	0,616
19/03/2007	5	0,806	0,792	-3,588	0,709	0,803	0,525	0,709	0,597	0,657
28/04/2007	6	0,675	0,714	0,211	0,736	0,700	0,719	0,211	0,729	0,217
24/05/2007	7	0,954	0,954	0,901	0,947	0,951	0,952	0,956	0,938	0,953
18/10/2007	8	0,518	0,569	-1,582	0,531	0,591	0,487	-1,582	0,606	-1,565
15/11/2007	9	0,849	0,857	0,457	0,855	0,855	0,839	0,855	0,841	0,830
12/12/2007	10	0,871	0,875	-1,549	0,812	0,861	0,706	-1,549	0,845	-1,539
07/05/2008	11	0,930	0,928	0,638	0,913	0,937	0,916	0,913	0,900	0,913
18/08/2008	12	0,871	0,874	-0,284	0,861	0,900	0,842	0,861	0,865	0,860
31/10/2008	13	0,933	0,936	0,802	0,924	0,931	0,931	0,802	0,928	0,802
04/08/2009	14	0,941	0,946	0,680	0,940	0,953	0,940	0,680	0,948	0,681
03/10/2009	15	0,975	0,972	0,939	0,963	0,970	0,970	0,963	0,961	0,965
04/12/2009	16	0,905	0,890	0,062	0,870	0,801	0,827	0,870	0,719	0,850
29/01/2010	17	0,961	0,957	0,634	0,953	0,912	0,921	0,950	0,825	0,947
07/04/2010	18	0,946	0,947	0,682	0,937	0,942	0,925	0,937	0,872	0,937
08/05/2010	19	0,970	0,972	0,948	0,970	0,970	0,970	0,971	0,961	0,972
23/06/2010	20	0,551	0,506	-4,885	0,289	0,454	0,034	0,289	-0,171	0,289
21/07/2010	21	0,829	0,835	-0,109	0,831	0,830	0,768	-0,109	0,814	-0,053
25/02/2011	22	0,972	0,972	0,871	0,965	0,969	0,965	0,965	0,949	0,964
03/04/2011	23	0,904	0,904	0,576	0,894	0,907	0,885	0,894	0,864	0,894
10/09/2011	24	0,930	0,930	0,902	0,913	0,931	0,927	0,931	0,915	0,932
20/10/2011	25	0,856	0,856	0,426	0,845	0,856	0,825	0,845	0,789	0,842
01/05/2012	26	0,824	0,845	0,665	0,860	0,845	0,855	0,665	0,855	0,667
20/06/2012	27	0,982	0,982	0,927	0,974	0,981	0,978	0,975	0,964	0,975
02/08/2012	28	0,938	0,937	0,571	0,931	0,939	0,914	0,931	0,884	0,930
10/01/2013	29	0,684	0,668	-1,594	0,584	0,681	0,500	-1,594	0,634	-1,589
22/02/2013	30	0,779	0,789	-3,763	0,738	0,756	0,532	0,738	0,601	0,668
22/03/2013	31	0,862	0,855	-0,406	0,820	0,859	0,785	0,820	0,806	0,808
03/06/2013	32	0,764	0,791	-0,661	0,788	0,785	0,737	-0,661	0,800	-0,601
02/07/2013	33	0,958	0,961	0,947	0,958	0,967	0,962	0,965	0,958	0,965
02/08/2013	34	0,947	0,945	0,742	0,930	0,944	0,930	0,930	0,893	0,929
24/09/2013	35	0,931	0,932	0,884	0,922	0,931	0,929	0,923	0,920	0,920
17/01/2014	36	0,889	0,902	0,037	0,895	0,878	0,878	0,895	0,887	0,887
22/03/2014	37	0,905	0,895	0,324	0,864	0,918	0,861	0,864	0,813	0,864
17/06/2014	38	0,957	0,958	0,950	0,954	0,959	0,958	0,957	0,955	0,958
04/09/2014	39	0,569	0,587	-0,653	0,578	0,620	0,547	-0,653	0,625	-0,543
03/10/2014	40	0,887	0,887	0,832	0,878	0,882	0,889	0,878	0,878	0,878
09/11/2014	41	0,404	0,420	-2,499	0,399	0,466	0,259	-2,499	0,435	-2,370
02/01/2015	42	0,908	0,906	0,691	0,900	0,906	0,902	0,900	0,898	0,900
23/02/2015	43	0,928	0,909	-0,714	0,850	0,894	0,773	0,850	0,615	0,850
04/04/2015	44	0,766	0,770	-1,994	0,721	0,773	0,590	-1,994	0,726	-1,248
31/05/2015	45	0,785	0,820	-0,550	0,817	0,803	0,757	-0,550	0,822	-0,541
20/06/2015	46	0,897	0,906	0,469	0,910	0,901	0,892	0,469	0,908	0,489
24/10/2015	47	0,962	0,962	0,928	0,956	0,956	0,964	0,926	0,960	0,938
			, -		,	,	, -	* -	,	,

 \mathbf{TABELA} F.3 – VALORES DE NS MODIFICADO PARA OS EVENTOS PREVISTOS – 2005 A 2015

Data do pico	ID	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Previsão média	Previsão por grupo	Previsão Euclidiana	Previsão Pertenc.
26/05/2005	1	0,050	0,168	-2,296	0,192	0,083	0,086	-2,296	0,175	-2,280
20/06/2005	2	0,156	0,244	-2,823	0,256	0,239	0,183	-2,823	0,302	-2,772
20/09/2005	3	0,173	0,192	-0,522	-0,027	0,217	0,182	0,101	0,054	0,013
26/02/2007	4	0,491	0,505	-5,871	0,430	0,626	0,284	-5,871	0,505	-1,611
19/03/2007	5	0,491	0,454	-11,041	0,236	0,482	-0,245	0,236	-0,057	0,100
28/04/2007	6	0,112	0,217	-1,159	0,277	0,179	0,232	-1,159	0,258	-1,143
24/05/2007	7	0,168	0,173	-0,793	0,046	0,109	0,127	0,202	-0,122	0,151
18/10/2007	8	0,114	0,207	-3,751	0,137	0,248	0,055	-3,751	0,275	-3,719
15/11/2007	9	0,154	0,201	-2,041	0,189	0,185	0,098	0,189	0,109	0,050
12/12/2007	10	0,248	0,274	-13,859	-0,098	0,192	-0,713	-13,859	0,095	-13,800
07/05/2008	11	0,316	0,295	-2,534	0,152	0,385	0,184	0,152	0,023	0,153
18/08/2008	12	0,311	0,328	-5,867	0,257	0,467	0,158	0,257	0,277	0,250
31/10/2008	13	0,041	0,082	-1,857	-0,101	0,000	0,000	-1,857	-0,030	-1,844
04/08/2009	14	0,333	0,389	-2,624	0,324	0,465	0,323	-2,624	0,414	-2,612
03/10/2009	15	0,160	0,082	-1,024	-0,239	0,012	0,013	-0,237	-0,311	-0,158
04/12/2009	16	-0,103	-0,282	-9,915	-0,515	-1,318	-1,018	-0,515	-2,273	-0,744
29/01/2010	17	-0,485	-0,645	-13,042	-0,801	-2,369	-2,028	-0,903	-5,728	-1,028
07/04/2010	18	0,281	0,303	-3,198	0,172	0,238	0,013	0,172	-0,685	0,169
08/05/2010	19	0,100	0,136	-0,569	0,094	0,095	0,082	0,131	-0,193	0,139
23/06/2010	20	-0,141	-0,256	-13,964	-0,807	-0,388	-1,457	-0,807	-1,976	-0,808
	21		0,160		*		-0,177	-4,629	0,054	-4,348
21/07/2010		0,134		-4,629 2,703	0,141	0,138		-	*	
25/02/2011	22	0,179	0,179	-2,793	-0,018	0,097	-0,045	-0,034	-0,513	-0,057
03/04/2011	23	0,139	0,138	-2,791	0,055	0,166	-0,029	0,055	-0,217	0,054
10/09/2011	24	0,121	0,130	-0,222	-0,092	0,141	0,085	0,133	-0,063	0,145
20/10/2011	25	0,128	0,127	-2,467	0,063	0,133	-0,057	0,063	-0,277	0,045
01/05/2012	26	0,169	0,265	-0,585	0,338	0,265	0,314	-0,585	0,313	-0,573
20/06/2012	27	0,489	0,492	-1,051	0,284	0,476	0,376	0,287	-0,007	0,309
02/08/2012	28	0,162	0,145	-4,848	0,055	0,171	-0,165	0,055	-0,582	0,053
10/01/2013	29	0,320	0,286	-4,585	0,105	0,314	-0,076	-4,585	0,212	-4,574
22/02/2013	30	0,090	0,131	-18,618	-0,081	-0,007	-0,929	-0,081	-0,643	-0,368
22/03/2013	31	0,228	0,185	-6,881	-0,008	0,209	-0,207	-0,008	-0,090	-0,079
03/06/2013	32	0,102	0,206	-5,308	0,196	0,184	0,003	-5,308	0,240	-5,081
02/07/2013	33	0,349	0,395	0,192	0,361	0,488	0,417	0,469	0,355	0,461
02/08/2013	34	0,267	0,250	-2,541	0,039	0,234	0,041	0,039	-0,466	0,025
24/09/2013	35	0,090	0,106	-0,521	-0,020	0,095	0,072	-0,008	-0,044	-0,051
17/01/2014	36	0,175	0,276	-6,148	0,223	0,092	0,091	0,223	0,159	0,160
22/03/2014	37	0,139	0,056	-5,100	-0,223	0,258	-0,257	-0,223	-0,689	-0,225
17/06/2014	38	0,114	0,146	-0,024	0,068	0,164	0,139	0,114	0,080	0,134
04/09/2014	39	0,114	0,151	-2,400	0,132	0,219	0,068	-2,400	0,230	-2,173
03/10/2014	40	0,030	0,035	-0,438	-0,044	-0,008	0,047	-0,044	-0,041	-0,044
09/11/2014	41	0,300	0,318	-3,109	0,294	0,372	0,130	-3,109	0,336	-2,957
02/01/2015	42	0,228	0,212	-1,592	0,165	0,213	0,175	0,165	0,145	0,165
23/02/2015	43	0,115	-0,115	-20,112	-0,845	-0,308	-1,794	-0,845	-3,746	-0,850
04/04/2015	44	0,270	0,282	-8,347	0,128	0,291	-0,281	-8,347	0,146	-6,018
31/05/2015	45	0,142	0,280	-5,193	0,267	0,211	0,029	-5,193	0,289	-5,160
20/06/2015	46	0,141	0,221	-3,423	0,247	0,175	0,099	-3,423	0,234	-3,252
24/10/2015	47	0,017	0,034	-0,836	-0,136	-0,113	0,069	-0,887	-0,014	-0,591

TABELA F.4 – VALORES DE MARE PARA OS EVENTOS PREVISTOS – 2005 A 2015

Data do pico	ID	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Previsão média	Previsão por grupo	Previsão Euclidiana	Previsão Pertenc.
26/05/2005	1	0,114	0,123	0,402	0,160	0,124	0,173	0,402	0,148	0,401
20/06/2005	2	0,104	0,094	0,290	0,106	0,091	0,118	0,290	0,094	0,287
20/09/2005	3	0,047	0,049	0,106	0,062	0,050	0,059	0,068	0,063	0,082
26/02/2007	4	0,055	0,057	0,244	0,062	0,050	0,067	0,244	0,054	0,143
19/03/2007	5	0,055	0,061	0,344	0,079	0,063	0,104	0,079	0,098	0,089
28/04/2007	6	0,150	0,165	0,518	0,217	0,152	0,226	0,518	0,197	0,515
24/05/2007	7	0,078	0,079	0,164	0,086	0,078	0,084	0,074	0,101	0,079
18/10/2007	8	0,137	0,141	0,407	0,152	0,132	0,162	0,407	0,135	0,405
15/11/2007	9	0,074	0,071	0,207	0,077	0,071	0,089	0,077	0,082	0,102
12/12/2007	10	0,060	0,072	0,367	0,097	0,075	0,119	0,367	0,085	0,366
07/05/2008	11	0,067	0,072	0,202	0,086	0,068	0,086	0,086	0,097	0,086
18/08/2008	12	0,059	0,057	0,262	0,063	0,050	0,080	0,063	0,073	0,066
31/10/2008	13	0,060	0,062	0,135	0,071	0,063	0,070	0,135	0,067	0,135
04/08/2009	14	0,060	0,054	0,211	0,060	0,054	0,071	0,211	0,059	0,210
03/10/2009	15	0,046	0,051	0,115	0,064	0,050	0,063	0,057	0,079	0,058
04/12/2009	16	0,051	0,055	0,199	0,064	0,070	0,077	0,064	0,103	0,067
29/01/2010	17	0,051	0,055	0,202	0,058	0,082	0,086	0,060	0,133	0,062
07/04/2010	18	0,070	0,072	0,199	0,081	0,079	0,093	0,081	0,120	0,081
08/05/2010	19	0,069	0,071	0,175	0,075	0,075	0,090	0,076	0,126	0,076
23/06/2010	20	0,005	0,104	0,371	0,123	0,110	0,143	0,123	0,120	0,123
	21	0,068	0,070	0,228	0,123	0,076	0,097	0,123	0,137	0,123
21/07/2010						,	,	-	· ·	
25/02/2011	22	0,048	0,050	0,139	0,058	0,054	0,062	0,060	0,079	0,061
03/04/2011	23	0,078	0,079	0,221	0,084	0,080	0,100	0,084	0,112	0,084
10/09/2011	24	0,070	0,071	0,115	0,086	0,071	0,080	0,075	0,089	0,077
20/10/2011	25	0,083	0,086	0,241	0,099	0,090	0,114	0,099	0,128	0,101
01/05/2012	26	0,115	0,127	0,320	0,147	0,127	0,150	0,320	0,143	0,318
20/06/2012	27	0,056	0,058	0,136	0,072	0,061	0,072	0,071	0,092	0,070
02/08/2012	28	0,073	0,080	0,269	0,092	0,078	0,107	0,092	0,126	0,092
10/01/2013	29	0,103	0,126	0,445	0,171	0,138	0,190	0,445	0,158	0,445
22/02/2013	30	0,060	0,058	0,306	0,060	0,058	0,081	0,060	0,073	0,070
22/03/2013	31	0,061	0,069	0,261	0,083	0,072	0,096	0,083	0,091	0,087
03/06/2013	32	0,101	0,097	0,314	0,092	0,095	0,109	0,314	0,090	0,307
02/07/2013	33	0,083	0,081	0,142	0,085	0,074	0,087	0,105	0,094	0,102
02/08/2013	34	0,064	0,072	0,192	0,090	0,072	0,095	0,090	0,121	0,090
24/09/2013	35	0,088	0,093	0,212	0,110	0,095	0,113	0,152	0,118	0,160
17/01/2014	36	0,068	0,065	0,257	0,075	0,068	0,082	0,075	0,077	0,079
22/03/2014	37	0,067	0,075	0,228	0,095	0,066	0,098	0,095	0,115	0,096
17/06/2014	38	0,092	0,093	0,176	0,098	0,084	0,100	0,084	0,124	0,084
04/09/2014	39	0,130	0,125	0,339	0,133	0,119	0,148	0,339	0,119	0,323
03/10/2014	40	0,097	0,094	0,178	0,101	0,092	0,099	0,101	0,105	0,102
09/11/2014	41	0,117	0,114	0,373	0,134	0,110	0,156	0,373	0,122	0,366
02/01/2015	42	0,074	0,085	0,206	0,103	0,086	0,104	0,103	0,108	0,103
23/02/2015	43	0,052	0,059	0,295	0,077	0,062	0,099	0,077	0,133	0,077
04/04/2015	44	0,078	0,079	0,347	0,093	0,081	0,119	0,347	0,093	0,296
31/05/2015	45	0,093	0,095	0,380	0,119	0,097	0,139	0,380	0,104	0,378
20/06/2015	46	0,088	0,086	0,253	0,093	0,086	0,104	0,253	0,087	0,246
24/10/2015	47	0,042	0,042	0,084	0,051	0,045	0,047	0,085	0,046	0,075

APÊNDICE G -ERROS RELATIVOS DAS ESTATÍSTICAS DOS EVENTOS

TABELA G.1 – ERROS RELATIVOS DAS VAZÕES MÍNIMAS PREVISTAS DOS EVENTOS – 2005 A 2015

Data do pico	ID	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Previsão média	Previsão por grupo	Previsão Euclidiana	Previsão Pertenc
26/05/2005	1	15,91%	41,05%	201,77%	76,44%	42,31%	75,50%	201,77%	65,25%	200,18%
20/06/2005	2	-10,77%	0,66%	81,28%	17,11%	0,64%	17,78%	81,28%	10,59%	79,26%
20/09/2005	3	7,31%	22,84%	134,85%	45,99%	25,09%	47,21%	120,01%	38,25%	133,70%
26/02/2007	4	-1,44%	4,11%	66,38%	13,94%	8,95%	18,39%	66,38%	15,86%	44,45%
19/03/2007	5	-0,94%	2,90%	55,40%	11,46%	9,20%	15,61%	11,46%	16,62%	15,52%
28/04/2007	6	10,14%	39,16%	227,59%	80,63%	43,92%	80,29%	227,59%	72,09%	225,51%
24/05/2007	7	-0,48%	6,48%	61,47%	16,31%	4,62%	17,68%	16,31%	16,21%	18,72%
18/10/2007	8	-29,15%	-16,93%	60,52%	-0,06%	-17,94%	-0,71%	60,52%	-7,27%	59,47%
15/11/2007	9	-14,77%	-7,15%	47,38%	3,71%	-9,24%	3,98%	3,71%	-0,41%	22,40%
12/12/2007	10	10,04%	26,88%	143,87%	51,25%	28,18%	52,04%	143,87%	42,62%	142,119
07/05/2008	11	19,39%	26,94%	91,61%	38,23%	26,16%	40,47%	38,23%	46,78%	38,23%
18/08/2008	12	-2,46%	2,63%	54,86%	10,63%	4,40%	14,01%	10,63%	13,06%	14,25%
31/10/2008	13	7,99%	23,96%	135,45%	47,02%	25,09%	47,92%	135,45%	38,82%	134,04%
04/08/2009	14	-15,14%	-0,03%	103,40%	22,08%	1,58%	22,38%	103,40%	15,05%	102,35%
03/10/2009	15	-1,37%	5,71%	65,23%	16,59%	5,69%	18,37%	16,59%	14,75%	28,51%
04/12/2009	16	-3,72%	0,26%	37,34%	3,08%	2,10%	10,22%	3,08%	10,86%	3,18%
29/01/2010	17	3,51%	3,96%	39,12%	4,90%	11,28%	12,56%	4,90%	19,95%	4,94%
07/04/2010	18	4,76%	8,02%	51,38%	14,00%	10,81%	17,86%	14,00%	21,38%	14,07%
08/05/2010	19	4,55%	9,65%	61,59%	18,35%	10,91%	21,01%	18,35%	23,13%	18,50%
23/06/2010	20	4,21%	8,62%	57,44%	16,29%	10,37%	19,38%	16,29%	21,54%	16,41%
21/07/2010	21	-4,71%	1,85%	67,43%	14,46%	6,14%	17,03%	67,43%	12,51%	61,28%
25/02/2011	22	2,90%	8,83%	65,47%	18,87%	10,17%	21,25%	18,87%	18,99%	23,76%
03/04/2011	23	8,14%	12,39%	64,88%	21,15%	16,27%	24,57%	21,15%	26,13%	21,37%
10/09/2011	24	22,44%	31,73%	105,86%	45,66%	30,90%	47,32%	45,66%	43,01%	57,43%
20/10/2011	25	0,94%	6,06%	59,38%	15,43%	8,29%	18,02%	15,43%	17,06%	17,41%
01/05/2012	26	21,63%	56,17%	272,56%	104,04%	60,34%	102,95%	272,56%	93,47%	270,06%
20/06/2012	27	3,28%	8,19%	61,01%	16,32%	10,35%	19,83%	16,32%	22,49%	16,58%
02/08/2012	28	12,82%	18,50%	71,19%	27,31%	17,95%	29,55%	27,31%	32,75%	27,37%
10/01/2013	29	7,18%	21,27%	134,78%	44,68%	27,13%	47,01%	134,78%	38,69%	134,17%
22/02/2013	30	-10,47%	-6,37%	42,93%	1,61%	-2,27%	5,09%	1,61%	4,15%	7,64%
22/02/2013	31	2,94%	7,31%	62,47%	17,98%	13,32%	20,80%	17,98%	20,83%	22,20%
03/06/2013	32	-19,31%	-11,81%	50,64%	0,43%	-10,25%	1,94%	50,64%	-3,24%	46,77%
02/07/2013	33	2,32%	11,28%	88,67%	25,75%	13,90%	28,38%	88,67%	22,35%	83,22%
02/07/2013	34	1,79%	6,21%	51,57%	13,22%	6,46%	15,85%	13,22%	19,51%	13,34%
24/09/2013	35	10,79%	25,78%	135,09%	48,10%	27,90%	49,53%	107,57%	40,32%	120,98%
17/01/2014	36	-8,44%	-2,46%	47,55%	7,26%	-2,88%	8,21%	7,26%	5,68%	12,53%
22/03/2014	37	2,25%	7,35%	53,54%	15,03%	6,34%	16,90%	15,03%	24,91%	15,04%
17/06/2014	38	3,74%	8,53%	57,31%	15,74%	9,22%	18,91%	15,74%	24,06%	15,94%
04/09/2014	39	-17,33%	-9,10%	50,86%	2,63%	-9,86%	3,44%	50,86%	-1,81%	42,97%
					-					
03/10/2014 09/11/2014	40	-6,75%	-2,40% 6.33%	42,48%	4,59% 5,97%	-1,44%	7,30%	4,59% 53,35%	7,35% 2,27%	5,37%
	41	-12,90% 22,60%	-6,33% 20,85%	53,35%	5,97%	-3,71%	7,28%	*	53,30%	49,62%
02/01/2015	42	-	29,85%	110,75%	50,59%	37,63%	50,41%	50,59%		51,40%
23/02/2015	43	3,75%	7,34%	47,25%	13,25%	7,37%	15,79%	13,25%	19,84%	13,29%
04/04/2015	44	-10,97%	-3,83%	56,84%	7,80%	-2,81%	9,41%	56,84%	4,75%	43,34%
31/05/2015	45	-2,75%	12,19%	116,65%	34,11%	13,61%	34,76%	116,65%	26,42%	115,55%
20/06/2015	46	-4,73%	4,67% 7,33%	77,01% 80,81%	18,51% 21,66%	5,10% 7,09%	20,11% 22,87%	77,01% 80,81%	14,02% 16,43%	70,80% 74,19%

TABELA G.2 – ERROS RELATIVOS DAS VAZÕES MÁXIMAS PREVISTAS DOS EVENTOS – 2005 A 2015

Data do pico	ID	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Previsão média	Previsão por grupo	Previsão Euclidiana	Previsão Pertenc.
26/05/2005	1	1,05%	1,70%	16,39%	3,38%	-0,07%	4,45%	16,39%	1,57%	16,38%
20/06/2005	2	-0,53%	-0,47%	9,23%	-1,09%	-1,62%	1,10%	9,23%	-1,45%	9,23%
20/09/2005	3	0,62%	0,69%	0,57%	-0,68%	1,17%	0,44%	1,17%	0,02%	1,05%
26/02/2007	4	-1,08%	-1,09%	10,26%	-2,67%	-0,90%	0,90%	10,26%	-0,11%	4,53%
19/03/2007	5	-0,47%	0,54%	20,42%	1,58%	0,00%	4,41%	1,58%	3,00%	2,01%
28/04/2007	6	-2,71%	-1,41%	11,40%	-0,01%	-6,18%	0,22%	11,40%	-3,09%	11,38%
24/05/2007	7	0,15%	0,39%	3,54%	1,50%	-0,60%	0,79%	-0,60%	2,07%	0,01%
18/10/2007	8	0,34%	1,85%	22,70%	4,43%	-0,31%	5,80%	22,70%	1,51%	22,69%
15/11/2007	9	1,22%	1,57%	13,42%	4,14%	2,43%	4,56%	4,14%	5,26%	6,61%
12/12/2007	10	3,06%	3,54%	18,85%	3,84%	4,58%	6,77%	18,85%	4,60%	18,84%
07/05/2008	11	1,64%	1,68%	8,09%	-0,38%	2,82%	2,77%	-0,38%	4,53%	-0,38%
18/08/2008	12	-1,26%	-1,29%	12,30%	-2,10%	-0,90%	1,33%	-2,10%	1,23%	-2,07%
31/10/2008	13	-1,94%	-2,17%	-0,34%	-3,80%	-4,00%	-2,45%	-0,34%	-3,78%	-0,34%
04/08/2009	14	1,87%	1,85%	8,62%	2,15%	1,78%	3,25%	8,62%	2,45%	8,61%
03/10/2009	15	3,32%	5,66%	5,27%	4,11%	8,04%	4,94%	8,04%	4,83%	7,43%
04/12/2009	16	4,29%	3,96%	8,81%	-0,07%	7,89%	4,98%	-0,07%	7,51%	-0,03%
29/01/2010	17	3,65%	4,31%	3,51%	0,08%	6,65%	3,63%	4,27%	3,47%	3,11%
07/04/2010	18	-0,29%	-0,86%	-0,28%	-4,83%	0,07%	-1,24%	-4,83%	-0,33%	-4,83%
08/05/2010	19	-3,22%	-2,78%	-5,30%	-4,70%	-2,84%	-3,80%	-2,84%	-4,81%	-2,94%
23/06/2010	20	3,11%	4,79%	26,88%	7,80%	3,69%	9,25%	7,80%	9,79%	7,80%
21/07/2010	21	2,44%	2,42%	13,17%	1,80%	3,40%	4,59%	13,17%	3,75%	12,86%
25/02/2011	22	-0,24%	-0,37%	-1,81%	-4,55%	0,03%	-1,53%	0,03%	-2,46%	-1,72%
03/04/2011	23	0,74%	0,61%	6,36%	-0,91%	0,31%	1,42%	-0,91%	2,17%	-0,91%
10/09/2011	24	-0,23%	0,29%	1,72%	0,89%	-0,30%	0,44%	-0,30%	1,15%	-0,29%
20/10/2011	25	3,84%	4,07%	14,73%	5,76%	7,00%	6,95%	5,76%	9,51%	6,66%
01/05/2012	26	2,23%	1,97%	10,32%	1,95%	2,78%	3,85%	10,32%	3,01%	10,30%
20/06/2012	27	-1,37%	-1,63%	-3,30%	-4,35%	-1,75%	-2,48%	-4,35%	-3,01%	-4,13%
02/08/2012	28	0,86%	0,52%	8,67%	2,11%	0,07%	2,45%	2,11%	3,61%	2,12%
10/01/2013	29	-2,58%	-2,06%	13,46%	-1,64%	-3,45%	0,75%	13,46%	-2,43%	13,45%
22/02/2013	30	-1,04%	-0,61%	18,45%	-0,64%	0,34%	3,30%	-0,64%	2,64%	0,34%
22/03/2013	31	-3,37%	-3,53%	8,05%	-4,87%	-3,80%	-1,50%	-4,87%	-2,19%	-4,47%
03/06/2013	32	1,39%	1,76%	19,37%	3,34%	2,12%	5,60%	19,37%	3,07%	19,26%
02/07/2013	33	-0,47%	-0,20%	-3,05%	-2,25%	0,07%	-1,18%	0,07%	-2,42%	0,07%
02/08/2013	34	-2,17%	-2,97%	1,11%	-4,09%	-2,41%	-2,51%	-4,09%	-0,82%	-4,05%
24/09/2013	35	0,77%	0,74%	2,49%	0,99%	1,12%	0,71%	1,12%	1,32%	1,09%
17/01/2014	36	-3,07%	-3,06%	7,04%	-3,15%	-5,11%	-1,47%	-3,15%	-3,09%	-2,92%
22/03/2014	37	0,09%	-0,24%	9,28%	0,21%	-0,88%	1,69%	0,21%	3,09%	0,21%
17/06/2014	38	-3,41%	-3,68%	-0,98%	1,51%	-3,10%	-2,21%	-3,10%	-0,90%	-3,01%
04/09/2014	39	-1,13%	-0,78%	12,55%	-0,70%	-2,07%	1,57%	12,55%	-1,19%	12,30%
03/10/2014	40	-0,44%	-0,46%	2,04%	-1,03%	-2,16%	-0,41%	-1,03%	-0,37%	-1,03%
09/11/2014	41	2,05%	2,86%	22,12%	4,80%	2,60%	6,88%	22,12%	3,78%	21,94%
02/01/2015	42	-3,14%	-3,56%	0,56%	-5,22%	-5,21%	-3,31%	-5,22%	-3,63%	-5,22%
23/02/2015	43	0,57%	0,98%	13,37%	0,71%	0,25%	3,18%	0,71%	4,99%	0,71%
04/04/2015	44	1,35%	1,94%	21,56%	2,57%	2,89%	5,82%	21,56%	4,05%	18,55%
31/05/2015	45	1,29%	1,78%	19,99%	3,51%	2,00%	5,67%	19,99%	2,65%	19,99%
20/06/2015	46	0,39%	0,49%	8,82%	-0,37%	-0,45%	1,77%	8,82%	-0,14%	8,80%
24/10/2015	47	-1,53%	-1,64%	1,37%	-0,05%	-2,98%	-1,32%	-0,82%	-0,50%	-1,57%

TABELA G.3 – ERROS RELATIVOS DAS VAZÕES MÉDIAS PREVISTAS DOS EVENTOS – 2005 A 2015

Data do		Modelo	Modelo	Modelo	Modelo	Modelo	Previsão	Previsão	Previsão	Previsão
pico	ID	1	2	3	4	5	média	por grupo	Euclidiana	Pertenc.
26/05/2005	1	-1,61%	0,83%	26,12%	4,53%	-1,22%	5,73%	26,12%	1,87%	26,05%
20/06/2005	2	-2,25%	-0,52%	21,77%	1,50%	-1,62%	3,77%	21,77%	-0,25%	21,62%
20/09/2005	3	-0,62%	-0,58%	3,15%	-2,28%	-0,53%	-0,17%	0,80%	-0,73%	1,55%
26/02/2007	4	-1,25%	-0,37%	21,01%	0,28%	0,97%	4,13%	21,01%	2,84%	11,77%
19/03/2007	5	-0,16%	1,80%	32,64%	5,06%	3,23%	8,51%	5,06%	7,68%	6,17%
28/04/2007	6	-5,13%	-1,49%	29,16%	3,45%	-4,57%	4,28%	29,16%	0,43%	29,02%
24/05/2007	7	-1,65%	-1,44%	6,49%	-1,64%	-2,30%	-0,11%	-1,57%	0,52%	-1,47%
18/10/2007	8	-2,27%	1,44%	36,07%	7,08%	-0,31%	8,40%	36,07%	3,21%	35,93%
15/11/2007	9	0,33%	0,97%	15,69%	1,88%	0,74%	3,92%	1,88%	3,13%	5,02%
12/12/2007	10	0,99%	3,21%	33,03%	7,05%	3,96%	9,65%	33,03%	5,58%	32,97%
07/05/2008	11	-0,44%	-0,02%	13,88%	-0,67%	0,44%	2,64%	-0,67%	4,32%	-0,63%
18/08/2008	12	-1,99%	-0,78%	22,95%	0,61%	-0,27%	4,10%	0,61%	3,61%	0,97%
31/10/2008	13	-0,81%	-0,77%	7,96%	-0,89%	-1,41%	0,82%	7,96%	-0,85%	7,95%
04/08/2009	14	-1,30%	-0,68%	15,43%	-0,36%	-0,18%	2,58%	15,43%	0,45%	15,41%
03/10/2009	15	-0,42%	-0,19%	2,90%	-1,81%	-0,16%	0,06%	-0,82%	0,44%	-0,53%
04/12/2009	16	2,24%	2,64%	18,35%	1,69%	5,48%	6,08%	1,69%	9,10%	2,29%
29/01/2010	17	3,58%	3,95%	17,52%	3,69%	7,13%	7,17%	3,94%	11,53%	4,34%
07/04/2010	18	1,83%	1,93%	14,81%	2,50%	3,56%	4,93%	2,50%	8,03%	2,52%
08/05/2010	19	-0,14%	0,19%	5,81%	0,08%	0,43%	1,27%	0,56%	2,92%	0,58%
23/06/2010	20	3,04%	5,20%	35,59%	9,22%	5,85%	11,78%	9,22%	13,40%	9,25%
21/07/2010	21	1,76%	2,27%	19,78%	3,12%	3,40%	6,07%	19,78%	4,43%	19,22%
25/02/2011	22	0,99%	1,05%	8,62%	-0,47%	1,94%	2,42%	0,45%	3,75%	0,62%
03/04/2011	23	0,45%	0,90%	15,79%	0,81%	1,30%	3,85%	0,81%	5,25%	0,82%
10/09/2011	24	-1,26%	-1,30%	2,00%	-2,34%	-1,70%	-0,92%	-1,80%	-0,67%	-1,59%
20/10/2011	25	1,75%	2,34%	18,30%	3,01%	3,51%	5,78%	3,01%	7,56%	3,25%
01/05/2012	26	-2,34%	-1,59%	14,36%	-0,89%	-1,49%	1,61%	14,36%	-0,11%	14,29%
20/06/2012	27	0,14%	0,34%	5,42%	-0,95%	1,14%	1,22%	-0,16%	2,42%	-0,11%
02/08/2012	28	0,72%	1,59%	20,64%	3,65%	1,40%	5,60%	3,65%	7,47%	3,67%
10/01/2013	29	1,83%	4,01%	37,14%	9,21%	5,79%	11,60%	37,14%	7,96%	37,11%
22/02/2013	30	-0,56%	1,01%	29,31%	3,36%	2,48%	7,12%	3,36%	6,06%	5,44%
22/02/2013	31	-0,46%	0,44%	22,51%	1,40%	1,56%	5,09%	1,40%	4,52%	2,24%
03/06/2013	32	-1,78%	0,4476	27,66%	3,29%	-0,17%	5,86%	27,66%	1,98%	27,03%
02/07/2013	33		-0,89%	1,05%	-2,44%	-0,17%	-0,88%	1,16%	-1,53%	1,00%
02/07/2013		-1,36% 0,74%	1,14%	1,05%	1,34%	1,99%	3,62%		6,40%	1,42%
24/09/2013	34 35	-1,45%	-1,34%	6,55%	-1,61%	-1,19%	0,19%	1,34% 2,73%	-0,23%	
										3,32% 2,84%
17/01/2014	36	-2,73%	-0,89%	22,23%	1,75%	-2,88%	3,50%	1,75%	1,30%	
22/03/2014	37	0,50% -2,84%	1,02% -2,57%	18,54%	2,82%	0,77%	4,73%	2,82%	7,21%	2,84%
17/06/2014	38			1,73%	-0,31%	-3,02%	-1,40%	-2,95%	0,07%	-2,83%
04/09/2014	39	-1,31%	0,81%	27,72%	3,89%	0,05%	6,23%	27,72%	2,34%	26,37%
03/10/2014	40	-2,75%	-2,56%	6,42%	-3,17%	-3,73%	-1,16%	-3,17%	-0,71%	-3,11%
09/11/2014	41	0,75%	2,89%	34,45%	7,34%	4,05%	9,90%	34,45%	6,23%	33,71%
02/01/2015	42	-1,60%	-1,16%	12,31%	-0,75%	-2,21%	1,32%	-0,75%	1,64%	-0,72%
23/02/2015	43	1,83%	3,22%	26,68%	5,21%	3,76%	8,14%	5,21%	11,55%	5,22%
04/04/2015	44	0,91%	2,93%	32,07%	6,26%	3,75%	9,19%	32,07%	6,25%	27,41%
31/05/2015	45	-1,27%	1,41%	31,60%	5,38%	0,96%	7,62%	31,60%	3,12%	31,52%
20/06/2015	46	-0,87%	0,46%	19,46%	2,12%	-0,77%	4,08%	19,46%	0,66%	19,06%
24/10/2015	47	-1,56%	-1,65%	3,73%	-2,39%	-2,37%	-0,85%	3,30%	-1,34%	2,20%

TABELA G.4 – ERROS RELATIVOS DOS DESVIOS-PADRÃO DAS VAZÕES PREVISTAS DOS EVENTOS – 2005 A 2015

Data do pico	ID	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Previsão média	Previsão por grupo	Previsão Euclidiana	Previsão Pertenc.
26/05/2005	1	6,68%	1,16%	-21,50%	-5,60%	-2,16%	-4,39%	-21,50%	-5,59%	-21,14%
20/06/2005	2	2,78%	-2,53%	-26,55%	-11,69%	-3,58%	-8,33%	-26,55%	-7,81%	-25,87%
20/09/2005	3	1,50%	0,18%	-15,16%	-4,71%	0,40%	-3,66%	-4,92%	-4,68%	-9,60%
26/02/2007	4	-1,14%	-5,08%	-30,00%	-15,22%	-7,65%	-11,87%	-30,00%	-10,99%	-24,26%
19/03/2007	5	-3,53%	-6,81%	-32,21%	-16,34%	-16,13%	-15,11%	-16,34%	-19,45%	-20,40%
28/04/2007	6	3,20%	-1,43%	-24,03%	-8,66%	-7,87%	-7,80%	-24,03%	-10,09%	-23,66%
24/05/2007	7	-2,34%	-3,98%	-16,61%	-6,57%	-4,52%	-6,88%	-5,86%	-7,45%	-6,36%
18/10/2007	8	16,22%	8,75%	-19,19%	-0,78%	6,14%	2,19%	-19,19%	0,25%	-18,41%
15/11/2007	9	4,72%	0,15%	-19,24%	-6,18%	2,75%	-3,68%	-6,18%	0,64%	-10,84%
12/12/2007	10	14,57%	7,91%	-17,26%	-1,16%	10,61%	2,80%	-17,26%	4,84%	-16,72%
07/05/2008	11	-3,23%	-7,05%	-28,89%	-16,10%	-5,31%	-12,16%	-16,10%	-11,23%	-16,29%
18/08/2008	12	-2,92%	-7,73%	-31,25%	-16,99%	-8,67%	-13,55%	-16,99%	-12,70%	-19,00%
31/10/2008	13	0,94%	-2,04%	-19,43%	-7,17%	-3,62%	-6,35%	-19,43%	-5,61%	-19,29%
04/08/2009	14	2,19%	-1,23%	-21,44%	-8,02%	-2,04%	-6,19%	-21,44%	-5,14%	-21,36%
03/10/2009	15	0,71%	0,29%	-13,69%	-3,64%	1,46%	-3,05%	0,48%	-5,91%	-0,74%
04/12/2009	16	9,41%	9,68%	-8,21%	5,17%	10,39%	4,19%	5,17%	2,30%	7,80%
29/01/2010	17	4,90%	4,95%	-10,67%	0,87%	7,27%	1,29%	2,74%	-2,53%	3,43%
07/04/2010	18	-0,96%	-3,67%	-20,50%	-8,40%	-2,63%	-7,43%	-8,40%	-8,55%	-8,38%
08/05/2010	19	-1,90%	-1,95%	-11,05%	-2,46%	-1,91%	-3,88%	-2,21%	-6,51%	-2,23%
23/06/2010	20	7,44%	4,22%	-13,09%	0,43%	1,81%	0,08%	0,43%	-0,64%	0,27%
21/07/2010	21	4,50%	0,30%	-22,38%	-7,69%	0,53%	-5,02%	-22,38%	-2,64%	-20,28%
25/02/2011	22	-1,06%	-2,29%	-19,27%	-8,21%	-2,56%	-6,73%	-4,80%	-9,71%	-6,38%
03/04/2011	23	-2,68%	-5,61%	-26,38%	-13,56%	-6,99%	-11,08%	-13,56%	-11,26%	-13,62%
10/09/2011	24	-1,66%	-2,38%	-11,49%	-2,96%	-1,83%	-4,17%	-2,52%	-4,34%	-3,51%
20/10/2011	25	1,91%	-1,14%	-19,64%	-6,84%	1,67%	-4,93%	-6,84%	-2,98%	-6,67%
01/05/2012	26	6,92%	1,30%	-22,11%	-6,90%	1,87%	-3,87%	-22,11%	-3,45%	-21,84%
20/06/2012	27	-1,86%	-3,25%	-19,85%	-9,10%	-3,50%	-7,56%	-7,26%	-11,13%	-7,50%
02/08/2012	28	-2,17%	-5,26%	-20,12%	-7,75%	-6,63%	-8,44%	-7,75%	-8,68%	-7,77%
10/01/2013	29	-0,65%	-5,72%	-32,75%	-16,56%	-11,71%	-13,71%	-32,75%	-15,45%	-32,52%
22/02/2013	30	4,94%	-0,47%	-30,25%	-13,06%	-3,83%	-8,55%	-13,06%	-8,21%	-18,82%
22/03/2013	31	-9,42%	-14,01%	-39,85%	-25,41%	-16,74%	-21,14%	-25,41%	-20,69%	-27,50%
03/06/2013	32	5,41%	-0,63%	-26,62%	-10,33%	-1,22%	-6,72%	-26,62%	-5,62%	-24,19%
02/07/2013	33	0,02%	-0,44%	-13,79%	-4,27%	-0,24%	-3,77%	-6,95%	-5,23%	-6,57%
02/08/2013	34	-0,79%	-3,71%	-21,25%	-9,02%	-0,44%	-7,17%	-9,02%	-8,29%	-8,83%
24/09/2013	35	0,04%	-1,84%	-15,18%	-4,55%	-1,86%	-4,76%	-11,49%	-4,28%	-12,13%
17/01/2014	36	-5,87%	-10,31%	-31,43%	-17,54%	-13,30%	-15,75%	-17,54%	-16,52%	-19,49%
22/03/2014	37	-3,13%	-7,86%	-25,71%	-12,08%	-7,37%	-11,31%	-12,08%	-12,50%	-12,16%
17/06/2014	38	-3,96%	-4,43%	-8,65%	-0,69%	-3,18%	-4,22%	-3,52%	-4,46%	-3,48%
04/09/2014	39	3,19%	-3,04%	-28,36%	-12,56%	-3,69%	-8,91%	-28,36%	-8,48%	-23,11%
03/10/2014	40	-2,17%	-4,27%	-20,35%	-9,13%	-6,37%	-8,50%	-9,13%	-9,27%	-9,29%
09/11/2014	41	10,12%	4,21%	-24,51%	-7,36%	0,81%	-3,52%	-24,51%	-3,68%	-21,48%
02/01/2015	42	-7,44%	-10,89%	-31,94%	-18,69%	-13,77%	-16,57%	-18,69%	-17,97%	-18,83%
23/02/2015	43	-7,48%	-10,85%	-32,94%	-19,73%	-14,24%	-17,09%	-19,73%	-19,12%	-19,78%
04/04/2015	44	9,80%	3,46%	-23,38%	-6,93%	4,85%	-2,53%	-23,38%	-0,17%	-15,74%
31/05/2015	45	11,85%	4,88%	-21,80%	-5,11%	5,04%	-1,09%	-21,80%	-0,95%	-21,34%
20/06/2015	46	2,53%	-2,04%	-24,05%	-9,80%	-3,21%	-7,34%	-24,05%	-6,48%	-22,09%
24/10/2015	47	-0,14%	-1,80%	-13,30%	-2,96%	-3,39%	-4,43%	-15,89%	-2,53%	-14,70%

TABELA G.5 – ERROS RELATIVOS DAS ASSIMETRIAS DAS VAZÕES PREVISTAS DOS EVENTOS – 2005 A 2015

Data do pico	ID	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Previsão média	Previsão por grupo	Previsão Euclidiana	Previsão Pertenc.
26/05/2005	1	-6,58%	-10,91%	-29,18%	-20,47%	-10,88%	-14,35%	-29,18%	-16,52%	-28,31%
20/06/2005	2	34,72%	18,05%	-30,70%	-1,28%	15,72%	9,88%	-30,70%	-8,55%	-20,56%
20/09/2005	3	-2,82%	-9,35%	-32,07%	-19,68%	-11,80%	-14,24%	-29,43%	-13,86%	-31,97%
26/02/2007	4	11,79%	10,02%	-4,12%	8,39%	2,89%	6,58%	-4,12%	0,60%	-1,80%
19/03/2007	5	52,36%	30,66%	-443,15%	-149,52%	-148,21%	-116,16%	-149,52%	-219,53%	-362,69%
28/04/2007	6	-7,08%	-19,93%	-68,32%	-39,98%	-15,52%	-27,70%	-68,32%	-31,28%	-65,58%
24/05/2007	7	134,78%	202,70%	585,17%	406,37%	220,61%	296,45%	265,54%	372,67%	322,03%
18/10/2007	8	114,27%	97,26%	49,65%	73,58%	110,36%	91,86%	49,65%	85,12%	55,95%
15/11/2007	9	-0,93%	-4,55%	-27,42%	-16,41%	-7,53%	-9,71%	-16,41%	-12,72%	-28,60%
12/12/2007	10	20,43%	3,08%	-39,48%	-11,26%	-7,83%	-4,84%	-39,48%	-26,94%	-32,41%
07/05/2008	11	-27,18%	-33,29%	-50,71%	-39,54%	-34,17%	-35,97%	-39,54%	-35,43%	-40,13%
18/08/2008	12	-50,31%	-62,97%	-116,30%	-81,19%	-67,41%	-73,24%	-81,19%	-75,11%	-106,38%
31/10/2008	13	12,10%	5,83%	-9,10%	0,18%	11,07%	4,78%	-9,10%	6,15%	-8,50%
04/08/2009	14	-38,81%	-21,01%	42,12%	10,47%	-27,61%	-9,72%	42,12%	-7,05%	40,55%
03/10/2009	15	-14,81%	-26,37%	-62,03%	-45,43%	-35,53%	-35,44%	-48,85%	-40,31%	-49,59%
04/12/2009	16	3,20%	46,81%	274,15%	208,84%	-282,62%	50,66%	208,84%	44,63%	180,76%
29/01/2010	17	0,06%	1,93%	-2,18%	-4,73%	-2,79%	-1,66%	0,58%	-5,51%	-0,90%
07/04/2010	18	16,78%	15,16%	34,57%	36,12%	16,99%	26,63%	36,12%	37,51%	36,07%
08/05/2010	19	0,74%	4,03%	22,08%	14,32%	9,93%	9,86%	11,28%	14,35%	11,29%
23/06/2010	20	-13,23%	-30,32%	-96,56%	-68,17%	-17,75%	-41,99%	-68,17%	-23,16%	-69,70%
21/07/2010	21	-0,03%	-2,41%	-18,07%	-9,52%	-8,46%	-6,83%	-18,07%	-9,90%	-12,18%
25/02/2011	22	211,25%	411,15%	1201,44%	801,11%	493,38%	587,53%	1280,68%	493,22%	1209,16%
03/04/2011	23	-41,58%	-53,77%	-113,41%	-82,53%	-59,15%	-67,43%	-82,53%	-58,63%	-83,22%
10/09/2011	24	-15,50%	-28,63%	-135,04%	-90,97%	-37,37%	-58,53%	-56,94%	-78,77%	-64,32%
20/10/2011	25	-33,30%	-42,03%	-111,41%	-79,94%	-66,72%	-62,24%	-79,94%	-63,55%	-87,61%
01/05/2012	26	-7,86%	-11,87%	-31,32%	-20,44%	-15,99%	-16,33%	-31,32%	-18,36%	-30,66%
20/06/2012	27	13,21%	8,50%	2,74%	14,30%	22,07%	12,49%	9,42%	27,63%	12,79%
02/08/2012	28	13,86%	17,14%	42,88%	31,09%	21,00%	24,09%	31,09%	25,37%	31,27%
10/01/2013	29	19,95%	16,98%	0,56%	12,29%	19,78%	15,65%	0,56%	13,43%	1,43%
22/02/2013	30	169,65%	173,95%	185,04%	179,43%	151,41%	171,03%	179,43%	153,29%	145,88%
22/03/2013	31	7,52%	5,07%	-13,44%	-6,26%	6,34%	0,85%	-6,26%	3,12%	-13,45%
03/06/2013	32	21,05%	19,32%	-23,91%	2,23%	-6,58%	4,83%	-23,91%	-12,50%	-9,10%
02/07/2013		2.13%			ŕ		ĺ		ŕ	
	33	, -	-1,01%	-15,77%	-7,87%	-3,28%	-4,90% 2,220/	-25,45%	0,85%	-24,96%
02/08/2013	34	9,10%	7,99%	-14,35%	-1,40%	-1,18%	3,23%	-1,40%	3,99%	-2,22%
24/09/2013	35	-25,95%	-49,00%	-187,28%	-127,89%	-66,06%	-86,36%	-139,59%	-97,97%	-149,89%
17/01/2014	36	0,99%	1,90%	13,77%	6,06%	12,63%	6,55%	6,06%	12,68%	9,85%
22/03/2014	37	-2,11%	-4,99%	-18,48%	-11,56%	-1,95%	-6,80%	-11,56%	-3,50%	-11,71%
17/06/2014	38	17,12%	20,47%	86,29%	70,11%	44,36%	47,03%	51,89%	66,95%	51,58%
04/09/2014	39	356,09%	293,94%	118,54%	224,13%	313,33%	272,38%	118,54%	201,99%	371,86%
03/10/2014	40	-56,58%	-76,27%	-163,98%	-118,04%	-73,31%	-94,31%	-118,04%	-86,39%	-119,79%
09/11/2014	41	-195,44%	-160,85%	-35,02%	-94,45%	-191,70%	-141,66%	-35,02%	-150,60%	-90,31%
02/01/2015	42	-8,80%	-12,12%	-28,95%	-19,31%	-16,47%	-16,56%	-19,31%	-14,70%	-20,16%
23/02/2015	43	10,94%	15,81%	29,70%	24,01%	4,73%	16,68%	24,01%	5,45%	24,18%
04/04/2015	44	29,11%	23,22%	-9,65%	9,38%	16,80%	16,08%	-9,65%	7,34%	17,22%
31/05/2015	45	-7,74%	-10,66%	-28,94%	-17,92%	-18,70%	-15,79%	-28,94%	-24,05%	-27,45%
20/06/2015	46	8,34%	3,72%	-12,32%	-2,53%	-0,99%	0,06%	-12,32%	-7,63%	-0,88%
24/10/2015	47	20,61%	7,80%	-90,28%	-55,57%	12,09%	-18,49%	-57,33%	-15,46%	-84,43%

TABELA G.6 – ERROS RELATIVOS DOS VOLUMES PREVISTOS DOS EVENTOS – 2005 A 2015

Data do pico	ID	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Previsão média	Previsão por grupo	Previsão Euclidiana	Previsão Pertenc.
26/05/2005	1	-1,61%	0,83%	26,12%	4,53%	-1,22%	5,73%	26,12%	1,87%	26,05%
20/06/2005	2	-2,25%	-0,52%	21,77%	1,50%	-1,62%	3,77%	21,77%	-0,25%	21,62%
20/09/2005	3	-0,62%	-0,58%	3,15%	-2,28%	-0,53%	-0,17%	0,80%	-0,73%	1,55%
26/02/2007	4	-1,25%	-0,37%	21,01%	0,28%	0,97%	4,13%	21,01%	2,84%	11,77%
19/03/2007	5	-0,16%	1,80%	32,64%	5,06%	3,23%	8,51%	5,06%	7,68%	6,17%
28/04/2007	6	-5,13%	-1,49%	29,16%	3,45%	-4,57%	4,28%	29,16%	0,43%	29,02%
24/05/2007	7	-1,65%	-1,44%	6,49%	-1,64%	-2,30%	-0,11%	-1,57%	0,52%	-1,47%
18/10/2007	8	-2,27%	1,44%	36,07%	7,08%	-0,31%	8,40%	36,07%	3,21%	35,93%
15/11/2007	9	0,33%	0,97%	15,69%	1,88%	0,74%	3,92%	1,88%	3,13%	5,02%
12/12/2007	10	0,99%	3,21%	33,03%	7,05%	3,96%	9,65%	33,03%	5,58%	32,97%
07/05/2008	11	-0,44%	-0,02%	13,88%	-0,67%	0,44%	2,64%	-0,67%	4,32%	-0,63%
18/08/2008	12	-1,99%	-0,78%	22,95%	0,61%	-0,27%	4,10%	0,61%	3,61%	0,97%
31/10/2008	13	-0,81%	-0,77%	7,96%	-0,89%	-1,41%	0,82%	7,96%	-0,85%	7,95%
04/08/2009	14	-1,30%	-0,68%	15,43%	-0,36%	-0,18%	2,58%	15,43%	0,45%	15,41%
03/10/2009	15	-0,42%	-0,19%	2,90%	-1,81%	-0,16%	0,06%	-0,82%	0,44%	-0,53%
04/12/2009	16	2,24%	2,64%	18,35%	1,69%	5,48%	6,08%	1,69%	9,10%	2,29%
29/01/2010	17	3,58%	3,95%	17,52%	3,69%	7,13%	7,17%	3,94%	11,53%	4,34%
07/04/2010	18	1,83%	1,93%	14,81%	2,50%	3,56%	4,93%	2,50%	8,03%	2,52%
08/05/2010	19	-0,14%	0,19%	5,81%	0,08%	0,43%	1,27%	0,56%	2,92%	0,58%
23/06/2010	20	3,04%	5,20%	35,59%	9,22%	5,85%	11,78%	9,22%	13,40%	9,25%
21/07/2010	21	1,76%	2,27%	19,78%	3,12%	3,40%	6,07%	19,78%	4,43%	19,22%
25/02/2011	22	0,99%	1,05%	8,62%	-0,47%	1,94%	2,42%	0,45%	3,75%	0,62%
03/04/2011	23	0,45%	0,90%	15,79%	0,81%	1,30%	3,85%	0,81%	5,25%	0,82%
10/09/2011	24	-1,26%	-1,30%	2,00%	-2,34%	-1,70%	-0,92%	-1,80%	-0,67%	-1,59%
20/10/2011	25	1,75%	2,34%	18,30%	3,01%	3,51%	5,78%	3,01%	7,56%	3,25%
01/05/2012	26	-2,34%	-1,59%	14,36%	-0,89%	-1,49%	1,61%	14,36%	-0,11%	14,29%
20/06/2012	27	0,14%	0,34%	5,42%	-0,95%	1,14%	1,22%	-0,16%	2,42%	-0,11%
02/08/2012	28	0,72%	1,59%	20,64%	3,65%	1,40%	5,60%	3,65%	7,47%	3,67%
10/01/2013	29	1,83%	4,01%	37,14%	9,21%	5,79%	11,60%	37,14%	7,96%	37,11%
22/02/2013	30	-0,56%	1,01%	29,31%	3,36%	2,48%	7,12%	3,36%	6,06%	5,44%
22/03/2013	31	-0,46%	0,44%	22,51%	1,40%	1,56%	5,09%	1,40%	4,52%	2,24%
03/06/2013	32	-1,78%	0,28%	27,66%	3,29%	-0,17%	5,86%	27,66%	1,98%	27,03%
02/07/2013	33	-1,36%	-0,89%	1,05%	-2,44%	-0,78%	-0,88%	1,16%	-1,53%	1,00%
02/08/2013	34	0,74%	1,14%	12,90%	1,34%	1,99%	3,62%	1,34%	6,40%	1,42%
24/09/2013	35	-1,45%	-1,34%	6,55%	-1,61%	-1,19%	0,19%	2,73%	-0,23%	3,32%
17/01/2014	36	-2,73%	-0,89%	22,23%	1,75%	-2,88%	3,50%	1,75%	1,30%	2,84%
22/03/2014	37	0,50%	1,02%	18,54%	2,82%	0,77%	4,73%	2,82%	7,21%	2,84%
17/06/2014	38	-2,84%	-2,57%	1,73%	-0,31%	-3,02%	-1,40%	-2,95%	0,07%	-2,83%
04/09/2014	39	-1,31%	0,81%	27,72%	3,89%	0,05%	6,23%	27,72%	2,34%	26,37%
03/10/2014	40	-2,75%	-2,56%	6,42%	-3,17%	-3,73%	-1,16%	-3,17%	-0,71%	-3,11%
09/11/2014	41	0,75%	2,89%	34,45%	7,34%	4,05%	9,90%	34,45%	6,23%	33,71%
02/01/2015	42	-1,60%	-1,16%	12,31%	-0,75%	-2,21%	1,32%	-0,75%	1,64%	-0,72%
23/02/2015	43	1,83%	3,22%	26,68%	5,21%	3,76%	8,14%	5,21%	11,55%	5,22%
04/04/2015	44	0,91%	2,93%	32,07%	6,26%	3,75%	9,19%	32,07%	6,25%	27,41%
31/05/2015	45	-1,27%	1,41%	31,60%	5,38%	0,96%	7,62%	31,60%	3,12%	31,52%
20/06/2015	46	-0,87%	0,46%	19,46%	2,12%	-0,77%	4,08%	19,46%	0,66%	19,06%
24/10/2015	47	-1,56%	-1,65%	3,73%	-2,39%	-2,37%	-0,85%	3,30%	-1,34%	2,20%

ANEXO 1 – AVALIAÇÃO DO COMPORTAMENTO DE MODELOS DE PREVISÃO DE CHEIAS NO RIO IGUAÇU, EM UNIÃO DA VITÓRIA, PARANÁ, UTILIZANDO O AJUSTE COM BASE EM AGRUPAMENTO

Trabalho completo publicado nos Anais do XXIV Simpósio Brasileiro de Recursos Hídricos.





XXIV SIMPÓSIO BRASILEIRO DE RECURSOS HIDRÍCOS

Avaliação do comportamento de modelos de previsão de cheias no rio Iguaçu, em União da Vitória, Paraná, utilizando o ajuste com base em agrupamento

Patrícia Cristina Steffen ¹; Nathalli Rogiski da Silva ²& Julio Gomes ³

Palavras-Chave – Previsão de cheias. Agrupamento de cheias. Horizontes de previsão.

INTRODUÇÃO

Desastres naturais podem ocasionar sérios impactos sociais, ambientais e financeiros (Nedel e Saunsen, 2012). Dentre eles, as cheias geram os maiores impactos sobre a sociedade, pois apresentam um maior número de ocorrências e produzem elevados números de mortos, feridos e desabrigados (Chu *et al.*, 2020). O Brasil, por sua vez, ocupa a 11ª posição entre os países com as maiores populações expostas a cheias urbanas (Verwey *et al.*, 2017).

As causas naturais das cheias são variadas, e, em primeiro lugar, estão os eventos hidrológicos intensos, com ocorrência rápida, seguidos de tsunamis, furacões, ciclones e outros (Freitas e Ximenes, 2012). Por outro lado, Castilho *et al.* (2005) afirmam que a ocupação irregular e desordenada de áreas ribeirinhas é a principal causa de inundações devido às ações antrópicas. Contudo, muitas vezes os fenômenos atuam simultaneamente e, independentemente de sua origem, as consequências podem ser, por vezes, catastróficas para as regiões onde ocorrem, fazendo-se necessário adotar medidas mitigatórias a fim de minimizar os impactos ocasionados.

Neste contexto, estão inseridos os sistemas de previsão e alerta de cheias, com o intuito de diminuir a implantação de medidas estruturais de mitigação e de proporcionar adequada relação entre a população e os eventos extremos máximos. Em outras palavras, são medidas economicamente mais viáveis e ambientalmente sustentáveis (Verwey *et al.*, 2017).

O presente trabalho visa avaliar o comportamento de modelos de previsão de cheias, no rio Iguaçu, em União da Vitória, no Estado do Paraná, com base no ajuste por grupos de eventos com características semelhantes. Entende-se que, utilizando-se a análise de grupos e aplicando-a a um modelo hidrológico, seria possível calibrar o modelo para cada grupo de eventos de cheia, e, a partir dos parâmetros calibrados, efetuar a previsão de vazões, obtendo-se melhor desempenho do modelo em um conjunto de eventos similar ao grupo utilizado para a calibração.

_

¹⁾ Mestra, Universidade Federal de Mato Grosso, patríciacsteffen@yahoo.com.br

²⁾ Mestra, Universidade Federal do Paraná, nrogiski@gmail.com

³⁾ Doutor, Universidade Federal do Paraná, jgomes.dhs@ufpr.br





ÁREA DE ESTUDO

A cidade de União da Vitória está localizada na região sudeste do estado do Paraná a cerca de 225 km da capital. A bacia hidrográfica em estudo é definida pela estação fluviométrica União da Vitória (65310000), no rio Iguaçu, apresentando uma área de drenagem de 24.200 km². Em União da Vitória, assim como em Porto União (município em Santa Catarina e vizinho à União da Vitória), houve uma expansão da área urbana em direção à planície de inundação do rio Iguaçu (COBRAPE, 20--), agravando as condições da população diante dos eventos de cheia, por meio das inundações de casas, interrupção de atividades que suportam as cidades e interdição das vias de acesso.

O rio Iguaçu cruza praticamente todo o estado do Paraná, no sentido Leste-Oeste, com uma extensão de 1.320 km, desaguando no rio Paraná (SEMA, 2010). Sua bacia hidrográfica possui uma área de drenagem total de 70.800 km², dividida entre os estados do Paraná e Santa Catarina, no Brasil, e a Argentina, com uma pequena parcela do território da bacia (Rocha, 2012). Na estação fluviométrica União da Vitória (65310000), a vazão média é de cerca de 500 m³/s (20 L/s/km²), tornando-se de quatro a seis vezes maior nos períodos de cheia. A Figura 1 apresenta a localização da bacia do rio Iguaçu dividida entre os estados do Paraná e Santa Catarina, bem como, a localização do município de União da Vitória.

53°0′W 52°0′W 51°0′W 50°0'W 49°0'W 23°0'S 24°0′S 25°0'S Legenda: Brasil 26°0'S Estado do Paraná Estado de Santa Catarina 27°0'S Município de União da Vitória Bacia hidrográfica do rio Iguaçu, no Brasil INFORMAÇÕES: 28°0'S Datum: SIRGAS 2000 100 200 km Local: Bacia do Rio Iguaçu Fonte dos shapefiles: 29°0'S Territórios brasileiros: IBGE, 2015; Bacia Hidrográfica: ANA, 2017.

Figura 1 - Bacia do rio Iguaçu no Brasil e localização da cidade de União da Vitória, Paraná

MATERIAL E MÉTODOS

Steffen e Gomes (2018) apresentam o agrupamento de cheias históricas observadas na estação fluviométrica União da Vitória (65310000) no período de 1931 a 2015, com base em sete indicadores





de cheia, definindo quatro grupos de severidade: baixa (G1); média (G2); alta (G3); e desastres ou catástrofes (G4), apresentados na Tabela 1. Os eventos utilizados para o ajuste e verificação dos modelos de previsão de vazão foram os mesmos utilizados por Steffen e Gomes (2018).

Tabela 1 – Agrupamento de cheias observadas na estação fluviométrica União da Vitória (65310000) (Steffen e Gomes, 2018)

	G	<i>31</i>			<i>G</i> 2				<i>G4</i>	
1931	1945	1963	1986	1937	1975	2007	1932	1982	2001	1935
1933	1949	1964	1991	1939	1976	2008	1938	1984	2005	1957
1934	1951	1967	1994	1948	1979	2011	1946	1987	2009	1983
1936	1952	1968	2002	1953	1980	2012	1947	1989	2010	1992
1940	1956	1969	2003	1961	1981		1950	1990	2013	1998
1941	1958	1974	2006	1965	1988		1954	1993	2015	2014
1942	1959	1977		1966	1996		1955	1995		
1943	1960	1978		1970	1999		1971	1997		
1944	1962	1985		1973	2004		1972	2000		

A Tabela 2 apresenta os parâmetros estatísticos da série de eventos de cheia utilizados no presente estudo, segundo os 4 grupos definidos em Steffen e Gomes (2018), onde observa-se um aumento nos valores de média e desvio-padrão conforme aumenta a severidade dos grupos. Observa-se também que, quando considerados como um único grupo (*Todos*), o conjunto único se situa entre os grupos *G2* e *G3*, em termos de média, e entre os grupos *G3* e *G4*, em termos de desvio-padrão.

Tabela 2 – Parâmetros estatísticos da série de eventos de cheia utilizados no estudo.

Parâmetro	Grupo									
Parametro	Todos	G 1	G 2	<i>G</i> 3	G 4					
Número de eventos	85	33	22	24	6					
Número total de observações	5618	1622	1696	1701	599					
Média (m³/s)	916,8	559,6	816,2	1114,3	1608,1					
Desvio-padrão (m³/s)	675,7	318,4	505,1	639,5	1079,1					

Ajuste e verificação dos modelos de previsão

Com os parâmetros estimados pelo método dos Mínimos Quadrados (Box e Jenkins, 2016), ajustou-se modelos autorregressivos de *ordem 1* (AR(1)) e de *ordem 2* (AR(2)) para cada grupo de eventos (*G1* a *G4*) e para um conjunto único de eventos (*Todos*), que contém os eventos dos grupos *G1* a *G4*, para os horizontes de previsão de 24h, 48h, 72h e 96h. A Tabela 3 apresenta os parâmetros estatísticos dos eventos de cheia utilizados no ajuste dos modelos de previsão, observados no período de 1931 a 1990. Novamente, observa-se que, quando considerados como um único grupo (*Todos*), o conjunto único se situa entre os grupos *G2* e *G3*, em termos de média, e entre os grupos *G3* e *G4*, em termos de desvio-padrão.





Os modelos de previsão de vazão, obtidos na fase de ajuste, foram verificados para um novo conjunto de eventos de cheia, observados no período de 1991 a 2015, considerando-se novamente os 4 grupos (*G1* a *G4*), apresentados na Tabela 1. A Tabela 4 apresenta os parâmetros estatísticos dos eventos de cheia utilizados na verificação dos modelos de previsão, observados no período de 1991 a 2015.

Tabela 3 – Parâmetros estatísticos da série de eventos de cheia utilizados no ajuste dos modelos AR(1) e AR(2) (período: 1931 a 1990)

Parâmetro	Grupo					
Parametro	Todos	<i>G1</i>	G 2	<i>G3</i>	G 4	
Número de eventos	60	28	15	14	3	
Número total de observações	3798	1425	1159	876	338	
Média (m³/s)	854,2	542,0	790,1	1092,0	1775,9	
Desvio-padrão (m³/s)	652,5	306,9	477,6	636,3	1065,7	

Tabela 4 – Parâmetros estatísticos da série de eventos de cheia utilizados na verificação dos modelos AR(1) e AR(2) (período: 1991 a 2015)

Davêm atva	Grupo					
Parâmetro	Todos	<i>G1</i>	G 2	<i>G3</i>	G 4	
Número de eventos	25	5	7	10	3	
Número total de observações	1820	197	537	826	260	
Média (m³/s)	1047,6	687,4	872,7	1138,0	1394,7	
Desvio-padrão (m³/s)	703,8	366,4	555,3	641,6	1055,2	

O desempenho dos modelos não foi avaliado para o conjunto único (*Todos*), porque o objetivo do trabalho é avaliar se determinado modelo, ajustado para um dado grupo de eventos de cheia, apresenta melhor desempenho na fase de verificação para eventos pertencentes ao mesmo grupo utilizado no ajuste. Contudo, é importante, para a análise dos resultados, comparar o valor dos parâmetros estatísticos dos eventos dos grupos *G1* a *G4* com os obtidos para o conjunto *Todos*. Uma vez mais, observa-se que o conjunto *Todos* se situa entre os grupos *G2* e *G3*, em termos de média, e entre os grupos *G3* e *G4*, em termos de desvio-padrão.

O desempenho dos modelos de previsão nas fases de ajuste e verificação foi avaliada pelos coeficientes de Nash-Sutcliffe (NS) e Kling-Gupta (KG) (Gupta *et al.*, 2009; Lawin *et al.*, 2019) e pelo erro percentual absoluto médio (MAPE) (Mendonça *et al.*, 2020).

RESULTADOS

A Tabela 5 apresenta os modelos de previsão de vazão, onde Δt indica o horizonte de previsão.





Tabela 5 – Modelos de previsão de vazões ajustados a partir dos grupos *G1* a *G4* e *Todos*.

Δt	Grupo	AR (1,0)	AR (2,0)
	Todos	$Q_t = 0.9914(Q_{t-1} - \bar{Q}) + \bar{Q}$	$Q_t = 1,6714(Q_{t-1} - \bar{Q}) - 0,6860(Q_{t-2} - \bar{Q}) + \bar{Q}$
24h	G1	$Q_t = 0.9780(Q_{t-1} - \bar{Q}) + \bar{Q}$	$Q_t = 1,5978(Q_{t-1} - \bar{Q}) - 0,6338(Q_{t-2} - \bar{Q}) + \bar{Q}$
	G2	$Q_t = 0.9881(Q_{t-1} - \bar{Q}) + \bar{Q}$	$Q_t = 1,6268(Q_{t-1} - \bar{Q}) - 0,6469(Q_{t-2} - \bar{Q}) + \bar{Q}$
	G3	$Q_t = 0.9877(Q_{t-1} - \bar{Q}) + \bar{Q}$	$Q_t = 1,6949(Q_{t-1} - \bar{Q}) - 0,7167(Q_{t-2} - \bar{Q}) + \bar{Q}$
	G4	$Q_t = 0.9916(Q_{t-1} - \bar{Q}) + \bar{Q}$	$Q_t = 1,5619(Q_{t-1} - \bar{Q}) - 0,5783(Q_{t-2} - \bar{Q}) + \bar{Q}$
	Todos	$Q_t = 0.9710(Q_{t-2} - \bar{Q}) + \bar{Q}$	$Q_t = 2,0670(Q_{t-2} - \bar{Q}) - 1,1056(Q_{t-3} - \bar{Q}) + \bar{Q}$
	G1	$Q_t = 0.9288(Q_{t-2} - \bar{Q}) + \bar{Q}$	$Q_t = 1,9058(Q_{t-2} - \bar{Q}) - 0,9991(Q_{t-3} - \bar{Q}) + \bar{Q}$
48h	G2	$Q_t = 0.9605(Q_{t-2} - \bar{Q}) + \bar{Q}$	$Q_t = 1,9712(Q_{t-2} - \bar{Q}) - 1,0236(Q_{t-3} - \bar{Q}) + \bar{Q}$
	G3	$Q_t = 0.9572(Q_{t-2} - \bar{Q}) + \bar{Q}$	$Q_t = 2,1465(Q_{t-2} - \bar{Q}) - 1,2054(Q_{t-3} - \bar{Q}) + \bar{Q}$
	G4	$Q_t = 0.9703(Q_{t-2} - \bar{Q}) + \bar{Q}$	$Q_t = 1,8874(Q_{t-2} - \bar{Q}) - 0,9300(Q_{t-3} - \bar{Q}) + \bar{Q}$
	Todos	$Q_t = 0.9435(Q_{t-3} - \bar{Q}) + \bar{Q}$	$Q_t = 2,2994(Q_{t-3} - \bar{Q}) - 1,3677(Q_{t-4} - \bar{Q}) + \bar{Q}$
	G1	$Q_t = 0.8647(Q_{t-3} - \bar{Q}) + \bar{Q}$	$Q_t = 2,1038(Q_{t-3} - \bar{Q}) - 1,2669(Q_{t-4} - \bar{Q}) + \bar{Q}$
72h	G2	$Q_t = 0.9240(Q_{t-3} - \bar{Q}) + \bar{Q}$	$Q_t = 2,1506(Q_{t-3} - \bar{Q}) - 1,2424(Q_{t-4} - \bar{Q}) + \bar{Q}$
	G3	$Q_t = 0.9145(Q_{t-3} - \bar{Q}) + \bar{Q}$	$Q_t = 2,4419(Q_{t-3} - \bar{Q}) - 1,5481(Q_{t-4} - \bar{Q}) + \bar{Q}$
	G4	$Q_t = 0.9411(Q_{t-3} - \bar{Q}) + \bar{Q}$	$Q_t = 2,0329(Q_{t-3} - \bar{Q}) - 1,1070(Q_{t-4} - \bar{Q}) + \bar{Q}$
	Todos	$Q_t = 0.9118(Q_{t-4} - \bar{Q}) + \bar{Q}$	$Q_t = 2,4699(Q_{t-4} - \bar{Q}) - 1,5718(Q_{t-5} - \bar{Q}) + \bar{Q}$
	G1	$Q_t = 0.7904(Q_{t-4} - \bar{Q}) + \bar{Q}$	$Q_t = 2,2573(Q_{t-4} - \bar{Q}) - 1,4999(Q_{t-5} - \bar{Q}) + \bar{Q}$
96h	G2	$Q_t = 0.8824(Q_{t-4} - \bar{Q}) + \bar{Q}$	$Q_t = 2,2676(Q_{t-4} - \bar{Q}) - 1,4027(Q_{t-5} - \bar{Q}) + \bar{Q}$
	G3	$Q_t = 0.8634(Q_{t-4} - \bar{Q}) + \bar{Q}$	$Q_t = 2,6427(Q_{t-4} - \bar{Q}) - 1,8031(Q_{t-5} - \bar{Q}) + \bar{Q}$
	G4	$Q_t = 0.9078(Q_{t-4} - \bar{Q}) + \bar{Q}$	$Q_t = 2,1335(Q_{t-4} - \bar{Q}) - 1,2425(Q_{t-5} - \bar{Q}) + \bar{Q}$

Resultados da etapa de ajuste dos modelos de previsão

Na etapa de ajuste, o modelo AR(2) apresentou resultados melhores do que o modelo AR(1) para todos os horizontes de previsão. Como era de se esperar, os melhores ajustes foram para o horizonte de previsão de 24 h e os piores ajuste foram para o horizonte de previsão de 96 h.

Para o horizonte de previsão de 24 h, o coeficiente NS variou entre 0,96 e 0,98 para o modelo AR(1) e entre 0,98 e 0,99 para o modelo AR(2). Já o coeficiente KG variou entre 0,97 e 0,99 para o modelo AR(1) e entre 0,98 e 0,99 para o modelo AR(2). Em relação ao MAPE, este variou entre 5,6% e 7.9% para o modelo AR(1) e 3,8% e 5,4% para o modelo AR(2).

Para o horizonte de previsão de 96 h, o coeficiente NS variou entre 0,65 e 0,84 para o modelo AR(1) e entre 0,77 e 0,89 para o modelo AR(2). Já o coeficiente KG variou entre 0,71 e 0,88 para o modelo AR(1) e entre 0,79 e 0,91 para o modelo AR(2). Em relação ao MAPE, este variou entre 21,4% e 31,7% para o modelo AR(1) e 18,0% e 25,8% para o modelo AR(2).

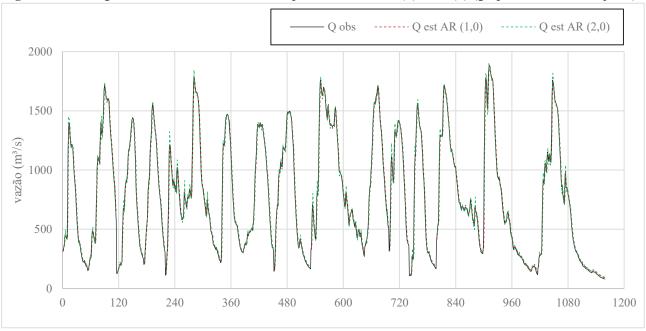
De modo geral, os piores índices de qualidade dos ajustes foram obtidos para os eventos do grupo *G1* e os melhores foram obtidos para os eventos do grupo *Todos*. Essa distinção fica maior, quanto maior for o horizonte de previsão. Contudo, a comparação entre os índices de qualidade dos ajustes dos diferentes grupos deve ser feita com reservas, uma vez que os índices de qualidade são calculados, considerando-se diferentes grupos de eventos.

Apesar do melhor desempenho, os modelos AR(2) apresentaram maior dificuldade para simular os picos dos hidrogramas, como pode ser observado na Figura 2, que apresenta, como exemplo, os eventos observados e estimados do grupo *G2* para o horizonte de previsão de 24h na fase de ajuste.





Figura 2 – Hidrogramas observados e estimados pelos modelos AR(1) e AR(2) (grupo G2, $\Delta t = 24$ h, ajuste)



Resultados da etapa de verificação dos modelos de previsão

Na etapa de verificação, os modelos de previsão produzidos para cada horizonte de previsão e para cada ordem de autorregressão, a partir dos grupos G1 a G4 e Todos, foram avaliados para conjuntos de eventos agrupados nos grupos G1 a G4, diferentes daqueles utilizados na etapa de ajuste. Nessa etapa, os modelos podem ser comparados diretamente entre si, porque os índices de qualidade são calculados para um mesmo conjunto de eventos.

As Tabelas 6 e 7 apresentam os valores do coeficiente de Nash-Sutcliffe (NS) para os modelos com horizonte de previsão de 24 h e 96 h, respectivamente, obtidos na etapa de verificação. As células destacadas em cinza representam os melhores desempenhos com base dos valores de NS. Observase, a partir dos valores de NS apresentados nas Tabelas 6 e 7, que os modelos AR(2) apresentaram um melhor desempenho quando comparados aos modelos AR(1). Além disso, o ganho no desempenho dos modelos de previsão, quando se compara o desempenho dos modelos obtidos a partir dos grupos *G1* a *G4* com o modelo obtido a partir do grupo *Todos*, que não há um ganho significativo de desempenho ao se ajustar o modelo para um grupo específico de eventos. Para o horizonte de previsão de 24 h, a explicação para essa conclusão é relativamente simples, porque o modelo ajustado para o grupo *Todos* apresenta valores de NS altos. No entanto, para o horizonte de 96 h, embora com uma melhoria de desempenho não muito significativa, os resultados sugerem um melhor desempenho quando o modelo é ajustado para um grupo específico de eventos.

Por outro lado, os resultados para o horizonte de previsão de 96 h evidenciam o fato de que modelos ajustados para um grupo com características distintas do grupo utilizado para a verificação tendem a produzir os piores desempenhos. Em termos práticos, é o equivalente a máxima de que





eventos "pequenos" tendem a subestimar eventos "grandes" e que eventos "grandes" tendem a superestimar eventos "pequenos", principalmente quando são utilizados modelos lineares. Em outras palavras, evidenciam a importância de uma amostra significativa de eventos para a calibração/ajuste e verificação de modelos hidrológicos.

As Tabelas 8 e 9 apresentam os valores do coeficiente de Kling-Gupta (KG) para os modelos com horizonte de previsão de 24 h e 96 h, respectivamente, obtidos na etapa de verificação. As células destacadas em cinza representam os melhores desempenhos com base dos valores de KG.

Tabela 6 – Valores dos coeficientes de Nash-Sutcliffe (NS) para o horizonte de previsão de 24 h, obtidos na etapa de verificação.

Tipo de	Grupo de	Modelo de Previsão					
Modelo	verificação	<i>G1</i>	G2	<i>G</i> 3	<i>G4</i>	Todos	
	G 1	0,959	0,959	0,959	0,959	0,959	
A D (1)	G2	0,971	0,971	0,971	0,971	0,971	
AR(1)	<i>G3</i>	0,978	0,979	0,979	0,979	0,979	
	G 4	0,977	0,977	0,977	0,977	0,977	
AR(2)	G1	0,978	0,978	0,977	0,975	0,978	
	G2	0,985	0,986	0,986	0,984	0,986	
	<i>G</i> 3	0,989	0,990	0,990	0,990	0,990	
	<i>G4</i>	0,989	0,991	0,991	0,990	0,991	

Tabela 7 – Valores dos coeficientes de Nash-Sutcliffe (NS) para o horizonte de previsão de 96 h, obtidos na etapa de verificação.

Tipo de	Grupo de	Modelo de Previsão					
Modelo	verificação	<i>G1</i>	G 2	<i>G</i> 3	G 4	Todos	
AR(1)	G1	0,611	0,614	0,602	0,542	0,606	
	G2	0,703	0,724	0,723	0,701	0,721	
	<i>G3</i>	0,724	0,770	0,775	0,767	0,773	
	G 4	0,710	0,742	0,746	0,745	0,743	
AR(2)	G 1	0,743	0,743	0,705	0,616	0,734	
	G2	0,756	0,785	0,775	0,750	0,781	
	<i>G3</i>	0,801	0,862	0,874	0,849	0,870	
	<i>G4</i>	0,772	0,815	0,824	0,815	0,819	

Tabela 8 – Valores dos coeficientes de Kling-Gupta (KG) para o horizonte de previsão de 24 h, obtidos na etapa de verificação.

Tipo de	Grupo de	Modelo de Previsão					
Modelo	verificação	<i>G1</i>	G 2	<i>G</i> 3	G 4	Todos	
AR(1)	<i>G1</i>	0,973	0,979	0,979	0,978	0,979	
	G2	0,974	0,983	0,983	0,983	0,984	
	<i>G</i> 3	0,974	0,985	0,985	0,987	0,987	
	G 4	0,972	0,983	0,983	0,986	0,985	
AR(2)	G1	0,981	0,988	0,983	0,972	0,985	
	G2	0,974	0,991	0,991	0,982	0,992	
	<i>G</i> 3	0,968	0,988	0,989	0,988	0,993	
	<i>G4</i>	0,967	0,987	0,990	0,991	0,993	





Tabela 9 – Valores dos coeficientes de Kling-Gupta (KG) para o horizonte de previsão de 96 h, obtidos na etapa de verificação.

Tipo de	Grupo de		M	odelo de Previs	são	
Modelo	verificação	G 1	G 2	<i>G</i> 3	<i>G4</i>	Todos
	<i>G1</i>	0,713	0,772	0,755	0,748	0,782
AD(1)	G2	0,732	0,812	0,799	0,806	0,829
AR(1)	<i>G</i> 3	0,735	0,829	0,820	0,843	0,851
	<i>G</i> 4	0,718	0,813	0,805	0,835	0,835
	G1	0,792	0,852	0,828	0,778	0,862
AD(2)	G2	0,764	0,856	0,850	0,826	0,877
AR(2)	<i>G</i> 3	0,753	0,870	0,874	0,873	0,905
	<i>G</i> 4	0,739	0,854	0,861	0,874	0,885

Observa-se, a partir dos valores de KG apresentados nas Tabelas 8 e 9, que os modelos AR(2) apresentaram um melhor desempenho quando comparados aos modelos AR(1). Além disso, verifica-se que o modelo ajustado a partir do grupo *Todos* apresentou melhor desempenho em relação aos demais, portanto não há ganho de desempenho ao se ajustar o modelo para um grupo específico de eventos.

Os resultados obtidos para KG, em parte, são explicados pela própria definição do coeficiente de Kling-Gupta (KG) que utiliza a razão entre as médias observadas e estimadas e os desvios-padrão observados e estimados, sugerindo que um modelo obtido de um conjunto mais "diverso" de eventos deve estimar de melhor forma os eventos observados. Importante destacar que o coeficiente de correlação, também utilizado no cálculo do coeficiente KG, não apresentou diferenças significativas entre os valores observados e estimados pelos distintos modelos.

As Tabelas 10 e 11 apresentam os valores do MAPE para os modelos com horizonte de previsão de 24 h e 96 h, respectivamente, obtidos na etapa de verificação. As células destacadas em cinza representam os melhores desempenhos com base dos valores de MAPE.

Novamente verifica-se, a partir dos valores do MAPE apresentados nas Tabelas 10 e 11, que os modelos AR(2) apresentaram um melhor desempenho quando comparados aos modelos AR(1). Além disso, verifica-se que os modelos ajustados a partir dos grupos *G1* e *Todos* apresentaram melhor desempenho em relação ao demais.

Os resultados obtidos para o MAPE podem ser explicados principalmente pela definição do próprio índice e pelo método utilizado para a estimativa dos parâmetros do modelo. O método dos mínimos quadrados, usado na estimativa dos parâmetros, trabalha com desvios (erros) absolutos entre os valores observados e estimados elevadas ao quadrado, enquanto o MAPE, em sua definição, utiliza desvios relativos em módulo. Desta forma, grupos com eventos menores (grupo GI) e grupos com mais eventos (grupo Todos) tenderiam a produzir ajustes em que a minimização das diferenças absolutas fossem menos "distantes" de ajustes produzidos pela minimização dos desvios relativos.





Tabela 10 – Valores do MAPE (%) para o horizonte de previsão de 24 h, obtidos na etapa de verificação.

Tipo de	Grupo de		M	odelo de Previs	são	
Modelo	verificação	<i>G1</i>	G 2	<i>G3</i>	<i>G4</i>	Todos
	<i>G1</i>	8,0	8,1	8,5	8,8	8,1
AD(1)	<i>G2</i>	7,7	7,8	8,2	8,5	7,7
AR(1)	<i>G3</i>	5,9	6,0	6,2	6,4	5,9
	<i>G4</i>	7,6	7,6	7,9	8,1	7,6
	<i>G1</i>	5,6	5,8	6,5	7,8	5,6
AD(2)	<i>G2</i>	5,8	5,8	6,5	7,6	5,6
AR(2)	<i>G</i> 3	4,0	3,9	4,3	5,1	3,8
	<i>G4</i>	5,7	5,6	5,9	6,7	5,5

Tabela 11 – Valores do MAPE (%) para o horizonte de previsão de 96 h, obtidos na etapa de verificação.

Tipo de	Grupo de		M	odelo de Previs	são	
Modelo	verificação	G1	G 2	<i>G</i> 3	<i>G4</i>	Todos
	G 1	26,7	28,3	33,1	38,0	28,1
AR(1)	G2	28,6	29,5	35,1	39,7	29,0
AK(1)	G 3	21,6	21,8	25,0	28,3	21,6
	G 4	26,8	28,2	31,5	34,5	28,2
	<i>G1</i>	21,1	23,5	29,4	36,8	22,9
AD(2)	G 2	24,4	25,1	31,5	38,3	24,1
AR(2)	<i>G</i> 3	17,6	17,0	20,0	25,6	16,3
	G 4	21,5	21,4	24,6	30,3	20,8

CONCLUSÕES

O objetivo do presente trabalho foi avaliar se o ajuste de um modelo de previsão de vazões com base no agrupamento de eventos pode produzir melhores estimativas de vazão para eventos classificados nos mesmos grupos utilizados para os ajustes dos modelos.

Os resultados obtidos mostraram que a conclusão sobre a eficiência do ajuste do modelo de previsão de vazões com base no agrupamento de eventos de cheias é função do próprio modelo utilizado, no caso, utilizou-se modelos autorregressivos de *ordem 1* e 2, dos índices usados para avaliar a adequação do ajuste e do método de estimativa de parâmetros do modelo. Portanto, para estudos futuros, recomenda-se aplicar a conceito que embasa o presente trabalho a outros modelos hidrológicos, por exemplo, modelos de transformação chuva-vazão, bem como, utilizar outros índices de qualidade de ajuste e métodos de estimativa de parâmetros do modelo de previsão.

REFERÊNCIAS BIBLIOGRÁFICAS

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C.; JUNG, G. M. L. (2016). "Time series analysis: forecasting and control". 5a. ed. [S.l.]: Wiley.

CASTILHO, A. S.; PINTO, É. J. A.; OLIVEIRA, G. S. R. (2005). "Utilização de modelo hidráulico associado a um SIG para a definição de áreas inundáveis na cidade de Governador Valadares, Minas Gerais" in Anais do XVI Simpósio Brasileiro de Recursos Hídricos, João Pessoa, nov. 2005.





CHU, H.; WU, W.; WANG, Q. J.; NATHAN, R.; WEI, J. (2020). "An ANN-based emulation modelling framework for flood inundation modelling: application, challenges and future directions". Environmental Modelling and Software, v. 124. DOI: 10.1016/j.envsoft.2019.104587.

COBRAPE – COMPANHIA BRASILEIRA DE PROJETOS E EMPREENDIMENTOS. (20--). *Produto 1.2 – Parte D – Levantamento e análise de eventos críticos*. Curitiba: 20--. Revisão final.

FREITAS, C. M.; XIMENES, E. F. (2012). "Enchentes e saúde pública: uma questão na literatura científica recente das causas, consequências e respostas para prevenção e mitigação". Ciência & Saúde Coletiva, v. 17, n. 6, pp. 1601–1616. DOI: 10.1590/s1413-81232012000600023.

GUPTA, H. V.; KLING, H.; YILMAZ, K. K.; MARTINEZ, G. F. (2009). "Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling". Journal of Hydrology, v. 377, pp. 80-91.

LAWIN, A. E.; HOUNGUÈ, R.; M'PO, Y. N'T.; HOUNGUÈ, N. R.; ATTOGOUINON, A.; AFOUDA, A. A. (2019). "Mid-Century Climate Change Impacts on Ouémé River Discharge at Bonou Outlet (Benin)". Hydrology, v. 6, n. 72, pp. 1-20.

MENDONÇA, L. M.; GOMIDE, I. S.; SOUSA, J. V.; BLANCO, C. J. C. (2021). "Modelagem chuvavazão via redes neurais artificiais para simulação de vazões de uma bacia hidrográfica da Amazônia". Revista de Gestão de Águas da América Latina, v. 18, e12, pp. 1–14. DOI: 10.21168/rega.v18e2.

NEDEL, A. S.; SAUNSEN, T. (2012). "Análise das inundações bruscas ocorridas na região sul do Brasil no período 1980 – 2009". Ciência e Natura, v. 34, n. 1, pp. 119–129.

ROCHA, P. S. M. (2012). Gestão em áreas de risco de enchentes: estudo de caso para União da Vitória – Paraná. 134 f. Dissertação (Mestrado em Gestão Ambiental), Universidade Positivo, Curitiba.

SEMA – SECRETARIA DE ESTADO DO MEIO AMBIENTE E RECURSOS HÍDRICOS. (2010). *Bacias hidrográficas do Paraná*. PEREIRA, M. C. B., SCROCCARO, J. L. (Orgs.), Curitiba.

STEFFEN, P. C.; GOMES, J. (2018). "Clustering of historical floods observed on Iguaçu River, in União da Vitória, Paraná". Revista Brasileira de Recursos Hídricos, v. 23, pp. 1–12. DOI: 10.1590/2318-0331.231820170107.

VERWEY, A.; KERBLAT, Y.; CHIA, B. (2017). Flood risk management at river basin scale: the need to adopt a proactive approach. Flórida: Urban Flood Community of Practice. DOI: 10.1596/27472.

ANEXO 2 – CARACTERÍSTICAS DE CHEIAS OBSERVADAS NA BACIA DO RIO IGUAÇU – PARTE 1: AGRUPAMENTO DE EVENTOS

Trabalho completo publicado nos Anais do XXV Simpósio Brasileiro de Recursos Hídricos.





XXV SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS

CARACTERÍSTICAS DE CHEIAS OBSERVADAS NA BACIA DO RIO IGUAÇU – PARTE 1: AGRUPAMENTO DE EVENTOS

Patrícia Cristina Steffen ¹& Júlio Gomes ²

Resumo: As informações extraídas de dados de vazões podem subsidiar diversos estudos na Engenharia de Recursos Hídricos. Uma aplicação é o agrupamento dos eventos de cheia, que pode ser desenvolvido a partir das próprias vazões e de outros parâmetros extraídos das vazões. O objetivo do agrupamento é classificar os eventos de cheia em grupos de eventos semelhantes entre si. Neste contexto, o presente trabalho realizou um agrupamento, denominado de cenário base, partindo dos dados de vazão de União da Vitória, e agrupamentos que combinaram as vazões máximas desta estação e outras cinco estações, localizadas a montante. Utilizou-se o algoritmo *fuzzy c-means* para realização do agrupamento e, em cada cenário analisado, as cheias foram classificadas em quatro grupos distintos entre si (grupos 1 a 4); o método classificou as cheias por meio da menor distância Euclidiana da observação até os centroides dos grupos. Fluviópolis e São Mateus do Sul apresentaram agrupamentos mais semelhantes ao de União da Vitória. Os eventos se movimentaram, sempre, a um grupo acima, ou abaixo, do grupo do cenário base, e as maiores alterações foram observadas entre os grupos 2 e 3; cinco eventos tiveram alteração em todas as simulações realizadas.

Palavras-Chave – Agrupamento de cheias. Caracterização de eventos. Fuzzy c-means.

INTRODUÇÃO

A complexidade da propagação das cheias nos rios torna necessária a avaliação da bacia hidrográfica a montante do ponto de interesse; aspectos hidrológicos das estações fluviométricas são diretamente afetados pelas estações de montante, e nenhuma estação é independente (Joo *et al.*, 2021). Fatores como o tempo de concentração e as características físicas da bacia, as condições do canal e as contribuições dos afluentes são relevantes para se investigar a influência das vazões observadas nas estações fluviométricas a montante sobre o local de estudo.

Dados observados de vazão fornecem informações importantes para diversas aplicações da Hidrologia e dos Recursos Hídricos (Joo *et al.*, 2021), e um dos principais focos de discussão da comunidade científica atualmente são os estudos que consigam descrever com precisão o comportamento histórico dos recursos hídricos (Graciano e Detzel, 2019). Dentre estas discussões estão o entendimento e a representação dos mecanismos de formação das cheias em uma bacia hidrográfica, que podem colaborar com a elaboração de sistemas de previsão. Wijayarathne e Coulibaly (2020) afirmam que medidas de mitigação e um sistema de previsão aprimorado são partes

¹⁾ Mestra, Universidade Federal de Mato Grosso, patricia.steffen@ufmt.br

²⁾ Doutor, Universidade Federal do Paraná, jgomes.dhs@ufpr.br





críticas de um sistema de gerenciamento de cheias, uma vez que um sistema de previsão preciso pode gerar previsões confiáveis em intervalos de tempo adequados.

Neste contexto, a análise de agrupamento tem sido utilizada para classificação de similaridades hidrológicas e da bacia hidrográfica, bem como, para previsão de dados hidrológicos, análise de frequência de cheias e, também, previsão de cheias (Joo *et al.*, 2021). Ademais, métodos de agrupamento têm sido utilizados na Hidrologia para identificação de diferenças de propriedades hidrológicas entre grupos (Wang *et al.*, 2015). Steffen e Gomes (2018) indicam que a análise de agrupamentos pode colaborar com o entendimento da formação das cheias.

O agrupamento, por sua vez, consiste na classificação dos eventos de cheia em grupos, cujos integrantes apresentam características semelhantes entre si e características distintas dos eventos pertencentes aos demais grupos formados; o agrupamento é precedido de uma seleção de variáveis, chamadas de indicadores, capazes de representar cada evento da série de dados. Os indicadores de cheia são definidos a critério do pesquisador e, neste trabalho, utilizou somente indicadores associados aos dados observados de vazão.

O objetivo do presente trabalho é agrupar eventos de cheia observados no rio Iguaçu em União da Vitória, como cenário base, e comparar o agrupamento obtido a outros agrupamentos realizados a partir de vazões máximas anuais observadas em estações fluviométricas localizadas a montante, com intuito de verificar se a incorporação de informações das estações a montante modifica o agrupamento obtido no cenário base. Os agrupamentos dão subsídios a posterior avaliação de eventos de cheia (Steffen e Gomes, 2023b), visto que as cheias agrupadas em um mesmo grupo tendem a ser similares entre si e menos semelhantes às demais. Com isso, acredita-se ser possível compreender, em parte, os mecanismos de formação de cheias na área de estudo escolhida, a bacia do rio Iguaçu a montante de União da Vitória, no Paraná.

ÁREA DE ESTUDO

A área de estudo compreende a bacia hidrográfica do rio Iguaçu a montante da Estação Fluviométrica União da Vitória (6531000) e foi escolhida em razão dos recorrentes eventos de cheia registrados historicamente no município de União da Vitória, no Paraná.

O município de União da Vitória se desenvolveu às margens do rio Iguaçu, que propiciou o desenvolvimento socioeconômico da região por meio do transporte de passageiros e mercadorias, da pesca, do lazer, da dessedentação e da geração de energia (Rocha, 2012), e sua população é majoritariamente urbana (IBGE, 2020). Desta forma, com a ocorrência das grandes cheias do rio Iguaçu, os impactos gerados sobre a cidade são significativos para a economia local, interrompendo as atividades que a suportam e interditando vias de acesso.





O rio Iguaçu cruza praticamente todo o estado do Paraná, no sentido Leste-Oeste, com uma extensão de 1.320 km, desaguando no rio Paraná. Localiza-se entre os paralelos 25° 00' e 27° 00' de latitude sul e os meridianos 49° 00' e 54° 30' de longitude oeste, e apresenta um relevo bastante irregular, com baixas declividades e vales amplos, em seu trecho superior, e altas declividades e vales rochosos estreitos, nos trechos médio e inferior (Mine e Tucci, 2002).

A bacia hidrográfica do rio Iguaçu possui uma área de drenagem total de 70.800 km², dividida entre os estados do Paraná e Santa Catarina, no Brasil, e a Argentina, com uma pequena parcela do território da bacia (Rocha, 2012). Ainda, há uma subdivisão desta área estabelecida pela Resolução Nº 49/2006/CERH/PR, definindo as três Unidades Hidrográficas de Gestão de Recursos Hídricos do Iguaçu: Baixo Iguaçu; Médio Iguaçu; e Alto Iguaçu, agrupado à Bacia do Ribeira. A estação fluviométrica União da Vitória (65310000) se situa no trecho Médio, com área de drenagem de, aproximadamente, 24.200 km² (SEMA, 2010). Na referida estação, a vazão média é de aproximadamente 500 m³/s (20 L/s/km²), tornando-se de quatro a seis vezes maior nos períodos de cheia; na maior cheia registrada na região, em 1983, a vazão de pico atingiu 5.157 m³/s (210 L/s/km²), 10 vezes maior que a vazão média da estação (COPEL, 1995 *apud* Rocha, 2012). A Figura 1 apresenta a localização da cidade de União da Vitória na bacia hidrográfica do rio Iguaçu.

-25.00 100 200 km -27.00 LEGENDA INFORMAÇÕES Fonte dos shapefiles - Território brasileiro: IBGE, 2015; União da Vitória - PR Estado do Paraná Datum: SIRGAS 2000 Bacia Hidrográfica: ANA, 2017; Estado de Santa Catarina Estações fluviométricas Local: Bacia do Rio Iguaçu no - Hidrografia: ANA, 2015 Bacias - Alto e Médio Iguaçu Brasil Fonte das coordenadas das estações Rio Iguacu - HidroWeb (ANA, 2022). Afluentes

Figura 1 – Bacia do rio Iguaçu no Paraná e em Santa Catarina e localização de União da Vitória, no Paraná.

MATERIAIS E MÉTODOS

Coleta e análise de dados fluviométricos

Foram investigadas as estações fluviométricas dos principais afluentes do rio Iguaçu, existentes na área de estudo. Dentre as estações encontradas, foram selecionadas aquelas apresentadas na Tabela 1, cujos dados foram coletados da plataforma HidroWeb da Agência Nacional das Águas e Saneamento Básico (ANA). O critério de seleção priorizou as estações nos afluentes com maiores





áreas de drenagem em cada região da bacia, bem como, a consistência dos dados e o tamanho da amostra. A localização das estações na área de estudo pode ser observada pelo ID da Tabela 1, que corresponde ao número de identificação das estações fluviométricas da Figura 1.

Dentre as Estações Fluviométricas da Tabela 1, observou-se que Santa Cruz do Timbó (65295000), São Bento (65155000) e Rio Negro (65100000) apresentaram períodos de falhas; a Estação São Bento (65155000) foi preenchida e estendida (até 2014) por meio da curva de permanência da Estação Fluviométrica Rio da Várzea dos Lima (65135000), a montante de São Bento, ambas no rio da Várzea. Os parâmetros estatísticos das vazões médias diárias foram calculados e são apresentados na Tabela 2.

Tabela 1. Estações Fluviométricas selecionadas, a montante de União da Vitória.

ID	Código	Nome da estação	Rio	Cidade	UF	Lat. Sul	Long. Oeste	A (km²)	Início-fim
1*	65010000	Fazendinha	Pequeno	São José dos Pinhais	PR	-25,5192	-49,1456	106	1956-2014
2*	65027000	Rodeio	Verde	Balsa Nova	PR	-25,5797	-49,5831	231	1973-2014
3	65035000	Porto Amazonas	Iguaçu	Porto Amazonas	PR	-25,5481	-49,8886	3620	1936-2005
4	65060000	São Mateus do Sul	Iguaçu	São Mateus do Sul	PR	-25,8758	-50,3897	6050	1931-2014
5*	65100000	Rio Negro	Negro	Rio Negro	PR	-26,1097	-49,8011	3450	1931-2014
6**	65155000	São Bento	da Várzea	Lapa	PR	-25,9456	-49,7914	2000	1931-2014
7	65180000	Salto Canoinhas	Canoinhas	Major Vieira	SC	-26,3722	-50,2881	772	1981-2014
8	65220000	Fluviópolis	Iguaçu	São Mateus do Sul	PR	-26,0192	-50,5925	18600	1964-2014
9*	65295000	Santa Cruz do Timbó	Timbó	Irineópolis	SC	-26,3839	-50,8783	2610	1975-2004
10	65310000	União da Vitória	Iguaçu	União da Vitória	PR	-26,2281	-51,0803	24200	1931-2014

Lat. – latitude; Long. – longitude; A – área de drenagem; * Estações com falhas; ** Estação preenchida.

Tabela 2. Estatística descritiva das vazões médias diárias das estações fluviométricas selecionadas.

ID	Nome da estação	tamanho da amostra (anos)	vazão média (m³/s)	vazão específica média (L/s/km²)	vazão mínima (m³/s)	vazão máxima (m³/s)	desvio- padrão (m³/s)	assimetria
1	Fazendinha	54	2,8	26,4	0,4	21,2	2,31	2,19
2	Rodeio	41	4,8	20,7	0,2	49,0	4,65	3,03
3	Porto Amazonas	Amazonas 70 68,3 18,9 3,4 116		1167,0	63,07	3,83		
4	São Mateus do Sul	84	107,0	17,7	11,7	1533,1	94,68	3,37
5	Rio Negro	72	71,9	20,9	7,7	953,9	68,05	3,30
6*	São Bento	84	35,4	17,7	6,0	564,0	28,31	3,35
7	Salto Canoinhas	34	18,9	24,5	0,8	254,9	22,79	2,83
8	Fluviópolis	51	395,8	21,3	34,5	3819,4	348,21	2,58
9	Santa Cruz do Timbó	26	78,7	30,2	4,5	706,0	82,61	2,22
10	União da Vitória	84	487,1	20,1	46,2	5156,7	439,09	2,56

^{*} Valores correspondentes à série preenchida





As correlações entre União da Vitória e as demais estações foram calculadas para diferentes lags de tempo para as séries de vazões médias diárias. As estações que apresentaram baixas correlações com União da Vitória foram descartadas para realização das etapas subsequentes.

Critérios de agrupamento e indicadores de cheia

Cada amostra é definida por um número *i* de indicadores de cheia, que definem a dimensionalidade do problema e são utilizados para o agrupamento; os dados da amostra (e dos indicadores) são identificados pelo ano de ocorrência do evento.

Foram utilizados dois indicadores de cheia para o agrupamento que representa o *cenário base*: vazão máxima anual em União da Vitória (*i1*) e taxa média de recessão em União da Vitória (*i2*), correspondentes às máximas cheias anuais observadas na estação. Para os demais agrupamentos, mencionados aqui como *agrupamentos adicionais*, também foram utilizados dois indicadores de cheia: vazão máxima anual em União da Vitória (*i1*) e vazão máxima na estação fluviométrica de montante para a cheia correspondente à máxima cheia anual observada em União da Vitória (*i2*). Portanto, para os agrupamentos adicionais, cada série de vazões máximas das estações com alta correlação foi agrupada à série de máximas anuais de União da Vitória, considerando os seus respectivos períodos simultâneos de dados. Os agrupamentos adicionais foram comparados ao cenário base para a análise dos resultados.

Devido às diferenças de magnitude das vazões e de unidade de medida (no caso da taxa média de recessão), os indicadores de cheia foram normalizados pelo método linear (Steffen e Gomes, 2018), considerando os períodos simultâneos entre as estações a cada simulação. Portanto, as amostras agrupadas possuíam valores no intervalo [0, 1].

Agrupamento de eventos anuais de cheia

Para o agrupamento de dados, utilizou-se o método *fuzzy c-means (fcm)*, disponível na biblioteca Python *skfuzzy*; o equacionamento do método é apresentado em Steffen e Gomes (2018). O método exige a definição prévia do número de grupos (maior que 2) para sua aplicação, por isso, foram adotados 4 grupos de eventos em todas as simulações, como resultado do trabalho de Steffen e Gomes (2018).

RESULTADOS

Correlações das vazões

Na Tabela 3 são apresentadas as correlações das vazões de União da Vitória com as vazões observadas nas estações fluviométricas selecionadas para diferentes *lags* de tempo, considerando as séries de vazões médias diárias. Para cada estação fluviométrica, foram destacadas as maiores correlações encontradas.





Tabela 3. Correlações entre as séries de vazões diárias de União da Vitória (65310000) e as das estações selecionadas, para diferentes *lags* de tempo.

lan		Estação Fluviométrica (ID)														
lag	1	2	3	4	5	6	7	8	9							
0	0,5406	0,6015	0,7845	0,8988	0,8414	0,8239	0,8192	0,9781	0,9359							
1	0,5706	0,6331	0,8081	0,8968	0,8575	0,8329	0,8380	0,9697	0,9450							
2	0,5875	0,6521	0,8190	0,8871	0,8620	0,8318	0,8430	0,9525	0,9340							
3	0,5962	0,6623	0,8217	0,8718	0,8591	0,8244	0,8398	0,9289	0,9148							
4	0,5990	0,6663	0,8190	0,8522	0,8506	0,8127	0,8309	0,9007	0,8904							
5	0,5974	0,6658	0,8117	0,8294	0,8375	0,7977	0,8173	0,8691	0,8625							

Observa-se, a partir da Tabela 3, que Fluviópolis (**ID** 8) obteve maior correlação com União da Vitória (**ID** 10) dentre todas as estações para todos os *lags*, seguida de Santa Cruz do Timbó (**ID** 9), que obteve a segunda maior. Santa Cruz do Timbó, contudo, apesar de ser a estação mais próxima (Figura 1) de União da Vitória, é a que apresenta série de dados mais curta (1975 a 2004, com falhas). Já Fluviópolis apresenta a maior área de drenagem dentre as estações avaliadas e é a segunda estação mais próxima (Figura 1), além de também se localizar no rio Iguaçu, enquanto Santa Cruz do Timbó se localiza no rio Timbó, afluente do rio Iguaçu. Em geral, quanto maior a distância entre a estação e União da Vitória, maior o *lag* onde se observa a correlação mais alta, o que se justifica pelo tempo de viagem da vazão entre a estação e União da Vitória. Em São Mateus do Sul (**ID** 4) esta situação não se verificou, o que se justifica por se localizar no próprio rio Iguaçu.

As séries das Estações Fluviométricas Fazendinha (**ID** 1) e Rodeio (**ID** 2) apresentaram menores correlações para todos os *lags* de tempo (inferiores a 0,67) e não foram consideradas para o agrupamento. As referidas estações são as mais distantes de União da Vitória e representam as menores contribuições (pequenas áreas de drenagem).

Agrupamento dos eventos anuais de cheia

O número dos grupos (1 a 4) representa, em ordem crescente, a magnitude das vazões máximas das cheias anuais observadas em União da Vitória. A Tabela 4 apresenta o resultado do agrupamento das máximas cheias anuais observadas na estação fluviométrica União da Vitória (cenário base), representadas pelo seu ano de ocorrência.

Tabela 4. Agrupamento para a estação fluviométrica União da Vitória (cenário-base).

		1				2	2			4		
1933	1944	1959	1968	1986	1931	1955	1979	2004	1932	1982	1999	1983
1934	1945	1960	1969	1991	1937	1957	1980	2005	1935	1987	2000	1992
1936	1949	1962	1974	1994	1939	1961	1981	2007	1938	1989	2001	2014
1940	1951	1963	1976	2002	1947	1965	1984	2008	1946	1990	2010	
1941	1952	1964	1977	2003	1948	1970	1988	2009	1954	1993	2011	
1942	1956	1966	1978	2006	1950	1973	1996	2012	1971	1995	2013	
1943	1958	1967	1985		1953	1975	1997		1972	1998	2015	





De forma semelhante foram produzidos agrupamentos, a partir do conjunto formado pelos dados de União de Vitória e cada uma das demais estações separadamente, perfazendo um total de 7 agrupamentos adicionais, cujos grupos *não* são aqui apresentados. A Tabela 5 apresenta uma análise da diferença entre os agrupamentos adicionais produzidos e o cenário base, onde é mostrado o número de anos em que houve mudança da classificação da cheia anual em comparação ao cenário base.

Tabela 5. Variação na classificação entre os agrupamentos adicionais e o cenário base.

ID	Código	Nome da estação*	Número de diferenças na classificação	Número de anos da amostra	Variação na classificação (%)
3	65035000	Porto Amazonas	16	70	22,9%
4	65060000	São Mateus do Sul	13	84	15,5%
5	65100000	Rio Negro	16	72	22,2%
6	65155000	São Bento	17	84	20,2%
7	65180000	Salto Canoinhas	16	34	47,1%
8	65220000	Fluviópolis	8	51	15,7%
9	65295000	Santa Cruz do Timbó	7	26	26,9%

^{*} corresponde ao nome da estação fluviométrica cujos dados foram agregados aos dados de União da Vitória para o agrupamento das cheias anuais.

Observa-se da Tabela 5 que as estações Salto Canoinhas (**ID** 7) e Santa Cruz do Timbó (**ID** 9) apresentaram as maiores diferenças de agrupamento quando comparadas ao cenário base. Especificamente Santa Cruz do Timbó (**ID** 9) não possui informações anuais para as cheias de 1983 e 1992 que são as maiores cheias observadas em União da Vitória (**ID** 10) para o período do estudo. Como consequência, o grupo 4 (com as maiores vazões) ficou apenas com 1 evento. Em razão destes resultados, ambas as estações também não foram consideradas nas demais análises. As estações São Mateus do Sul (**ID** 4) e Fluviópolis (**ID** 8) apresentaram relativamente, nesta ordem, os agrupamentos mais similares ao cenário base.

A Tabela 6 apresenta, como exemplo, a variação nos agrupamentos de São Mateus do Sul (**ID** 4) e Fluviópolis (**ID** 8) em comparação ao cenário base.

Tabela 6. Variação entre os grupos gerados para as simulações de São Mateus do Sul e Fluviópolis.

				a Vitória o base)	
	Grupo	1	2	3	4
S	1	30			
União da Vitória + ão Mateu do Sul	2	4	24	6	
União d Vitória + São Mate do Sul	3		3	14	
_ ×	4				3

			União da Vitória (cenário base)							
	Grupo	1	2	3	4					
	1	15	1							
União da Vitória + Fluviópolis	2	1	14	4						
União da Vitória + luviópoli	3		2	11						
	4				3					

As diagonais principais das matrizes indicam a quantidade de eventos classificados nos agrupamentos adicionais nos mesmos grupos do cenário base. Os elementos fora da diagonal principal indicam a quantidade de eventos classificados nos agrupamentos adicionais em grupos





distintos do cenário base. De modo geral, as maiores diferenças ocorrem para os grupos 2 e 3, para todas as estações fluviométricas analisadas nesta fase do estudo. Esta situação se repetiu para os demais agrupamentos adicionais.

A Tabela 7 apresenta um panorama da migração dos eventos de máxima cheia anual dos grupos obtidos no cenário base para os grupos obtidos nos agrupamentos adicionais.

Tabela 7. Migração dos grupos do cenário base para grupos dos agrupamentos adicionais

Migração de Grupo		F	Estação Fluvi	ométrica (II))	
(do cenário base para o agrupamento adicional)	Ano	3	4	5	6	8
	1942	X				
	1945				X	
	1966	X	X		X	
Grupo 1 para Grupo 2	1976	X	X			X
	1986		X			
	1994		X			
	1931			X	X	
	1937			X	X	
	1950	X		X		
	1953	X			X	
Grupo 2 para Grupo 1	1965			X		
	1996	X				
	2004	X			X	
	2008				X	X
	1955		X			
	1957	X	X	X	X	
Grupo 2 para Grupo 3	1984			X		
	1997			X		X
	2009		X	X	X	X
	1932		X		X	
	1938	X	X		X	
	1972	X	X	X	X	X
	1982	X		X		
	1987			X		
Grupo 3 para Grupo 2	1989	X	X	X	X	X
	1990	X		X		
	1995			X		
	1999	X	X		X	X
	2000	X		X	X	
	2001	X	X	X	X	X
Grupo 4 para Grupo 3	2014				X	

[&]quot;X" – simulação com alteração de grupo; "---" – estação sem observação no referido ano.

Observa-se, a partir da Tabela 7, que as alterações de cada evento, de modo geral, se deram sempre à mesma maneira; por exemplo, o ano de 1966 pertencia ao grupo 1 (cenário base) e passou a pertencer ao grupo 2 em todas os agrupamentos adicionais em que houve alteração; isto demonstra a aplicabilidade do método para identificação das similaridades entre as vazões. Ainda, os anos em





destaque na Tabela 7 representam eventos com alteração para todas as simulações que possuem o referido dado observado.

Uma vez que a determinação do pertencimento do evento (máxima cheia anual) a um grupo é feita com base na menor distância Euclidiana até cada um dos centroides do agrupamento formado, faz-se também necessária a observação das coordenadas dos centroides, conforme a Tabela 8; cada par de coordenadas representa um grupo, em ordem (1 a 4). As distâncias dos centroides dos agrupamentos adicionais para os centroides do cenário base são apresentadas na Tabela 9.

Tabela 8. Centroides dos grupos definidos no cenário base e nos agrupamentos adicionais.

	Cenári	o base	esta	estações		esta	ções	esta	ções	est	ações	estações	
Grupo	esta 6531000	,	653100 650350	()		65310000(UV) 65060000(SM)		65310000(UV) 65100000(RN)		65310000(UV) 65155000(SB)		65310000(UV) 65220000(FL)	
	I_1	I_2	Quv	QPA		Quv	Qsм	Quv	QRN	Quv	QsB	Quv	QFL
1	0,1055	0,0031	0,1063	0,0763		0,1209	0,0852	0,1175	0,1498	0,1089	0,0816	0,1158	0,1041
2	0,2584	0,1726	0,2298	0,1823		0,2907	0,1909	0,2776	0,3050	0,2651	0,1790	0,2583	0,2609
3	0,3789	0,4570	0,3735	0,3393		0,4212	0,4822	0,4118	0,4354	0,4218	0,3448	0,3971	0,4325
4	0,7852	0,7965	0,7257	0,8239		0,7874	0,8544	0,7834	0,9434	0,8685	0,8524	0,7840	0,9076

Tabela 9. Distância dos centroides dos agrupamentos adicionais para os centroides do cenário base.

Grupo	estações 65310000(UV) 65035000(PA)	estações 65310000(UV) 65060000(SM)	estações 65310000(UV) 65100000(RN)	estações 65310000(UV) 65155000(SB)	estações 65310000(UV) 65220000(FL)
1	0,0732	0,0835	0,1472	0,0786	0,1015
2	0,0302	0,0371	0,1338	0,0093	0,0883
3	0,1178	0,0492	0,0394	0,1201	0,0305
4	0,0655	0,0579	0,1469	0,1003	0,1111
distância média	0,0391	0,0297	0,0626	0,0438	0,0443

Apesar de São Mateus do Sul e Fluviópolis apresentarem resultados mais similares em relação aos eventos agrupados, observando a distância média (Tabela 9), as menores distâncias foram para São Mateus do Sul e Porto Amazonas; esta, apesar de apresentar as menores distâncias, teve um dos maiores números de alterações no agrupamento (Tabela 5).

CONCLUSÕES

O objetivo do trabalho foi elaborar um agrupamento como cenário base e compará-lo à agrupamentos adicionais, relativos às estações fluviométricas a montante de União da Vitória. O agrupamento que representa o cenário base considerou as vazões máximas anuais em União da Vitória e suas respectivas taxas médias de recessão; os agrupamentos adicionais consideraram a mesma vazão máxima anual, combinada à máxima anual de cada estação a montante.

Considerando unicamente a quantidade de eventos que não alteraram seu grupo, São Mateus do Sul e Fluviópolis retornaram um agrupamento mais semelhante ao cenário base. A maior





quantidade de alterações na classificação ocorreu entre os grupos 2 e 3; os eventos de 1957, 1972, 1989, 2001 e 2009, que alteraram de grupo em todas as simulações, estão presentes nestes grupos.

A classificação de cada evento a um grupo, ou outro, ocorre por meio do cálculo da distância Euclidiana do dado até os centroides; o evento se classifica ao grupo mais próximo, mas, segundo os princípios do método *fuzzy c-means*, não deixa de compartilhar características com os demais grupos.

As amostras e a convergência do método fizeram com que os centroides calculados nas simulações e no cenário base fossem diferentes; os centroides de São Mateus do Sul e Porto Amazonas apresentaram menor distância média até os centroides do cenário base. Porto Amazonas, no entanto, apresentou um dos resultados mais divergentes do cenário base em termos de agrupamento. A posição dos novos centroides, se comparadas aos indicadores dos eventos agrupados, pode explicar as diferenças observadas, conforme discutido em Steffen e Gomes (2023b).

REFERÊNCIAS BIBLIOGRÁFICAS

GRACIANO, M.; DETZEL, D. H. M. (2019). "Análise de tendências nas séries de vazões na bacia do rio Iguaçu" in Anais do XXIII Simpósio Brasileiro de Recursos Hídricos, Foz do Iguaçu, Nov. 2019.

IBGE – Instituto Brasileiro de Geografia e Estatística (2020). Disponível em: http://www.ibge.gov.br/home>. Acesso em: 30 jun. 2020.

JOO, H.; KIM, H. S.; KIM, S.; SIVAKUMAR, B. (2021). "Complex networks and integrated centrality measure to assess the importance of streamflow stations in a River basin". Journal of Hydrology, v. 598, pp. 1–12. DOI: 10.1016/j.jhydrol.2021.126280.

MINE, M. R. M.; TUCCI, C. E. M. (2002). "Gerenciamento da produção de energia e controle de inundação: Foz do Areia no rio Iguaçu". Revista Brasileira de Recursos Hídricos, v. 7, n. 3, pp. 85-107. DOI: 10.21168/rbrh.v7n3.p85-107.

ROCHA, P. S. M. (2012). Gestão em áreas de risco de enchentes: estudo de caso para União da Vitória – Paraná. 134 f. Dissertação (Mestrado em Gestão Ambiental), Universidade Positivo, Curitiba.

SEMA – Secretaria de Estado do Meio Ambiente e Recursos Hídricos. (2010). *Bacias hidrográficas do Paraná*. PEREIRA, M. C. B., SCROCCARO, J. L. (Orgs.), Curitiba.

STEFFEN, P. C.; GOMES, J. (2018). "Clustering of historical floods observed on Iguaçu River, in União da Vitória, Paraná". Revista Brasileira de Recursos Hídricos, v. 23, pp. 1–12. DOI: 10.1590/2318-0331.231820170107.

STEFFEN, P. C.; GOMES, J. (2023b). "Características de cheias observadas na bacia do rio Iguaçu – parte 2: avaliação de eventos observados" in Anais do XXV Simpósio Brasileiro de Recursos Hídricos, Aracaju, Nov., 2023.

WANG, L. N.; CHEN, X. H.; SHAO, Q. X.; LI, Y. (2015). "Flood indicators and their clustering features in Wujiang River, South China". Ecological Engineering, v. 76, pp. 66–74. DOI: 10.1016/j.ecoleng.2014.03.018.

WIJAYARATHNE, D. B.; COULIBALY, P. (2020). "Identification of hydrological models for operational flood forecasting in St. John's, Newfoundland, Canada". Journal of Hydrology: Regional Studies, v. 27, pp. 1–16. DOI: 10.1016/j.ejrh.2019.100646.

ANEXO 3 – CARACTERÍSTICAS DE CHEIAS OBSERVADAS NA BACIA DO RIO IGUAÇU – PARTE 2: AVALIAÇÃO DE EVENTOS OBSERVADOS

Trabalho completo publicado nos Anais do XXV Simpósio Brasileiro de Recursos Hídricos.





XXV SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS

CARACTERÍSTICAS DE CHEIAS OBSERVADAS NA BACIA DO RIO IGUAÇU – PARTE 2: AVALIAÇÃO DE EVENTOS OBSERVADOS

Patrícia Cristina Steffen ¹& Júlio Gomes ²

Resumo: A indicação de características de cheias permite classificá-las em grupos de eventos semelhantes entre si, e distintos entre os grupos. Desta forma, os resultados podem ser utilizados para avaliar os impactos de uma cheia sobre determinada região, partindo da identificação de peculiaridades dos eventos dentro de um grupo obtido. Neste contexto, o objetivo deste trabalho foi interpretar a posição dos eventos de cheia em agrupamentos previamente estabelecidos. Para a área de estudo, bacia do rio Iguaçu em União da Vitória, estado do Paraná e considerando os agrupamentos definidos, a avaliação dos eventos de cheia foi realizada por meio da representação gráfica dos elementos dos grupos de cada classificação realizada; cada gráfico foi comparado a um cenário base para o período simultâneo de dados entre as estações em questão. Os principais eventos que sofreram alteração de grupo foram avaliados por meio de comparação das posições de seus indicadores no cenário base e nas novas classificações. Os resultados obtidos mostraram que os eventos de cheia localizados próximos às fronteiras de seus grupos no cenário base foram os mais sensíveis à mudança nas demais classificações, pois apresentaram maiores distâncias ao centroide do grupo a que pertencem, quando comparados aos demais integrantes do grupo.

Palavras-Chave – Caracterização de eventos. Agrupamento de cheias.

INTRODUÇÃO

A análise de agrupamento tem como finalidade a classificação de dados, de modo que cada dado "seja semelhante aos outros no agrupamento com base em um conjunto de características escolhidas" (Hair Junior et al., 2009). Para Bezdek et al. (1983), a utilização destas características no agrupamento permite reduzir a quantidade de dados em uma amostra, já que objetos semelhantes estão em um mesmo grupo. As similaridades são medidas por meio da matriz de proximidade, que indica o grau de familiaridade de um objeto com o grupo (Sato-Ilic e Jain, 2006).

Neste cenário também está inserida a lógica difusa existente no algoritmo *fuzzy c-means*, utilizado para realização do agrupamento. Nesta abordagem, a pertinência dos dados aos grupos permite valores no intervalo de 0 a 1, indicando que cada dado se relaciona com todos os grupos, em maior ou menor grau – quanto maior a relação, mais o valor se aproxima de 1 –, ao contrário do conceito clássico de conjuntos (Sato-Ilic e Jain, 2006).

Contudo, a classificação demanda um objetivo e, no caso dos eventos de cheia, a intenção é agrupá-las para, inicialmente, poder identificar suas particularidades dentro de um grupo e, desta forma, o agrupamento de cheias pode ser utilizado para avaliar os impactos de um evento sobre

-

¹⁾ Mestra, Universidade Federal de Mato Grosso, patricia.steffen@ufmt.br

²⁾ Doutor, Universidade Federal do Paraná, jgomes.dhs@ufpr.br





determinada região (Wang *et al.*, 2015). Por isso, visto que a pertinência dos dados a cada grupo é definida pela distância Euclidiana dos dados até os centroides das classes, o processo de caracterização dos eventos pode ter início com a avaliação da posição dos dados com relação aos centros. Ademais, uma vez que a escolha de indicadores para análise de agrupamento é critério do pesquisador, esta abordagem favorece uma escolha mais adequada das variáveis que definem uma região.

Neste contexto, o presente trabalho visa realizar uma avaliação preliminar de eventos de cheia agrupados previamente (Steffen e Gomes, 2023a), por meio da interpretação da posição dos eventos de cheia em relação a um cenário base e aos centroides dos grupos nas demais classificações realizadas. A partir desta análise, foram identificadas as estações que podem fornecer outros indicadores de cheia que melhorem a compreensão da dinâmica das cheias na região de estudo, bacia hidrográfica do rio Iguaçu em União da Vitória, estado do Paraná.

MÉTODOS

Para a área de estudo e os agrupamentos de cheia descritos em Steffen e Gomes (2023a), a avaliação dos eventos de cheia foi realizada a partir da representação gráfica dos grupos de cada classificação, comparadas, uma a uma, a um cenário base para o período simultâneo entre os pares; ou seja, foram elaborados gráficos em pares, em que cada agrupamento adicional foi comparado ao cenário base.

Conforme apresentam Steffen e Gomes (2023a), o cenário base foi o agrupamento definido pelo uso da vazão máxima anual em União da Vitória (i_l) e da taxa média de recessão em União da Vitória (i_2), como indicadores de cheia; os agrupamentos adicionais foram estabelecidos, tendo como indicadores de cheia a vazão máxima anual em União da Vitória (i_l) e a vazão máxima na estação fluviométrica de montante da cheia correspondente à máxima cheia anual observada em União da Vitória (i_2). As estações fluviométricas a montante de União da Vitória (65310000), consideradas para o presente estudo, foram: Porto Amazonas (65035000), São Mateus do Sul (65060000), Rio Negro (65100000), São Bento (65155000) e Fluviópolis (65220000).

Os principais anos que sofreram alteração de grupo foram avaliados por meio de comparação das posições de seus indicadores no cenário base e nas novas classificações (agrupamentos adicionais), bem como, pela comparação dos indicadores destes eventos com as coordenadas dos centroides em cada agrupamento adicional. Também é importante ressaltar que esta comparação entre os indicadores é possível devido à normalização dos dados, a partir da qual foram eliminadas as magnitudes e unidades de medida.

Ainda, a numeração dos grupos é possível na discussão devido à ordem crescente dos centroides adotada, considerando as vazões de União da Vitória (i_I) ; ou seja, para todos os agrupamentos





realizados, todos os grupos 1 correspondem a eventos de mesma magnitude em União da Vitória, assim como todos os grupos 2, e assim sucessivamente.

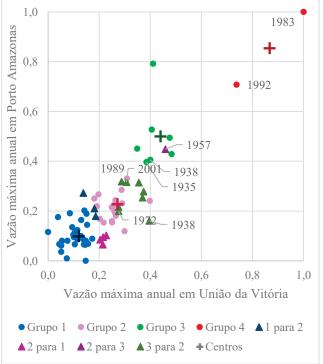
RESULTADOS

As Figuras 1 a 5 representam os eventos agrupados no cenário base (à direita nas figuras) e nos agrupamentos adicionais (à esquerda nas figuras) para efeito de comparação; os anos identificados indicam as principais alterações observadas quando efetuadas as comparações com o cenário base. Os marcadores em formato de cruz indicam os centroides dos grupos, enquanto os círculos coloridos, os eventos agrupados (cada cor corresponde a um grupo); os triângulos indicam os eventos que sofreram alteração de grupo, quando comparados ao cenário base. Ainda, é importante ressaltar que os triângulos das Figuras 1 a 5 são apresentados como "*x para y*" nas legendas, indicando que aquele evento de cheia migrou do grupo *x* do cenário base para o grupo *y* do agrupamento adicional.

As variações das posições dos centroides também podem ser observadas nas Figuras 1 a 5: os centroides de todos os agrupamentos adicionais sofreram alteração em suas coordenadas, quando comparados às coordenadas dos centroides do cenário base. Cada coordenada está associada a um indicador de cheia, o que explica as modificações nas características dos agrupamentos adicionais.

Nota-se também que os anos 1957, 1972, 1989, 2001 e 2009 apresentaram alterações em todos os agrupamentos adicionais com dados no período, quando comparados ao cenário base. Destaca-se que Porto Amazonas não possui o evento de 2009, enquanto Fluviópolis não possui o evento de 1957.

Figura 1. Comparação do agrupamento com vazões máximas de União da Vitória (65310000) e Porto Amazonas (65035000) (à esquerda) com o agrupamento do cenário base (à direita).



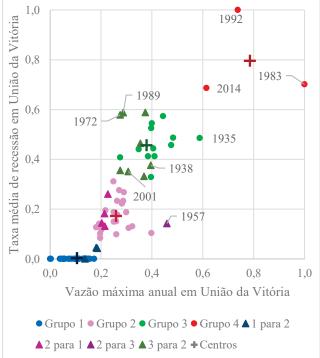






Figura 2. Comparação do agrupamento com vazões máximas de União da Vitória (65310000) e São Mateus do Sul (65060000) (à esquerda) com o agrupamento do cenário base (à direita).

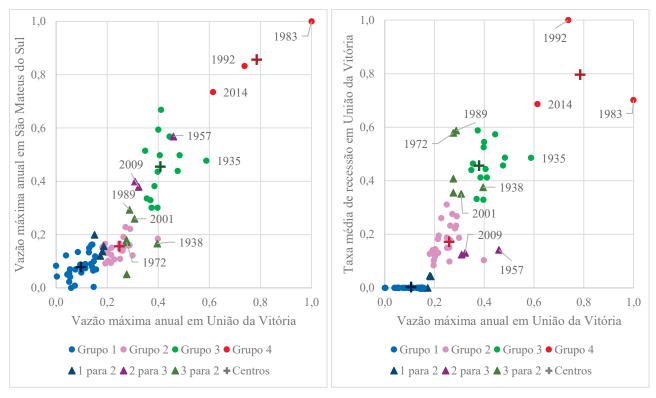


Figura 3. Comparação do agrupamento com vazões máximas de União da Vitória (65310000) e Rio Negro (65100000) (à esquerda) com o agrupamento do cenário base (à direita).

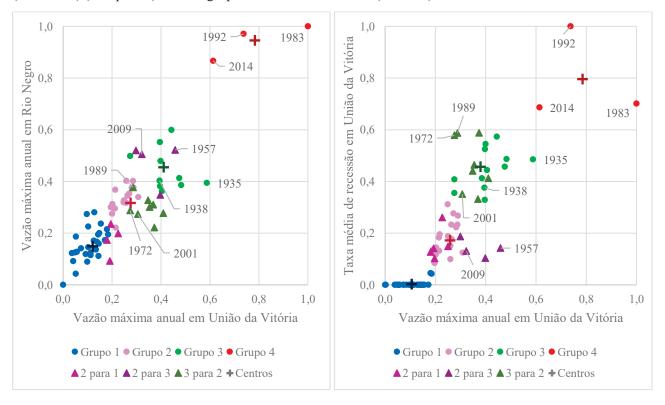






Figura 4. Comparação do agrupamento com vazões máximas de União da Vitória (65310000) e São Bento (65155000) (à esquerda) com o agrupamento do cenário base (à direita).

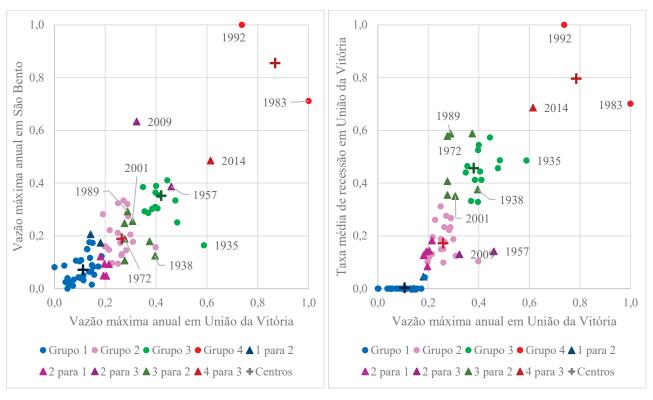
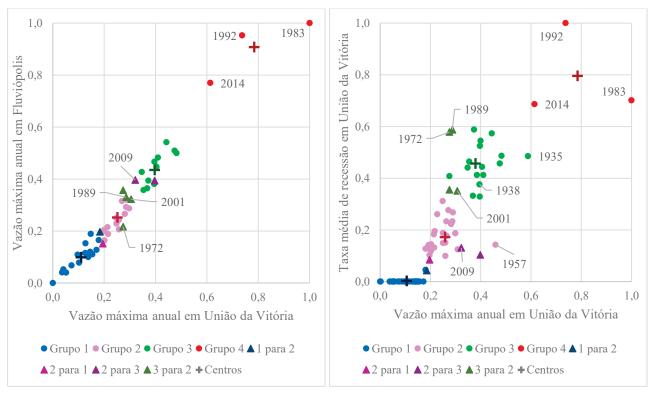


Figura 5. Comparação do agrupamento com vazões máximas de União da Vitória (65310000) e Fluviópolis (65220000) (à esquerda) com o agrupamento do cenário base (à direita).





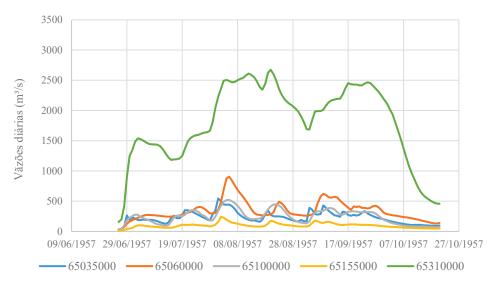


O significado hidrológico das alterações dos eventos dentre os grupos pode ser explicado pelos indicadores de cheia utilizados em cada cenário: enquanto a taxa média de recessão – indicador no cenário base – atribui peso ao volume dos hidrogramas e à duração dos eventos de cheia, as vazões máximas das estações de montante – indicadores nos demais cenários – representam a ocorrência de eventos críticos nos afluentes, o que remete ao contexto de previsão de cheias.

Verifica-se ainda, a partir das Figuras 1 a 5, que, de um modo geral, os deslocamentos dos eventos entre os grupos, quando comparados os agrupamentos adicionais e o agrupamento do cenário base, deram-se pela maior distância dos eventos ao centroide do seu grupo no agrupamento definido no cenário base (à direita nas figuras). No novo agrupamento (à esquerda nas figuras), estes eventos foram deslocados para uma região com maior proximidade de seu novo centro, ou seja, em geral, eventos de um grupo posicionados inicialmente em uma região de fronteira com outro grupo (ou extremidade do grupo), apresentaram mudança no agrupamento, posicionando-se mais próximos do centro no novo grupo.

Um exemplo, é o evento de 1957, que inicialmente estava no grupo 2 (2 para 3 na legenda) distante de seu respectivo centro; em todas as simulações, o mesmo evento passou a pertencer ao grupo 3 (2 para 3 na legenda), mais próximo do respectivo centroide. Com fins ilustrativos, a Figura 6 representa os hidrogramas dos eventos de cheia nas estações fluviométricas com dados observados para o referido evento de cheia.

Figura 6. Vazões diárias observadas nas estações fluviométricas analisadas para o evento de cheia de 1957.



Observa-se, a partir da Figura 6, que o evento de 1957 é composto por três ondas de cheia e que seu pico em União da Vitória (65310000) ocorreu durante a segunda onda, o que ocasionou um prolongamento do seu tempo de recessão. Desta forma, sua taxa média de recessão (um dos indicadores do cenário base) foi relativamente pequena, de forma que o evento se aproximasse mais do centroide do grupo 2, como pode ser observado no agrupamento do cenário base (Figuras 1 a 5).





Contudo, ao se agrupar considerando as vazões máximas das estações a montante de União da Vitória e a vazão máxima da estação União da Vitória, o evento 1957 se movimentou para o grupo 3, em razão dos valores normalizados de vazão máxima das estações a montante de União da Vitória resultarem superiores ao valor normalizado da taxa média de recessão em União da Vitória para a cheia de 1957, como pode ser observado na comparação entre os agrupamentos adicionais e o agrupamento do cenário base (Figuras 1 a 5).

Os anos de 1972 e 1989, também alocados na fronteira de seu grupo no cenário base, passaram do grupo 3 para o grupo 2, visto que estes eventos apresentavam valores normalizados de taxas de recessão (*i*₂) (0,5785 e 0,5875) maiores do que os valores normalizados das vazões máximas das estações a montante de União da Vitória (Q_{estação}), conforme a Tabela 1.

Tabela 1. Coordenadas dos indicadores de cheia para os eventos de 1972 e 1989.

'	19	72	19	Grupo		
Cenário Base	Quv	i_2	Quv	i_2	3	
Cenario dase	0,2757	0,5785	0,2874	0,5875	3	
Agrupamentos Adicionais	Quv	Qestação	Quv	Qestação		
Fluviópolis	0,2737	0,2167	0,2855	0,3302		
São Mateus do Sul	0,2757	0,1717	0,2875	0,2930	2	
São Bento	0,2757	0,1892	0,2875	0,2931	2	
Rio Negro	0,2737	0,2880	0,2855	0,3773		
Porto Amazonas	0,2757	0,1993	0,2875	0,3181		

O ano de 2009 (que também se destacou por ter alterado de grupo em todas as simulações com dados) estava relativamente próximo ao centroide do grupo 2 no cenário base, conforme Figuras 1 a 5. Ao se deslocar para o grupo 3, permaneceu a uma distância próxima do novo centroide, exceto para o agrupamento com as vazões máximas de União da Vitória e São Bento como indicadores de cheia (Figura 4 à esquerda). O ano de 2001, no cenário base estava posicionado na fronteira do grupo 3 (Figuras 1 a 5), e com seu deslocamento para o grupo 2 nos agrupamentos adicionais, sua distância ao novo centroide não se alterou significativamente. O referido evento teve "melhor" posicionamento (menor distância ao centroide do grupo) nos agrupamentos de Rio Negro (Figura 3), São Bento (Figura 4) e Fluviópolis (Figura 5).

A Figura 7 apresenta os hidrogramas das estações fluviométricas analisadas no estudo durante o evento de 2014; observa-se que, mesmo mantidas as magnitudes das vazões das estações na figura, o pico de cada evento se destaca dentre suas vazões mais baixas; isso não acontece em todos os eventos avaliados. O evento de 2014, historicamente, apresentou a terceira maior vazão observada em União da Vitória, assim como ocorrido nas estações de Fluviópolis, São Mateus do Sul e Rio





Negro, mas em São Bento, o evento ocupou a quarta posição. A cheia de 2014 foi um evento do grupo 4 que passou a pertencer ao grupo 3, o que ocorreu apenas na simulação com São Bento (Figura 4).

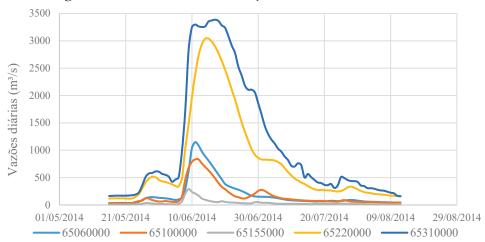


Figura 7. Evento de 2014 nas estações fluviométricas analisadas.

Avaliando os indicadores normalizados dos agrupamentos, para o evento de 2014, observa-se que São Bento apresentou, inclusive, a vazão normalizada mais baixa dentre todas as estações no mesmo ano; isso justifica a alteração de grupo do evento no agrupamento com São Bento. Reforça-se que Porto Amazonas não dispunha de dados para este evento, por isso, os eventos do grupo 4 neste agrupamento (Figura 1 à esquerda) mudaram suas posições em relação ao centroide.

No entanto, outros eventos apresentaram alterações de grupo e pioraram seu posicionamento, ou seja, aumentaram a sua distância em relação ao centroide do novo grupo, como, por exemplo, o evento de 1938; isto ocorreu em todas as *suas* mudanças de grupo (Figuras 1, 2 e 4). Observando-se os valores da Tabela 2 e as Figuras 1, 2 e 4 (à esquerda), pode-se dizer que as classificações de 1938 para Porto Amazonas, São Mateus do Sul e São Bento estão distantes dos grupos 2 e 3, mas podem compartilhar características com ambos, embora tenham sido alocadas no grupo 2 nos agrupamentos adicionais.

Tabela 2. Indicadores e grupos de pertencimento do evento de 1938.

	19	Commo		
Cenário Base	Quv	I_2	Grupo	
Cenario dase	0,3955	0,3760	3	
Agrupamentos adicionais	Quv	Qestação		
São Mateus do Sul	0,3955	0,1663	2	
São Bento	0,3955	0,1249	2	
Rio Negro	0,3939	0,4023	3	
Porto Amazonas	0,3955	0,1608	2	

Por fim, é válido avaliar o comportamento dos eventos de 1983 e 1992 nos agrupamentos realizados, apesar de terem se mantido no grupo 4 em todas as classificações. O evento de 1983 é um





evento historicamente conhecido pelos registros extremos de vazões atingidas. O pico da cheia de 1983 chegou a 5.156 m³/s, 10 vezes maior que a vazão média observada em União da Vitória (Steffen e Gomes, 2023a); o evento de 1992 é o segundo maior evento em termos de vazão em União da Vitória (pico de 3.953 m³/s). As vazões normalizadas de 1983 atingiram o limite superior do intervalo de normalização (0 a 1) para todas as estações, exceto São Bento (0,7114), que atingiu o valor normalizado igual a 1 para a cheia de 1992. Nesta estação, portanto, as cheias de 1983 e 1992 inverteram suas posições, segundo as vazões máximas.

CONCLUSÕES

As alterações dos agrupamentos adicionais observadas podem ser justificadas pelas mudanças das posições dos centroides dos agrupamentos adicionais, quando comparadas às dos centroides do cenário base.

Hidrologicamente, a migração dos eventos dentre os grupos se justifica pelos indicadores de cheia utilizados em cada cenário: a taxa média de recessão se relaciona ao volume e à duração dos eventos; as vazões máximas representam a relação com os eventos críticos dos afluentes.

Os eventos classificados no cenário base, localizados próximos às fronteiras de seus grupos, foram os mais sensíveis à mudança nos agrupamentos adicionais, por apresentarem maiores distâncias ao centroide do grupo a que pertencem quando comparados aos demais integrantes do grupo. Para os eventos de cheias observados em 1957, 1972 e 1989 no rio Iguaçu em União da Vitória esta constatação ficou mais evidente.

Eventos deslocados para outros grupos nos agrupamentos adicionais ainda podem compartilhar informações com os demais, conforme preconiza a lógica difusa, principalmente com suas classes de origem. Um exemplo é o evento de cheia observado no ano de 2009.

O evento de cheia observado no ano de 1938 apresentou alteração de grupo em três classificações, contudo, sua distância com relação aos novos centroides aumentou. Sua alocação em outro grupo pode ter ocorrido pela convergência do método e, por isso, o evento compartilha característica com os dois grupos.

O evento de cheia observado no ano de 2014 teve uma única mudança de grupo, em São Bento, que se deu pela ocorrência de uma vazão normalizada consideravelmente inferior às das demais estações. Pelo mesmo motivo, o grupo dos eventos de cheia observados em 1983 e 1992 (que compõem o grupo de maiores vazões no cenário base), teve seu centroide alterado de posição, apesar dos eventos não terem se deslocado entre os grupos.

Por fim, é necessário evidenciar que o cenário base poderia ser construído com outros indicadores de cheia, podendo se tornar, inclusive, um problema de múltiplas dimensões, como, por exemplo, nas aplicações de Steffen e Gomes (2018) e Vivan (2020). Contudo, a utilização de dois





indicadores facilita a interpretação, pela possibilidade de uma visualização gráfica simplificada do problema. Ainda, os indicadores utilizados (vazões máximas) poderiam ser considerados em conjunto para realização de um único agrupamento; isto retornaria outro cenário de múltiplas dimensões, porém, uma dificuldade aqui encontrada, além da representação, foi a simultaneidade dos dados entre todas as estações, o que reduziria significativamente o tamanho da amostra.

Contudo, apesar das limitações, a abordagem facilita o entendimento dos mecanismos de cheia na bacia estudada e, diante dos resultados, pode-se afirmar que os agrupamentos nos quais foram incorporadas as vazões máximas observadas nas estações Fluviópolis, Porto Amazonas e Rio Negro foram aqueles que mais se assemelharam ao agrupamento inicial com dados somente de União da Vitória, agrupamento considerado como cenário base para efeito de comparação, enquanto que os agrupamentos nos quais foram incorporadas as vazões máximas observadas nas estações São Mateus do Sul e São Bento foram os que mais se diferenciaram do cenário base.

Em termos hidrológicos, assumindo, por hipótese, que o cenário base representasse o agrupamento das cheias em termos de criticidade, então os agrupamentos nos quais foram incorporadas as vazões máximas observadas nas estações Fluviópolis, Porto Amazonas e Rio Negro representariam os agrupamentos que melhor reproduziriam a criticidade dos eventos observados em União da Vitória como eventual subsídio à modelagem hidrológica para a bacia em estudo.

REFERÊNCIAS BIBLIOGRÁFICAS

BEZDEK, J. C.; EHRLICH, R.; FULL, W. (1984). "FCM: the fuzzy c-means clustering algorithm". Computers & Geosciences, v. 10, n. 2-3, pp. 191-203.

HAIR JUNIOR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E.; TATHAM, R. L. (2009). *Análise multivariada de dados*. 6. ed. Porto Alegre: Bookman. Tradução: Adonai Schlup Sant'Anna.

SATO-ILIC, M.; JAIN, L. C. (2006). *Innovations in fuzzy clustering*. Heidelberg: Springer.

STEFFEN, P. C.; GOMES, J. (2018). "Clustering of historical floods observed on Iguaçu River, in União da Vitória, Paraná". Revista Brasileira de Recursos Hídricos, v. 23, pp. 1–12. DOI: 10.1590/2318-0331.231820170107.

STEFFEN, P. C.; GOMES, J. (2023a). "Características de cheias observadas na bacia do rio Iguaçu – parte 1: agrupamento de eventos" in Anais do XXV Simpósio Brasileiro de Recursos Hídricos, Aracaju, Nov., 2023.

VIVAN, A. A. (2020). Lógica difusa aplicada ao agrupamento de cheias históricas observadas no rio Iguaçu na cidade de União da Vitória: comparação entre métodos de agrupamento k-means e fuzzy c-means. 283 f. Dissertação (Mestrado em Engenharia de Recursos Hídricos e Ambiental), Universidade Federal do Paraná, Curitiba.

WANG, L. N.; CHEN, X. H.; SHAO, Q. X.; LI, Y. (2015). "Flood indicators and their clustering features in Wujiang River, South China". Ecological Engineering, v. 76, pp. 66-74. DOI: 10.1016/j.ecoleng.2014.03.018.

ANEXO 4 – EVALUATING DIFFERENT STRATEGIES FOR MACHINE LEARNING TRAINING APPLIED TO FLOW FORECASTING BASED ON CLUSTERING OF FLOOD EVENTS

Artigo científico submetido à Revista Brasileira de Recursos Hídricos.



EVALUATING DIFFERENT STRATEGIES FOR MACHINE LEARNING TRAINING APPLIED TO FLOW FORECASTING BASED ON CLUSTERING OF FLOOD EVENTS

Journal:	al: RBRH, Brazilian Journal of Water Resources				
Manuscript ID	RBRH-2024-0087				
Manuscript Type:	Original Article				
Date Submitted by the Author:	29-Aug-2024				
Complete List of Authors:	Steffen, Patrícia; Universidade Federal de Mato Grosso, Engenharia Civil, ICET/CUA Gomes, Julio; Universidade Federal do Parana, Departamento de Hidraulica e Saneamento Kaviski, Eloy; Universidade Federal do Parana Detzel, Daniel; UFPR, Department of Hydraulics and Sanitation				
Keyword:	Artificial Intelligence techniques, Data-driven intelligent models, Fuzzy C-Means algorithm, Hydrological modeling, Support Vector Regression algorithm				

SCHOLARONE™ Manuscripts

1 EVALUATING DIFFERENT STRATEGIES FOR MACHINE

2 LEARNING TRAINING APPLIED TO FLOW FORECASTING

BASED ON CLUSTERING OF FLOOD EVENTS

4 Abstract

- Hydrological modeling has advanced with the rise of data-driven models, largely due to the widespread use of Artificial Intelligence (AI) tools. Given the complexity of modeling process, researchers are exploring new approaches to better understand hydrological events development. This paper presents a novel hydrological modeling approach for discharge prediction based on flood events clustering. Combined with Machine Learning techniques, river flow simulation is optimized through increased data similarity within clustered events. Using daily mean discharge data from 1964 to 2015 in União da Vitória (Iguaçu River basin, Paraná State, Brazil), the Fuzzy C-Means algorithm clustered flood events into three groups. So, five models were trained: one for the complete series, one for all flood events, and one for each cluster. The Support Vector Regression algorithm was used to develop AI models, that had better performance in predicting discharge categories for which they were trained and showed similar efficiency in predicting the entire series when using a 1-day forecast window.
- 18 Keywords: Artificial Intelligence techniques. Data-driven intelligent models. Fuzzy C-
- 19 Means algorithm. Hydrological modeling. Support Vector Regression algorithm.

20 Resumo

- 21 A modelagem hidrológica tem registrado avanços com o aumento das aplicações de
- 22 modelos de análise de séries temporais, que vem se difundindo devido à utilização de
- 23 técnicas de Inteligência Artificial (AI). Diante da complexidade envolvida ao processo de

modelagem, pesquisadores buscam desenvolver abordagens capazes de melhor compreender o processo de formação de eventos hidrológicos. O presente trabalho apresenta uma nova abordagem para modelagem hidrológica, baseada em uma classificação inicial de cheias. Combinada às técnicas de Aprendizagem de Máquinas, a pesquisa visa otimizar a simulação de vazões, por meio do aumento da similaridade dos dados após o agrupamento. Partindo das vazões médias diárias de 1964 a 2015 em União da Vitória (bacia hidrográfica do rio Iguaçu, estado do Paraná, Brasil), o algoritmo Fuzzy C-Means agrupou eventos de cheia em três grupos. Assim, foram treinados cinco modelos: um de série completa, um com todos os eventos de cheia e um para cada grupo formado. O algoritmo de Regressão por Vetores de Suporte foi aplicado para obtenção dos modelos de AI, que apresentaram melhor desempenho ao prever vazões de mesma categoria para as quais foram treinados e apresentaram desempenhos similares para a previsão de todas as séries para um horizonte de previsão de um dia.

Palavras-chave: Técnicas de Inteligência Artificial. Modelos inteligentes de análise de séries temporais. Algoritmo Fuzzy C-Means. Modelagem hidrológica. Algoritmo de Regressão por Vetores de Suporte.

1. Introduction

- 41 Hydrological modeling has achieved significant breakthroughs due to the computational
- 42 revolution (Singh, 2018). In this context, hydrological forecasting is generally conducted
- using two primary approaches (Mohammadi et al., 2020; Snieder et al., 2020; Wang et
- al., 2021; Difi et al., 2023): (i) traditional models and (ii) data-driven models.
- 45 In classical hydrological modeling, traditional models analyze hydrological
- characteristics and physically describe the confluence of flows, giving their parameters a
- 47 certain degree of physical relevance (Ding et al., 2020). However, the calibration of these

models is based on observed data and requires a high level of familiarity with the model from the researcher (Ding et al., 2020; Ebtehaj and Bonakdari, 2022). Moreover, incorrect parameter estimation can significantly increase errors in this class of models (Luppichini et al., 2019; Luppichini et al., 2022). Ebtehaj and Bonakdari (2022), while acknowledging that traditional models have produced good results for flood forecasting, also highlight several limitations. These include the manual collection of data, which is often stored in a disorganized manner, leading to insufficiently sized data series – especially problematic in remote regions where data availability is already limited. Luppichini et al. (2022) further note that physically-based models face challenges due to inherent heterogeneity of natural systems. On the other hand, with the growth of hydrological infrastructure, increased data availability, and advances in computational development (Singh, 2018; Bai et al., 2021), the capacity of empirical models has expanded, leading to the emergence of so-called data-driven models – or time series models. These models use system data to identify connections between variables (input, internal, and output) without explaining the system's physical behavior. In other words, the information source for these models is the data series itself (Adnan et al., 2019; Fathian et al., 2019; Ding et al., 2020; Bai et al., 2021; Ebtehaj and Bonakdari, 2022). Time series models capture both linear and non-linear relationships between discharges and their mathematical parameters (Zhang et al., 2018; Adnan et al., 2019). Commonly used approaches include the least squares method, multiple linear regression, and models such as Autoregressive (AR), Moving Average (MA), and Autoregressive Moving Average (ARMA). However, these methods typically only account for the linear relationships between inputs and outputs, often resulting in outputs that are insufficient

and fail to address the inherent non-linearity in the systems (Khosravi et al., 2021).

To address these disadvantages, automated time series models using Artificial Intelligence (AI) have been developed and explored for hydrological modeling and discharge forecasting (Bai et al., 2021; Khosravi et al., 2021). These efforts have focused on Computational Intelligence (CI), which can complement or replace traditional physical models (Fathian et al., 2019; Bai et al., 2021). Such models, often referred to as datadriven intelligent models (Ding et al., 2020; Khosravi et al., 2021), are notable for their ability to handle large data series and accommodate data at different scales (Bai et al., 2021; Khosravi et al., 2021). Among the most popular CI techniques are neural networks, fuzzy rule-based systems, and genetic algorithms (Mahdavi-Meymand et al., 2023). Among CI methods, Machine Learning (ML) techniques involve the automatic detection of significant patterns in data series. These techniques are categorized based on different learning objectives (Xu and Liang, 2021), including classification, regression, and clustering (Ibrahim et al., 2022; Mahdavi-Meymand et al., 2023). Over the last two decades, these techniques have yielded promising results for hydrological problems, including: (i) Support Vector Machines (SVM) and Support Vector Regression (SVR) (Li et al., 2019; Chen et al., 2021; Niu and Feng, 2021; Kim et al., 2022; Difi et al., 2023; Sharma et al., 2023); (ii) Random Forest (RF) (Fathian et al., 2019; Saadi et al., 2019; Schoppa et al., 2020; Desai and Ouarda, 2021; Sharma et al., 2023); (iii) Artificial Neural Networks (ANN) (Zhang et al., 2018; Fathian et al., 2019; Snieder et al., 2020; Brêda et al., 2021; Chen et al., 2021; Lima and Scofield, 2021); (iv) Long Short-Term Memory (LSTM) (Ding et al., 2020; Chen et al., 2021; Kim et al., 2022; Sharma et al., 2023; Zakhrouf et al., 2023); (v) Extreme Learning Machine (ELM) (Li et al., 2019; Ribeiro et al., 2020; Yaseen et al., 2020; Niu and Feng, 2021; Ebtehaj and Bonakdari, 2022; Feng et al., 2022; Difi et al., 2023); (vi) Multivariate Adaptive Regression Splines (MARS)

(Niu and Feng, 2021; Sharma et al., 2023); and (vii) Adaptive Neuro-Fuzzy Inference System (ANFIS) (Mohammadi et al., 2020; Shukla et al., 2022; Samantaray et al., 2023). Most applications involve Artificial Neural Networks (ANNs) and their variations, such as Extreme Learning Machines (ELM) and Long Short-Term Memory (LSTM), as well as Support Vector Machines (SVM) and its regression variant, Support Vector Regression (SVR). Generally, research has focused on comparing these techniques. Recently, LSTM, ELM, and SVM (or SVR) have been frequently compared in studies (Li et al., 2019; Chen et al., 2021; Niu and Feng, 2021; Kim et al., 2022; Difi et al., 2023; Sharma et al., 2023). Other applications of these models in hydrology can be found in Adnan et al. (2019), Kurian et al. (2020), Islam et al. (2021), and Ebtehaj and Bonakdari (2022). Additionally, Mahdavi-Meymand et al. (2023) provide a literature review of the last decade's publications applying Machine Learning techniques to various hydrological objectives. For discharge forecasting, Adnan et al. (2019) highlight that recent decades have seen successful applications of methods such as ANN, ANFIS, MARS, SVM, genetic algorithms (GA), and their hybrid models. These methods are effective in identifying nonlinearity in flood formation processes. In this work, SVR algorithm was chosen for its satisfactory results and straightforward computational implementation. However, selecting the model is only one aspect; it also requires consideration of other details in forecasting hydrological modeling. Thus, this manuscript evaluates a hydrological modeling approach for discharge prediction, supported by previous flood events clustering and flood events characterization. In this approach, events within the same cluster share similar characteristics and differ from those in other clusters. By clustering events with similar features, the approach constrains the predicted event's characteristics to those of its respective group (Joo et al., 2021), thereby reducing prediction variability.

In this context, Xu and Liang (2021) noted that machine learning methods for clustering and classification are often employed to support regression methods. However, these approaches typically serve as alternatives to regression techniques rather than integrating group hydrological modeling.

No application like the one proposed in this paper was found in the literature. However, Joo et al. (2021) noted that cluster analysis has been widely used in hydrology. They presented various studies that utilized cluster analysis for applications such as forecasting ungauged hydrological data, flood frequency analysis, hydrological modeling, flood forecasting, and classifying hydrological and catchment similarities.

According to Tarasova et al. (2019), classifying flood events provides a better understanding of flood generation mechanisms, which are not always clearly defined at the catchment scale. There are three main categories for flood generation mechanisms: hydroclimatic, characterized by large-scale circulation patterns and atmospheric conditions at the event's onset; hydrological, defined by the catchment's precipitation patterns and its antecedent conditions; and hydrograph-based, which considers formation mechanisms through their effects on hydrograph characteristics (Tarasova et al., 2019). Tarasova et al. (2019) reviewed existing classifications of flood events, discussing their validity, limitations, and transferability across temporal and spatial scales.

To aid in understanding the mechanisms of flood formation, clustering emphasizes similar characteristics among events within the same group. When used as a preliminary step in discharge forecasting, this two-step framework narrows the focus of forecasts typically produced by a single model. Consequently, the further an event is from a cluster of observed floods, the less it is likely to resemble the model associated with that group. In other words, when predicting the occurrence of an event, it is expected that the event will closely match its own model and differ more significantly from others.

Thus, the primary goal of cluster analysis is to use data characteristic variables to group data into classes in the most appropriate way, ensuring that similar objects are in the same class, thereby reducing the number of observations in a sample (Morettin and Singer, 2023). This process also requires defining the similarity within a class and the dissimilarity between classes, which is a complex task (Ezugwu et al., 2022). In this study, we use the Fuzzy C-Means (FCM) algorithm to assign each element to a group from a predefined number of groups (Mosavi et al., 2021). Hence, the objective of this paper is to develop a new hydrological modeling approach for discharge prediction based on the previous characterization and clustering of flood events. This approach aims to optimize the simulation of extreme maximum and minimum river flows by reducing the amplitude of the clustered data. Additionally, the forecasting process is automated, thereby eliminating the inherent disadvantages of statistical time series models and reducing processing time. The proposed hydrological modeling approach is demonstrated through a real-world application. Daily mean discharges from 1964 to 2015 in União da Vitória gauging station (Iguaçu River basin, Paraná State, Brazil) were used, and the SVR algorithm was applied to develop AI models for 1-day discharge forecasting. The FCM algorithm clustered flood events into three groups, number previous defined. Hence, five models were trained: one for the Complete Series, one for All Flood Events, and one for each of the three Clusters. The results depend on the clustering, the chosen hydrological model, and the forecast

2. Case Study

horizon. So, the proposed approach is allowed for the use of various hydrological models.

The chosen study area encompasses the Iguaçu River basin up to the União da Vitória gauging station (65310000), selected due to the historically recurrent flood events in the basin, particularly in the city of União da Vitória, Paraná State, Brazil (Figure 1).

The Iguaçu River flows almost entirely across the state of Paraná, in an east-west direction, spanning 1,320 km before discharging into the Paraná River (Mine and Tucci, 2002). The Iguaçu River basin covers a total drainage area of 70,800 km², divided between the states of Paraná and Santa Catarina in Brazil, with a small portion extending into Argentina (Rocha, 2012).

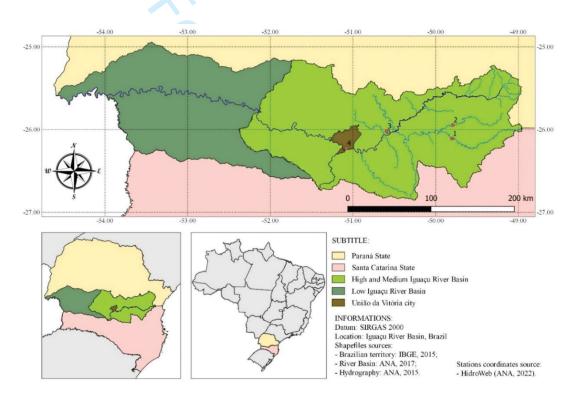


Figure 1. Location of União da Vitória city in Paraná State and Iguaçu River Basin, along with the positions of gauging stations.

The Iguaçu River is formed by the confluence of the Iraí and Atuba Rivers in Curitiba Metropolitan Region. Upstream of União da Vitória, the main contributors to the Iguaçu River are the Potinga River on the right bank, and the Negro and Timbó Rivers on the left bank. Other tributaries, including the Verde, Itaqui, da Vargem, and Claro Rivers (on the right bank), and the Passa Dois and Paciência Rivers (on the left bank), provide smaller

contributions to the Iguaçu River upstream of the União da Vitória gauging station. At this point, the drainage area is approximately 24,200 km² (ANA, 2023), accounting for more than one-third of the entire basin area.

Based on the identified tributaries of the Iguaçu River, existing gauging stations in the river basin were investigated using data from the National Water and Sanitation Agency database (ANA, 2023). The selection of stations primarily considered factors such as sample size, location, data quality, and the consistency of their records (Tozzi and Fill, 2020). In addition to União da Vitória (65310000), data were collected from Fluviópolis (65220000) on the Iguaçu River, Rio Negro (65100000) on the Negro River, and São Bento (65155000) on the Várzea River. São Bento station data were filled in using information from Rio da Várzea dos Lima station, which is located on the same river.

Figure 1 displays these stations with their IDs: Rio Negro (ID 1), São Bento (ID 2), Fluviópolis (ID 3), and União da Vitória (ID 4). Table 1 provides their descriptive statistics.

Table 1. Descriptive statistics of the gauging stations (ANA, 2023).

Code	ID	Stations name	Periods	n	\overline{Q} (m ³ /s)	q	Min	Max	S (m ³ /s)	а
stations	ID	Stations name	1 crious	(years)	Q (III /3)	$(L/s/km^2)$	(m^3/s)	(m^3/s)	5 (m /s)	u
65100000	1	Rio Negro	1931-2015	72	73.0	21.2	7.7	953.9	68.7	3.24
65155000	2	São Bento	1931-2015	84	35.8	17.9	6.0	482.3	28.8	2.84
65220000	3	Fluviópolis	1964-2015	51	399.7	21.5	34.5	3819.4	346.5	2.53
65310000	4	União da Vitória	1931-2015	84	492.2	20.3	46.2	5156.7	439.5	2.51

n – series size; \bar{Q} – mean discharge; q – mean specific flow rate; S – standard-deviation; a – skewness.

To account for the travel time of water from stations ID 1, ID 2, and ID 3 to station ID 4, a correlation analysis was performed (Table 2). Among all stations and lags, including lag 0, ID 3 exhibited the highest correlation with ID 4, although this correlation gradually decreases as the lag increases. Station ID 3 is the closest to ID 4 (Figure 1) and is in the Iguaçu River. However, with a distance about 100 km between these stations, the estimated travel time is less than 1 day and cannot be precisely defined on a daily scale.

Table 2. Correlations between the mean daily discharges at União da Vitória (ID 4) and the upstream gauging stations for different time lags. The highest correlations are highlighted.

Lag	Gauging Stations (ID)								
Lug	1	2	3						
0	0.8411	0.8246	0.9781						
1	0.8572	0.8300	0.9696						
2	0.8617	0.8266	0.9523						
3	0.8588	0.8172	0.9286						

Station ID 1 exhibited a higher correlation with ID 4 at lag 2, which can be attributed to the distance between the stations and its location on the Negro River. ID 1 consistently showed higher correlations with ID 4 than ID 2 across all lags. This is because ID 2 is situated on a tributary of the Negro River, which explains its lower correlation.

3. Material and Methods

The methodological processes are summarized in Figure 2 and are detailed in the following sections.

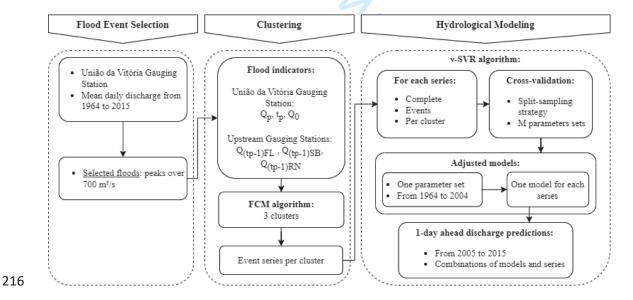


Figure 2. Workflow for 1-day-ahead discharge prediction at União da Vitória.

3.1 Flood events selection

Flood events were selected from the mean daily discharge series of the ID 4 gauging station, observed from 1964 to 2015, which corresponds to the simultaneous data record

period among all the river stations considered in the study. All observed floods with peaks exceeding 700 m³/s were considered, which corresponds to 20% on the flow duration curve. The mean daily discharge at ID 4 has a flow duration of 35% (about 500 m³/s), whereas the overflow threshold of the main channel of Iguaçu River at this location is 1,387 m³/s, corresponding to 5% (Steffen and Gomes, 2018).

The events were delineated by identifying their respective start, end, and peak dates, along with their peak and initial discharges. This selection yielded a total of 194 events with

with their peak and initial discharges. This selection yielded a total of 194 events with these characteristics. Over 52 years of observation, nearly 30% of the events (57 events) resulted in overbank flows (peak discharge > 1,387 m³/s), while 54.1% of the peak discharges were equal to or exceeded 1,000 m³/s. Additionally, almost 46% of the events (89 events) had peak discharges between 700 m³/s and 1,000 m³/s.

3.2 Clustering

Fuzzy C-Means (FCM) algorithm application requires flood indicators, which are the characteristic variables of the events used to group the data. For the events selected at ID 4 and the upstream stations, six indicators were determined: peak flow (Q_p), peak time (t_p), and initial flow (Q_0) at ID 4, as well as observed discharges at the upstream stations, which are adjusted for the peak time (t_p) at ID 4 and the lags defined from the correlation analysis – lag 1 to ID 3 (Q_{tp-1})_{ID-3}, lag 1 to ID 2 (Q_{tp-1})_{ID-2}, and lag 2 to ID 1 (Q_{tp-2})_{ID-1}. In FCM, fuzzy logic is incorporated through an uncertainty parameter, which determines the degree of similarity of events to each group. Clustering is considered fuzzy when the uncertainty parameter exceeds 1 (Askari, 2021; Mosavi et al., 2021). Using linearly standardized indicators (Mosavi et al., 2021), FCM is defined by the following iterative process (Steffen and Gomes, 2018; Askari, 2021):

1. The application of linear normalization to the indicators is performed using Equation (1), where x_{ij} and x_{ij}^0 are the normalized and non-normalized values of the *i*-th observation of the *j*-th indicator, respectively, and $x_{j\ min}^0$ and $x_{j\ max}^0$ are the minimum and maximum values of the *j*-th indicator, respectively.

$$x_{ij} = \frac{x_{ij}^{o} - x_{j \, min}^{o}}{x_{j \, max}^{o} - x_{j \, min}^{o}} \tag{1}$$

- 2. The number of clusters was predefined as c = 3, and the degree of uncertainty was set to r = 1.25.
- 3. An initial random fuzzy partition matrix (U₀) is defined. This matrix, with *j* rows and *c* columns, represents the degree of membership of the *j*-th sample observation to each of the *c* predefined clusters.
 - 4. The cluster centroids (V) are calculated, with their coordinates representing a weighted average of the indicators for each cluster. Equation (2) defines the cluster centroids, where v_{kj} is the coordinate of the *j*-th indicator of the *k*-th centroid, and μ_{ki} is the *i*-th element of the membership matrix U for the *k*-th cluster.

258
$$v_{kj} = \frac{\sum_{k=1}^{c} (\mu_{ki})^r x_{ij}}{\sum_{k=1}^{c} (\mu_{ki})^r}$$
 (2)

5. The Euclidean distances from the events to the centroids (**D**) are calculated using Equation (3), where d_{ik} represents the Euclidean distance of the *i*-th observation to the *k*-th cluster, and *d* denotes the total number of flood indicators, which defines the clustering dimension.

$$d_{ik} = \|x_i - v_k\|^2 = \sqrt{\sum_{j=1}^d (x_{ij} - v_{kj})^2}$$
 (3)

6. The objective function (**J**) is calculated using Equation (4), where *n* represents the total number of input observations.

$$J = \sum_{k=1}^{c} \sum_{i=1}^{n} (\mu_{ki})^{r} (d_{ik})^{2}$$
 (4)

7. The fuzzy partition matrix (U) is updated using Equation (5).

268
$$\mu_{ki} = \left[\sum_{t=1}^{c} \left(\frac{d_{ik}}{d_{tk}}\right)^{\frac{2}{r-1}}\right]^{-1}$$
 (5)

- 8. Verification of the algorithm's stopping criterion: the process stops when the absolute difference between the maximum degrees of similarity of the observations (between the last two stages of the iterative process) is less than the maximum tolerated ($\varepsilon_t = 10^{-5}$), indicating that the minimum value of the objective function has been reached.
- 9. If the stopping criterion is not met, return to Step 4 and repeat the process until the condition is satisfied.

3.3 Goodness of fit measures

- 277 To evaluate the adjustments and forecasts, the following metrics were applied: Nash-
- 278 Sutcliffe Efficiency (NS) (Unduche et al., 2018; Lawin et al., 2019; Liang et al., 2019;
- 279 Althoff et al., 2021; Brêda et al., 2021; Mosavi et al., 2021; Moura et al., 2022), Kling-
- Gupta Efficiency (KG) (Gupta et al., 2009; Althoff et al., 2021; Moura et al., 2022;
- Lappicy and Lima, 2023) and Mean Absolute Relative Error (MARE) (Staudinger et al.,
- 282 2011; Unduche et al., 2018; Liang et al., 2019; Althoff et al., 2021; Moura et al., 2022;
- Letessier et al., 2023). These metrics were chosen to evaluate different phases of the
- 284 hydrographs, ensuring a comprehensive assessment of the model's performance.

3.4 v-SVR algorithm

- Support Vector Machines (SVM) have gained popularity over the last four decades due
- 287 to their efficiency and processing speed in the realm of modern statistical learning

(Ibrahim et al., 2022). The evolution of the SVM approach is well-documented, with Freitas (2016) providing a comprehensive history of its development – from the early algorithms for pattern recognition in the 1950s to its current form, grounded in Statistical Learning Theory.

Support Vector Machines (SVM) is a classification method rooted in the Vapnik-Chervonenski theory, which was developed to address supervised learning problems. SVM is part of the generalized linear classification family (Ibrahim et al., 2021; Roy and Chakraborty, 2023), and, when adapted for non-linear regression, is known as Support Vector Regression (SVR). This adaptation is particularly effective for flow forecasting tasks. Based on Vapnik's ε-indifferent cost function, Schölkopf et al. (2000) introduced a significant modification to the SVR algorithm, known as ν-SVR. This version automatically minimizes the ε parameter, making the model less dependent on manually adjusting this parameter through optimization algorithms.

In v-SVR algorithm, v is the parameter responsible for controlling the number of support vectors – or the number of errors in the solution (Schölkopf et al., 2000); considering that $v \in (0,1]$, the closer to 1, the greater the number of support vectors and, consequently, the greater the tolerance to errors (Chang and Lin, 2001; Chang and Lin, 2022). From the set of parameters, the v-SVR algorithm aims to minimize Equation (6) (Schölkopf et al., 2000; Chang and Lin, 2001; Chang and Lin, 2022):

307
$$\tau(\mathbf{w}, \xi^*, \varepsilon) = \frac{1}{2} \|\mathbf{w}\|^2 + C \left(\nu \varepsilon + \frac{1}{m} \sum_{i=1}^m (\xi_i + \xi_i^*) \right)$$
 (6)

308 subject to:

$$((\mathbf{w} \cdot x_i) + b) - y_i \le \varepsilon + \xi_i \quad (7)$$

310
$$y_i - ((\mathbf{w} \cdot x_i) + b) \le \varepsilon + \xi_i^*, \quad (8)$$

where parameter C is the constant that defines the degree of cost penalty when there is a training error – C is also called as the regularization parameter and must be defined a priori; ξ_i and ξ_i^* are greater than zero and correspond to the slack variables that, respectively, specify the highest and lowest training errors, subject to tolerance; and, ε is the tolerance – greater than zero.

Therefore, to obtain the regression equation capable of predicting flows, NuSVR class of *sklearn.svm* module of Python was used, which was implemented using the LIBSVM library by Chang and Lin (2022). The parameters for calibration in this paper are: (i) the kernel function, (ii) the kernel coefficient, called Gamma (γ), (iii) the ν (Nu) parameter defined in the interval (0, 1], delimiting an upper limit for the training errors fraction, and a lower limit for the support vectors fraction, and (iv) the regularization parameter C.

- The Radial Basis Function (RBF) was selected as kernel function. RBF is less difficult numerically, allows non-linear samples to be mapped into a higher dimensional space and has fewer parameters to be calibrated if compared to the polynomial function (Freitas, 2016).
- In the calibration phase, C and γ parameters were combined, one by one, to identify the best combination among the 64 possibilities, considering:

328
$$C = [0.01, 0.05, 0.1, 0.5, 1.0, 2.0, 5.0, 10.0]$$

329
$$\gamma = [0.001, 0.01, 0.05, 0.1, 0.33, 0.5, 1.0, 5.0]$$

- To the other parameters of the NuSVR class, their respective standardized specifications were assigned, including the v parameter, whose default value is 0.5.
 - Given the parameters and correlations (Table 2) previously discussed, the SVR model for predicting the discharge at União da Vitória one day ahead uses the following input

variables: the discharges from Rio Negro using lag 2 (Q_{tp-2})_{RN}, and the discharges from São Bento (Q_{tp-1})_{SB}, Fluviópolis (Q_{tp-1})_{FL}, and União da Vitória (Q_{tp-1})_{UV}, using lag 1.

3.5 Train, test, and cross-validation

- For data series categories under analysis, SVR algorithm trained and tested different models: one for the complete series, one for all events series; and one for each event of the three clusters series.
- The period from 1964 to 2004 was subdivided into training and testing sets using a split sampling strategy based on a 20-year window (Hallouin et al., 2020). This approach divides the data into approximately equal intervals for training and testing stages, ensuring that the largest flood events are included in at least one of these stages.
 - Eight samples were generated to account for the largest historical flood event in 1983. The division was made as follows: Sample 1 training period from 1964 to 1984, and testing period from 1985 to 2004; Sample 2 training period from 1969 to 1989, and testing periods from 1990 to 2004 and from 1964 to 1968; Sample 3 training period from 1974 to 1994, and testing periods from 1995 to 2004 and from 1964 to 1973; Sample 4 training period from 1979 to 1999, and testing periods from 2000 to 2004 and from 1964 to 1978; Sample 5 training period from 1985 to 2004, and testing period from 1964 to 1984; Sample 6 training periods from 1989 to 2004 and 1964 to 1968, and testing period from 1969 to 1988; Sample 7 training periods from 1994 to 2004 and from 1964 to 1973, and testing period from 1974 to 1993; and Sample 8 training periods from 1999 to 2004 and from 1964 to 1978, and testing period from 1979 to 1998.
 - For each sample and each of the 64 combinations, the NuSVR algorithm was executed for the training period, generating the models later applied to the corresponding testing

period. The 32 best parameter sets were selected, and their models were applied to the testing period. The metrics for evaluating the training and test results were calculated, along with their main statistics: mean, maximum, and minimum flows; standard deviation; skewness; and volume ratio (simulated/observed).

3.6 Adjusting models

The parameter set selected to validate the models was obtained by averaging the quality measures results from the eight training and test samples. For each parameter set, the median of the averages was determined. Then, for each model, the parameter sets with performance lower than the median were identified. Finally, the parameter set that was repeated across all models was selected. This same set was used for all models.

Therefore, to avoid overfitting, the selected parameter set was the one with metrics close to the medians but capable of satisfactorily estimating discharges at ID 4, considering all five categories of models. When the performances of different parameter sets were close to each other, the set with the lowest C and γ values was adopted. Increasing these parameter values can speed up processing but does not necessarily improve performance.

Based on the established values of C and γ , each type of model was calibrated for the entire period of the sampling strategy (from 1964 to 2004). Simulations were then carried out for all series using each model, i.e., each type of calibrated model was applied to all samples: Complete, All Events, and by group.

Descriptive statistics and goodness-of-fit measures were calculated for each combination of model and simulated series. However, since each series has different sizes, proper comparisons were only possible by considering one series obtained from different models, rather than comparing simulations of different series using a single model.

3.7 Discharge prediction

Each calibrated model was applied to all discharge samples for the period from 2005 to 2015. Once again, descriptive statistics and goodness-of-fit measures were calculated for each combination of model and predicted series. A comparison analysis was possible only for one series due to size limitations encountered during the model adjustment phase.

4 Results and Discussions

4.1 Flood clustering

Table 3 shows the statistical parameters of the maximum similarity degrees of the events, with an ideal value of 1.0. The position of the events in the groups is determined by identifying the event's highest similarity degree (U) with the group. The largest number of events is found in Cluster 2; Cluster 3 had the lowest minimum value and average of the maximum similarity degrees, as well as the highest amplitude and standard deviation.

Table 3. Statistical parameters of the maximum similarity degrees (U).

Parameter	Cluster 1	Cluster 2	Cluster 3
Number of observations	70	102	22
Minimum	0.55	0.53	0.49
Maximum	1.00	1.00	1.00
Amplitude	0.45	0.47	0.51
Mean	0.96	0.95	0.88
Standard-deviation	0.10	0.11	0.15

Due to the minimum and maximum values observed in Table 3, Table 4 shows the distribution of the similarity degrees (U) between these extremes, through frequencies of occurrence for certain U intervals for each event series. Overall, 153 events (more than 78%) exhibited a degree of belonging to their group greater than 0.95, while 17 events (around 9%) were less than 0.70. For the events in Clusters 1 and 2, more than 80%

presented U values over 0.95. In contrast, events in Cluster 3 showed lower results, with around 54% having a degree of belonging greater than 0.95, 18% below 0.70, and approximately 9% in each of the other intervals.

Table 4. Frequencies of the intervals of maximum similarity degrees (U).

				U interva	ls		
		[0.95; 1.00]	[0.90; 0.95)	[0.80; 0.90)	[0.70, 0.80)	< 0.70	Sum
ts	Absolut frequency	153	6	12	6	17	194
All Events	Relative frequency	78.9%	3.1%	6.2%	3.1%	8.7%	100%
ter	Absolut frequency	59	1	5	0	5	70
Cluster 1	Relative frequency	84.3%	1.5%	7.1%	0.0%	7.1%	100%
ter	Absolut frequency	82	3	5	4	8	102
Cluster 2	Relative frequency	80.4%	2.9%	4.9%	3.9%	7.9%	100%
ter	Absolut frequency	12	2	2	2	4	22
Cluster 3	Relative frequency	54.5%	9.1%	9.1%	9.1%	18.2%	100%

Lastly, the coordinates of the centroids of each group are shown in Table 5. The observed peak flow at União da Vitória (Qp) and the observed lagged flows at upstream stations $[(Q_{tp-1})_{FL}, (Q_{tp-1})_{SB}, (Q_{tp-2})_{RN})]$ in Cluster 3 were higher than the respective coordinates in Clusters 1 and 2. Except for the initial flow (Q0) coordinate values, the other indicators remained close between their pairs for Clusters 1 and 2. The initial flow coordinate (Q0) for Cluster 2 was higher than the same coordinate for the other groups.

Table 5. Centroid coordinates of Clusters 1 to 3.

Cluster	t _p UV (days)	Qp UV (m ³ /s)	Q ₀ UV (m ³ /s)	Q _{tp-1} FL (m ³ /s)	Q _{tp-1} SB (m ³ /s)	Q _{tp-2} RN (m ³ /s)
1	12	1075.2	246.6	669.7	64.3	148.3
2	14	1087.9	434.3	730.4	66.9	150.0
3	41	2490.1	365.9	1934.8	130.2	353.2

Considering the entire data period (1964 to 2015), the statistical parameters (Table 6) were determined for the different data sets. It is observed that Clusters 1 and 2 present similar values for both mean and standard deviation. Additionally, the mean and standard deviation for "all events" data set fall between the values observed for Clusters 2 and 3.

Table 6. Descriptive statistics of the data series from the entire period (1964 - 2015).

	Data set										
	Complete	All events	Cluster 1	Cluster 2	Cluster 3						
Number of events	_	194	70	102	22						
Series size	18991	7356	2078	3508	1770						
Minimum	6.9	114.0	114.0	314.2	148.1						
Maximum	5156.7	5156.7	1939.5	1915.1	5156.7						
Mean	513.0	961.5	836.6	842.3	1344.3						
Standard-deviation	461.40	511.54	349.56	315.55	744.30						
Skewness	3.24	2.29	0.59	0.92	1.52						

In summary, clustering effectively identified similar patterns and characteristics within the events of each group. This resulted in reduced mean, standard deviation, and skewness for Clusters 1 and 2, while increasing the mean and standard deviation for Cluster 3 due to the greater magnitude of its events.

4.2 Cross-validation

Considering all eight samples from cross-validation, mean values of KG, NS, and MARE were calculated for the 64 parameter sets. For training stage, Figure 3 displays the results variation as heatmaps. Generally, C parameter above 2.0 yielded reasonable results across all γ values, except for Complete Series Model (Figure 3a), which showed reasonable indexes for a broader range of parameter sets, including those with C values below 2.0.

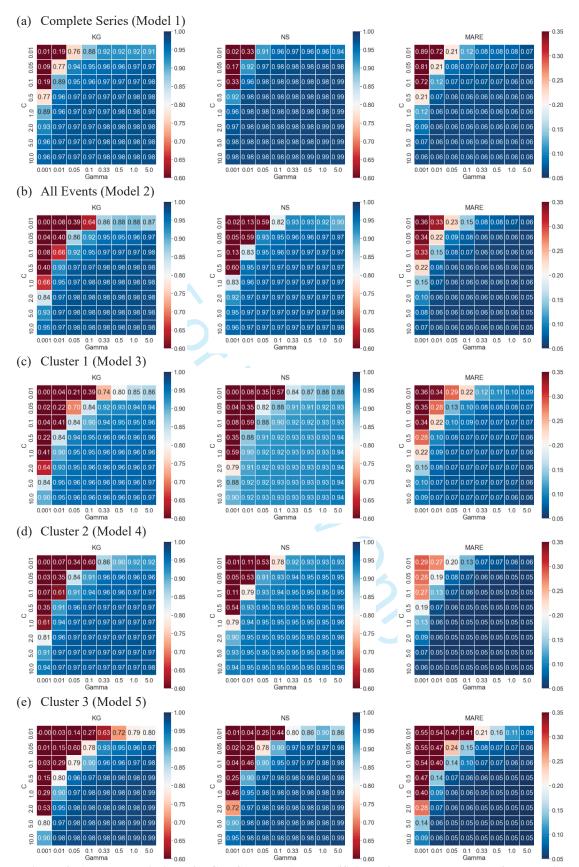


Figure 3. Average variation of KG, NS, and MARE coefficients for parameter sets of data series during the training phase.

Figure 3 shows that lower values of C and γ generally result in worse performance, with better results observed towards the bottom right corner, indicating increased C and γ parameters. The diagonal from the bottom left corner (C = 10.0 and γ = 0.001) to the top right corner (C = 0.01 and γ = 5.0) represents more uniform results. However, while increasing C and γ improves performance up to a point, further increases lead to minimal improvement and only enhance processing speed.

The variability between parameter sets and samples increased with model complexity.

For Complete Series Model (Figure 3a), the range of results across different parameter

sets was smaller compared to Cluster 3 Model (Figure 3e). Additionally, the results for

Complete Series Model showed more uniformity across samples than those for Cluster 3.

Since Figure 3 is based on the average of the sample coefficients, the results are

significantly influenced by the specific events included in each sample due to variations

in model input data. In other words, as the differences among the model's input data (i.e.,

different data samples) increase, the KG, NS, and MARE coefficients tend to diverge

from their averages.

In this context, the poorest performance of parameter sets was observed in models by group (Figures 3c, d, and e), particularly for Clusters 1 and 3. This trend, along with the heatmaps' behavior, underscores the importance of cross-validation and the need to minimize the risk of underfitting or overfitting. Consequently, parameter sets with values located in the center of the heatmaps generally produce better results in training phase.

Based on the results of training phase (Figure 3), parameter sets within the first and third quantiles were selected for testing phase. For each training sample, 32 parameter sets were selected, and then applied to their respective testing samples. The parameter sets that were common across all samples within a single model are shown in Figure 4.

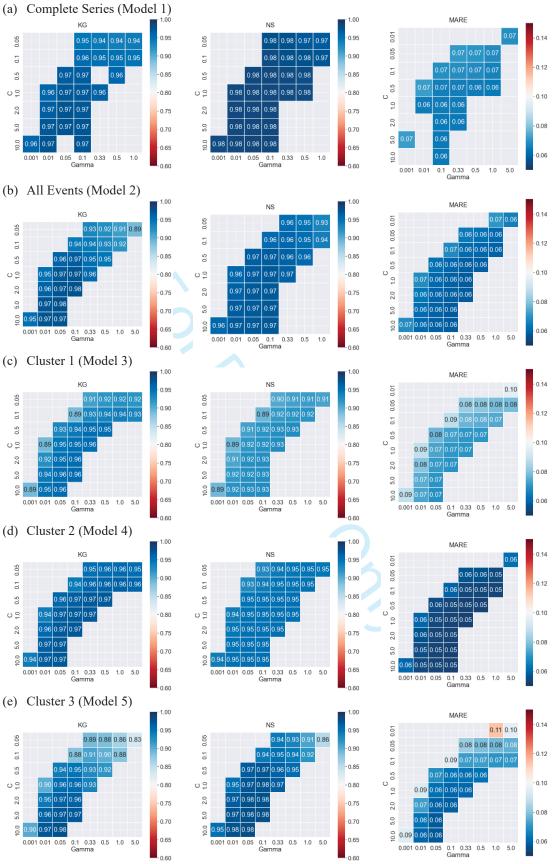


Figure 4. Average variation of KG, NS, and MARE coefficients for parameter sets of data series during the testing phase.

As expected, this procedure led to a reduction in the number of outliers. Despite the smaller number of combinations compared to the training phase, the test period exhibited greater variability in the results for all models. For Clusters 1 and 3 Models (Figures 4c and e), the variability between the corresponding parameter sets in both phases was even more pronounced, indicating that, for instance, Cluster 3 Model – dealing with extreme flow events – is more sensitive to the selection of events considered at each stage.

To illustrate susceptibility to the choice of events at each stage, Table 7 ranks the peak discharges of Cluster 3 events at União da Vitória for the period from 1964 to 2004. Table 8 shows the peak dates of the events included in each sample. Table 8 helps interpret the sample metrics by indicating which Cluster 3 events were used in each test phase sample. Highlighted colors in both tables represent the presence of events from Table 7 in the samples listed in Table 8.

Table 7. Ranking of Cluster 3 peak discharges at União da Vitória from 1964 to 2004.

Ord.	Peak date	Peak discharge (m³/s)	Ord.	Peak date	Peak discharge (m³/s)
1 st	07/18/1983	5156.7	9 th	11/27/1982	2263.0
$2^{\rm nd}$	06/08/1992	3953.6	10 th	01/27/1990	2196.0
3^{rd}	10/06/1993	2786.2	11 th	05/28/1987	2167.4
4 th	10/12/1998	2751.7	12 th	10/11/2001	1975.6
5 th	01/22/1995	2453.3	13 th	01/01/1981	1887.1
6^{th}	05/02/1998	2447.2	14 th	07/14/1999	1832.9
7^{th}	01/13/1971	2428.5	15 th	02/04/1997	1808.7
8 th	09/25/2000	228.0			

The highest observed peak (on July 18th, 1983) appeared in Samples 5 to 8, while the lowest peak (on February 04th, 1997) was present in Samples 1, 2, 3, and 8. In other words, both the maximum and minimum extreme values are found simultaneously only in Sample 8 during the test phase. This variability in the samples was also observed during the training phase.

471 Table 8. Sampling strategy for Cluster 3 Events during the test phase.

Cluster 3 peak date events											
Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Sample 7	Sample 8				
05/28/1987	01/27/1990	01/13/1971	01/13/1971	01/13/1971	01/13/1971	01/01/1981	01/01/1981				
01/27/1990	06/08/1992	01/22/1995	09/25/2000	01/01/1981	01/01/1981	11/27/1982	11/27/1982				
06/08/1992	10/06/1993	02/04/1997	10/11/2001	11/27/1982	11/27/1982	07/18/1983	07/18/1983				
10/06/1993	01/22/1995	05/02/1998		07/18/1983	07/18/1983	05/28/1987	05/28/1987				
01/22/1995	02/04/1997	10/12/1998			05/28/1987	01/27/1990	01/27/1990				
02/04/1997	05/02/1998	07/14/1999				06/08/1992	06/08/1992				
05/02/1998	10/12/1998	09/25/2000				10/06/1993	10/06/1993				
10/12/1998	07/14/1999	10/11/2001					01/22/1995				
07/14/1999	09/25/2000						02/04/1997				
09/25/2000	10/11/2001						05/02/1998				
10/11/2001							10/12/1998				

4.3 Adjusted models

The set of parameters with C = 0.5 and $\gamma = 0.33$ was centered in the heatmaps and appears in all models shown in Figure 4. For the reasons previously mentioned, these parameter values were adopted across all models.

Model evaluation measures for the period from 1964 to 2004 (training and test phases) are shown in Table 9. The results along the diagonals (in bold) indicate the outcomes when simulating the same series used to train the model. The results underlined represent the models with metrics lower than 0.90 for KG and NS, and higher than 0.10 for MARE.

As before, better results were expected when a series was estimated using its own model – trained with that specific category of series – compared to simulations conducted with models trained on different data classes (off-diagonal values).

Table 9. KG, NS, and MARE coefficients for final models' simulations from 1964 to 2004.

			S	imulated series	6	
		Complete	All Events	Cluster 1	Cluster 2	Cluster 3
	M1	0.975	0.984	0.949	0.971	0.982
	M2	0.954	0.978	0.934	0.975	0.983
KG	М3	<u>0.585</u>	<u>0.817</u>	0.960	0.930	0.882
	M4	0.885	0.949	0.958	0.972	0.962
	M5	0.948	0.977	0.933	0.973	0.980
	M1	0.985	0.974	0.923	0.952	0.982
	M2	0.983	0.973	0.919	0.951	0.982
S	М3	<u>0.766</u>	0.905	0.931	0.953	0.963
	M4	0.968	0.971	0.930	0.953	0.979
	M5	0.982	0.973	0.917	0.950	0.981
	M1	0.063	0.062	0.084	0.058	0.053
	M2	0.139	0.058	0.088	0.059	0.055
MARE	М3	0.896	0.205	0.069	0.067	<u>0.144</u>
M	M4	<u>0.275</u>	0.075	0.070	0.052	0.064
	M5	0.150	0.066	0.088	0.060	0.053

First, analyzing the results for each data series, Table 9 shows that: (i) the <u>Complete Series</u> was best simulated by Models 1 (its own model), 2, and 5; (ii) the <u>All Events Series</u> was well estimated by all models, though Model 3 showed poorer results; (iii) <u>Cluster 1 Series</u> were well represented by all models; (iv) <u>Cluster 2 Series</u> were also well represented by all models; and (v) <u>Cluster 3 Series</u> were well represented by all models except Model 3. Second, analyzing the results for each model, Table 9 shows that: (i) <u>Models 1, 2, and 5</u> provided good results for all series; (ii) <u>Model 3</u> provided good results only for Clusters 1 and 2; and (iii) <u>Model 4</u> provided good results for all series except the Complete Series. Models' behavior can also be statistically observed, as shown in Table 10, where simulated minimum discharges differed significantly from the observed ones, even in

series forecasted by their own models. Simulated maximum discharges remained close to the observed ones. For the other statistics, the main errors were identified in simulations performed by Models 3 and 4 for the Complete and All Events series.

Table 10. Statistical relationships between simulated series by final models and observed series, from 1964 to 2004.

	-	Simulated / observed series											
		Complete	All Events	Cluster 1	Cluster 2	Cluster 3			Complete	All Events	Cluster 1	Cluster 2	Cluster 3
	M1	1.19	0.90	0.85	1.05	1.07	on	M1	0.98	0.99	1.04	0.98	0.99
Е	M2	2.16	1.07	0.94	1.07	1.36	viati	M2	0.96	0.99	1.05	0.99	0.99
Minimum	M3	<u>8.26</u>	3.03	1.49	1.34	3.20	Standard - deviation	M3	0.80	<u>0.86</u>	0.99	0.93	0.89
Mir	M4	<u>3.52</u>	<u>1.60</u>	1.09	1.14	1.77	ndare	M4	0.91	0.95	1.02	0.99	0.97
	M5	2.31	1.25	0.98	1.10	1.35	Star	M5	0.95	0.98	1.05	0.99	0.99
	M1	1.00	1.00	0.95	0.94	1.00		M1	1.02	0.98	0.79	0.93	0.99
ш	M2	0.99	0.99	0.95	0.94	0.99	SS	M2	1.07	1.00	0.80	0.93	0.99
Maximum	M3	0.95	0.96	0.97	0.96	0.96	Skewness	M3	1.32	<u>1.20</u>	1.04	1.14	1.17
Ma	M4	0.99	0.99	0.96	0.96	0.99	Ske	M4	<u>1.19</u>	<u>1.11</u>	0.95	1.03	1.10
	M5	0.98	0.98	0.95	0.94	0.98		M5	1.08	1.01	0.79	0.95	0.99
	M1	0.98	1.00	1.00	0.99	0.99		M1	0.98	1.00	1.00	0.99	0.99
	M2	1.02	0.99	1.01	1.00	0.99	9	M2	1.02	0.99	1.01	1.00	0.99
Mean	M3	<u>1.36</u>	<u>1.12</u>	0.98	1.01	1.04	Volume	M3	<u>1.36</u>	1.12	0.98	1.01	1.04
2	M4	1.07	1.00	0.98	0.99	0.98	>	M4	1.07	1.00	0.98	0.99	0.98
	M5	1.02	1.00	1.01	1.00	0.99		M5	1.02	1.00	1.01	1.00	0.99

Thus, the estimations for Complete and All Events Series using all five models were compared. Model 3 performed poorly in predicting both series, such that:

1. KG values were reduced due to the greater distances between the simulated mean discharges and standard deviations compared to the respective observed parameters. This is because the KG coefficient is calculated based on the mean discharges, variability (standard deviation), and dynamics (correlation).

- 2. NS values were lower due to the difficulty in predicting mean discharges, as the coefficient prioritizes maximum flows over mean flows. This behavior is also reflected in the volume ratio, which is greater than 1.0, indicating an overestimation of the peaks, especially the lower ones.
- MARE values increased due to the greater differences between simulated minimum and mean flows compared to the observed parameters, as MARE prioritizes low and mean flows.

4.4 Discharge prediction

All five data samples were predicted by the five models for the 2005 to 2015 period with a one-day lead time. The measures used to evaluate the forecasts of the series are shown in Table 11.

Table 11. KG, NS, and MARE coefficients for series predictions from 2005 to 2015.

			P	redicted series		
		Complete	All Events	Cluster 1	Cluster 2	Cluster 3
	M1	0.973	0.976	0.974	0.968	0.967
	M2	0.949	0.936	0.970	0.970	0.964
KG	М3	<u>0.619</u>	0.820	0.958	0.925	0.875
	M4	0.892	0.949	0.972	0.974	0.960
	M5	0.944	0.967	0.969	0.967	0.967
	M1	0.984	0.968	0.950	0.948	0.964
	M2	0.982	0.967	0.949	0.948	0.964
SZ	М3	0.794	0.903	0.955	0.951	0.942
	M4	0.968	0.964	0.954	0.951	0.958
	M5	0.981	0.968	0.948	0.948	0.963
	M1	0.068	0.068	0.082	0.067	0.058
	M2	0.141	0.064	0.084	0.068	0.060
MARE	М3	0.873	0.201	0.072	0.074	0.113
Z	M4	0.275	0.082	0.071	0.062	0.072
	M5	<u>0.156</u>	0.073	0.085	0.069	0.057

In Table 11, the main diagonals – highlighted in bold – indicate the results obtained when forecasting the same series used to train the model. The underlined results represent models with metrics lower than 0.90 for KG and NS, and higher than 0.10 for MARE. The same combinations of series and models highlighted in Table 9 are also prominent in Table 11; the forecast measures performed similarly to the simulations in previous stage. As observed in final models, the performance of the forecasts for the main diagonals (Table 11) was satisfactory when compared within their matching columns. Simulating a series with a model other than its own was expected to perform worse than using the model that was trained with that specific series. Once again, Models 3 and 4 stand out for having the worst performances when forecasting Complete and All Events Series. Models' behavior can also be statistically observed (Table 12). A small variation between the results of the final validation and the forecast was noted, indicating that the observed differences in statistical relationships between the two phases were minimal. In other words, the simulated statistics well represented the observed ones, except for the minimum discharges. Greater variability was observed in mean discharges, standard deviations, skewness, and volumes for simulations made by Model 3, particularly for the Complete and All Events Series. Thus, as observed in previous stage, the estimations for Complete and All Events Series using all five models were exclusively compared and highlighted in Table 12. Model 3 performed poorly in predicting Complete and All Events Series, such that: (i) the relationship between means and standard deviations explains the behavior of KG; (ii) the relationship between volumes and means accounts for NS values; and (iii) the relationship between minimum and mean discharges suggests changes in MARE. Therefore,

improvements in statistical relationships also indicate improvements in quality metrics.

Table 12. Statistical relationships between predicted and observed series, from 2005 to 2015.

	-	Predicted / observed series											
	-	Complete	All Events	Cluster 1	Cluster 2	Cluster 3			Complete	All Events	Cluster 1	Cluster 2	Cluster 3
	M1	1.24	1.22	1.16	1.11	1.04		M1	0.98	0.98	1.01	0.98	0.98
_	M2	2.23	1.43	1.28	1.13	1.21	iation	M2	0.95	0.98	1.02	0.99	0.97
Minimum	М3	<u>8.62</u>	3.73	1.91	1.46	2.36	Standard - deviation	М3	<u>0.80</u>	0.85	0.97	0.93	0.88
Mir	M4	<u>3.64</u>	2.04	1.45	1.25	1.45	ndard	M4	0.91	0.95	1.01	1.00	0.97
	M5	2.44	1.60	1.29	1.18	1.19	Sta	M5	0.95	0.97	1.02	0.98	0.98
-	M1	0.96 0.97 1.00 0.95 0.96		M1	1.00	0.95	0.74	0.92	0.93				
_	M2	0.96	0.96	1.00	0.95	0.96		M2	1.05	0.98	0.72	0.91	0.98
Maximum	М3	0.99	0.99	0.97	0.97	0.99	Skewness	М3	1.27	1.27	1.00	1.09	1.46
May	M4	1.01	1.02	0.99	0.98	1.01	Ske	M4	<u>1.17</u>	1.17	0.95	1.00	1.31
	M5	0.97	0.97	0.99	0.95	0.97		M5	1.06	1.01	0.71	0.92	1.02
	M1	0.99	1.00	1.00	1.00	0.99		M1	0.98	1.00	1.00	1.00	0.99
	M2	1.02	0.98	1.01	1.00	0.99		M2	1.02	0.98	1.01	1.00	0.99
Mean	M3	<u>1.32</u>	<u>1.10</u>	0.98	1.01	1.04	Volume	М3	1.32	<u>1.10</u>	0.98	1.01	1.04
Z	M4	1.06	1.00	0.98	0.99	0.98	Λ	M4	1.06	1.00	0.98	0.99	0.98
	M5	1.02	1.00	1.01	1.01	0.98		M5	1.02	1.00	1.01	1.01	0.98

Based on the previous results, the hydrographs obtained from forecasts for the Complete and All Events Series – simulated by their own models and by the models per group – are shown in Figures 5 and 6. For the Complete Series (Figure 5), as observed for the 1964 to 2004 period, Model 5's forecasts were very close to those of Model 1, and consequently, to the observed flows – all three hydrographs overlapped. In contrast, Models 3 and 4, while not diverging significantly from peak discharges, displayed a notable offset from low flows.

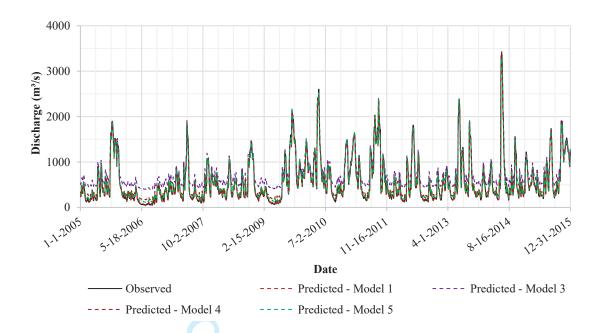


Figure 5. Predicted and observed hydrographs from 2005 to 2015 – Complete Series.

For the All Events Series (Figure 6), Model 5's forecasts closely matched Model 2's for both peak and low flows, with both hydrographs overlapping the observed data. However, Model 3 overestimated smaller peaks and underestimated larger ones. Compared to the final validation stage, the All Events Series predicted by Model 3 showed the greatest behavioral difference, along with a slight decline in performance metrics.

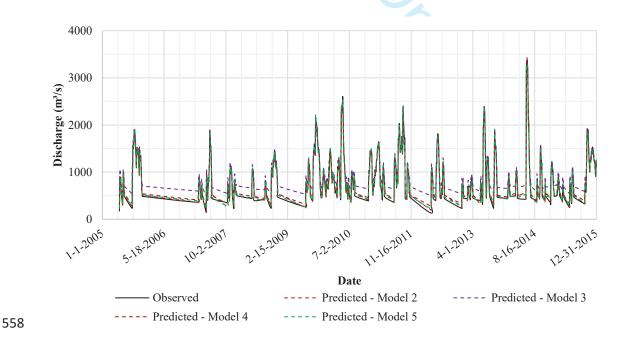


Figure 6. Predicted and observed hydrographs from 2005 to 2015 – All Events Series.

5 Conclusions

This paper proposes a new hydrological modeling approach for discharge prediction, based on previous flood events characterization through flood events clustering. The proposed hydrological modeling approach is illustrated with a real-world application.

Daily mean discharges from 1964 to 2015 in União da Vitória (Iguaçu River basin, Paraná State, Brazil) were used, and the SVR algorithm was applied to create AI models for 1-day discharge forecasts. The results obtained are a function of the clustering itself, the chosen hydrological model, and the adopted forecast horizon. Once again, we highlight that the proposed approach allows using different hydrological models.

Due to the prior application of the FCM algorithm, the results revealed that all events share similarities with each cluster to varying degrees. This overlap may negatively impact the performance of the cluster models.

Therefore, clustering effectively helps identify similar patterns and characteristics within event groups. It reduces averages, standard deviations, and skewness for the first clusters while increasing these parameters for Cluster 3, due to its events' greater magnitude. The centroid coordinates further highlight the larger scale of events in the last group.

Training, testing, and prediction periods were subdivided to distribute major events across these periods. The metric results, facing the sampling strategy, revealed that the presence or absence of certain events in the training or testing phases often led to significant differences between samples. During the testing stage, due to the smaller number of parameter sets, some samples showed less variability in their results, except for Models 1, 2, and 5, particularly in Samples 5 to 8.

As expected, both in the definition of final models and during forecasting stage, models performed better when simulating series from the same category used for calibration. Model 3 performed worse with Complete and All Events Series, which is justified by the disparity between the mean discharges of simulated and observed flows. This disparity impacted all three metrics, with the standard deviation having a greater influence on KG. However, these results may have been influenced by the events within each group generated during clustering, a limitation of the research. To reduce these impacts, one option is to investigate the final fuzzy partition matrix from FCM to weight events with low membership degrees, increasing the sharing information with other clusters. In the same context, another approach could be to eliminate these events from the analysis, keeping only those events that predominantly look like a single cluster. However, the latter consideration could significantly reduce the number of events available for defining the calibration and forecasting models by group, especially for Cluster 3. Other research limitations include the process of defining the start and end of events, which is not well established in the literature. In this work, it was done empirically based on the local characteristics of the União da Vitória gauging station. This issue directly influences the determination of flood indicators, which could lead to changes in clustering and, consequently, in the modeling process. Additionally, incorporating rainfall-related indicators could result in similar changes. Topics related to the characteristics of the Iguaçu River can also be discussed as limitations, including: (i) the topographical conditions of the Iguaçu River Basin, particularly between Fluviópolis and União da Vitória, where the river flows through a

main channel constrained by low and flat banks, that are more prone to overflow during

high runoff volumes (SEC-CORPRERI, 1999); and (ii) the proximity of both gauging stations, which necessitates more frequent discharge data than daily measurements.

On the other hand, all research limitations mentioned here can be seen as opportunities for further development. The approach of combining clustering with hydrological modeling can be applied to both traditional and data-driven models.

In the context of flow data scale, we suggest applying the methodological processes presented here to a study area with compatible data availability. This means that the larger the river basins under analysis, the smaller the data interval should be.

Based on flood indicators and a certain number of clusters as suggested in this research, a forecasting system based on groups of events could perform by reducing the forecasting possibilities, such that: (i) at the initial instant (t₀), a flood event is identified; (ii) at the next time point (t₁), the peak time and flow of the ongoing event are determined, and forecasts for the following time instant are made using the three models per group; (iii) as the event progresses (t₂, t₃, and the next ones), peak time and flow are updated when there is an overflow, and the forecasts tend to be closer to a particular group, restricting the forecasting possibilities.

Therefore, studying the process of forecasting by groups in real-time is necessary and represents a very intriguing challenge, so that, once the start of an event has been identified, the observed flows during the flood development can be used to identify the similarities of the events with groups, restricting the characteristics of the flood in progress. One possibility is using the models by group integrated with a fuzzy analysis to forecast the discharges.

6 References

- 628 ADNAN, R. M.; LIANG, Z.; TRAJKOVIC, S.; ZOUNEMAT-KERMANI, M.; LI, B.;
- 629 KISI, O. (2019). Daily streamflow prediction using optimally pruned extreme learning
- 630 machine. *Journal of Hydrology*, 577, 1-11. DOI: <u>10.1016/j.jhydrol.2019.123981</u>.
- 631 ALTHOFF, D.; RODRIGUES, L. N.; BAZAME, H. C. (2021). Uncertainty
- quantification for hydrological models based on neural networks: the dropout ensemble.
- 633 Stochastic Environmental Research and Risk Assessment, 35, 1051-1067. DOI:
- 634 10.1007/s00477-021-01980-8.
- 635 ASKARI, S. (2021). Fuzzy C-Means clustering algorithm for data with unequal cluster
- 636 sizes and contaminated with noise and outliers: Review and development. Expert Systems
- *With Applications*, 165, 1-27. DOI: <u>10.1016/j.eswa.2020.113856</u>.
- BAI, P.; LIU, X.; XIE, J. (2021). Simulating runoff under changing climatic conditions:
- A comparison of the long short-term memory network with two conceptual hydrologic
- 640 models. *Journal of Hydrology*, 592, 1-11, DOI: <u>10.1016/j.jhydrol.2020.125779</u>.
- BRÊDA, J. P. L. F.; PAIVA, R. C. D.; PEDROLLO, O. C.; PASSAIA, O. A.;
- 642 COLLISCHONN, W. (2021). Modeling coordinated operation of multiple hydropower
- reservoirs at a continental scale using artificial neural network: the case of Brazilian
- 644 hydropower system. Revista Brasileira de Recursos Hídricos, 26(12), 1-12. DOI:
- 645 10.1590/2318-0331.262120210011.
- 646 CHANG, C.-C.; LIN, C.-J. (2001). Training v-support vector classifiers: Theory and
- algorithms. Neural Computation, 13(9), 2119–2147.
- 648 CHANG, C.-C.; LIN, C.-J. (2022). LIBSVM: a library for Support Vector Machines.
- 649 Available at: https://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf.
- 650 CHEN, S.; REN, M. SUN, W. (2021). Combining two-stage decomposition based
- 651 machine learning methods for annual runoff forecasting. Journal of Hydrology, 603, 1-
- 652 18. DOI: 10.1016/j.jhydrol.2021.126945.
- DESAI, S.; OUARDA, T. B. M. J. (2021). Regional hydrological frequency analysis at
- 654 ungauged sites with random forest regression. *Journal of Hydrology*, 594, 1-8. DOI:
- 655 10.1016/j.jhydrol.2020.125861.
- DIFI, S.; ELMEDDAHI, Y.; HEBAL, A.; SINGH, V. P.; HEDDAM, S.; KIM, S.; KISI,
- O. (2023). Monthly streamflow prediction using hybrid extreme learning machine

- optimized by bat algorithm: a case study of Cheliff watershed, Algeria. Hydrological
- 659 Sciences Journal, 68(2), 189-208. DOI: 10.1080/02626667.2022.2149334.
- DING, Y.; ZHU, Y.; FENG, J.; ZHANG, P.; CHENG, Z. (2020). Interpretable spatio-
- temporal attention LSTM model for flood forecasting. *Neurocomputing*, 403, 348-359.
- 662 DOI: <u>10.1016/j.neucom.2020.04.110</u>.
- 663 EBTEHAJ, I.; BONAKDARI, H. (2022). A reliable hybrid outlier robust non-tuned rapid
- machine learning model for multi-step ahead flood forecasting in Quebec, Canada.
- *Journal of Hydrology*, 614, 1-21. DOI: <u>10.1016/j.jhydrol.2022.128592</u>.
- 666 EZUGWU, A. E.; IKOTUN, A. M.; OYELADE, O. O.; ABUALIGAH, L.;
- AGUSHAKA, J. O.; EKE, C. I.; AKINYELU, A. A. (2022). A comprehensive survey of
- 668 clustering algorithms: State-of-the-art machine learning applications, taxonomy,
- 669 challenges, and future research prospects. Engineering Applications of Artificial
- *Intelligence*, 110, DOI: <u>10.1016/j.engappai.2022.104743</u>.
- 671 FATHIAN, F.; MEHDIZADEH, S.; SALES, A. K.; SAFARI, M. J. S. (2019). Hybrid
- models to improve the monthly river flow prediction: Integrating artificial intelligence
- and non-linear time series models, and future research prospects. *Journal of Hydrology*,
- 674 575, 1200–1213. DOI: 10.1016/j.jhydrol.2019.06.025.
- 675 FENG, Z.; SHI, P.; YANG, T.; NIU, W.; ZHOU, J.; CHENG, C. (2022). Parallel
- 676 cooperation search algorithm and artificial intelligence method for streamflow time series
- 677 forecasting. *Journal of Hydrology*, 606, 1–14. DOI: <u>10.1016/j.jhydrol.2022.127434</u>.
- 678 FREITAS, C. (2016). Uso de técnicas de classificação de dados na operação de sistemas
- 679 de reservatórios de usinas hidrelétricas durante cheias (Tese de Doutorado), Programa
- de Pós-Graduação em Engenharia de Recursos Hídricos e Ambiental, Universidade
- 681 Federal do Paraná, Curitiba.
- 682 GUPTA, H. V.; KLING, H.; YILMAZ, K. K.; MARTINEZ, G. F. (2009). Decomposition
- of the mean squared error and NSE performance criteria: Implications for improving
- 684 hydrological modelling. *Journal of Hydrology*, 377(1-2), 80-91. DOI:
- 685 10.1016/j.jhydrol.2009.08.003.
- 686 HALLOUIN, T.; BRUEN, M.; O'LOUGHLIN, F. E. (2020). Calibration of hydrological
- 687 models for ecologically relevant streamflow predictions: a trade-off between fitting well

- 688 to data and estimating consistent parameter sets?. Hydrology and Earth System Science,
- 689 24, 1031-1054. DOI: 10.5194/hess-24-1031-2020.
- 690 IBRAHIM, K. S. M. H.; HUANG, Y. F.; AHMED, A. N.; KOO, C. H.; EL-SHAFIE, A.
- 691 (2021). A review of the hybrid artificial intelligence and optimization modelling of
- 692 hydrological streamflow forecasting. *Alexandria Engineering Journal*, 61(1), 279-303.
- 693 DOI: <u>10.1016/j.aej.2021.04.100</u>.
- 694 ISLAM, A. R. M. T.; TALUKDAR, S.; MAHATO, S.; KUNDU, S.; EIBEK, K. U.;
- 695 PHAM, Q. B.; KURIQI, A.; LINH, N. T. T. (2021). Flood susceptibility modelling using
- advanced ensemble machine learning models. Geoscience Frontiers, 12, 1-18. DOI:
- 697 <u>10.1016/j.gsf.2020.09.006</u>.
- 698 JOO, H.; LEE, M.; KIM, J.; JUNG, J.; KWAK, J.; KIM, H. S. (2021). Stream gauge
- 699 network grouping analysis using community detection. Stochastic Environmental
- *Research and Risk Assessment*, 35, 781-795. DOI: <u>10.1007/s00477-020-01916-8</u>.
- 701 KHOSRAVI, K.; GOLKARIAN, A.; BOOIJ, M. J.; BARZEGAR, R.; SUN, W.;
- YASEEN, Z. M.; MOSAVI, A. (2021). Improving daily stochastic streamflow prediction:
- 703 comparison of novel hybrid data-mining algorithms. Hydrological Sciences Journal,
- 704 66(9), 1457-1474. DOI: <u>10.1080/02626667.2021.1928673</u>.
- 705 KIM, D.; LEE, J.; KIM, J.; LEE, M.; WANG, W.; KIM, H. S. (2022). Comparative
- analysis of long short-term memory and storage function model for flood water level
- forecasting of Bokha stream in NamHan River, Korea. *Journal of Hydrology*, 606, 1-13.
- 708 DOI: 10.1016/j.jhydrol.2021.127415.
- 709 KURIAN, C.; SUDHEER, K. P.; VEMA, V. K.; SAHOO, D. (2020). Effective flood
- 710 forecasting at higher lead times through hybrid modelling framework. *Journal of*
- *Hydrology*, 587, 1-13. DOI: <u>10.1016/j.jhydrol.2020.124945</u>.
- 712 LAPPICY, T.; LIMA, C. H. R. (2023). Enhancing monthly streamflow forecasting for
- 713 Brazilian hydropower plants through climate index integration with stochastic methods.
- 714 Revista Brasileira de Recursos Hídricos, 28(48), 1-7. DOI: 10.1590/2318-
- 715 0331.282320230118.
- 716 LAWIN, A. E.; HOUNGUE, R.; PO, Y.; HOUNGUE, N. R.; ATTOGOUINON, A.;
- 717 AFOUDA, A. A. (2019). Mid-Century Climate Change Impacts on Ouémé River

- 718 Discharge at Bonou Outlet (Benin). *Hydrology*, 6(72), 1-20. DOI:
- 719 10.3390/hydrology6030072.
- 720 LETESSIER, C.; CARDI, J.; DUSSEL, A.; EBTEHAJ, I.; BONAKDARI, H. (2023).
- 721 Enhancing Flood Prediction Accuracy through Integration of Meteorological Parameters
- in River Flow Observations: A Case Study Ottawa River. *Hydrology*, 10(164), 1-23. DOI:
- 723 <u>10.3390/hydrology10080164</u>.
- LI, X.; SHA, J.; WANG, Z-L. (2019). Comparison of daily streamflow forecasts using
- 725 extreme learning machines and the random forest method. Hydrological Sciences
- 726 Journal, 64(15), 1857-1866. DOI: 10.1080/02626667.2019.1680846.
- 727 LIANG, Z.; XIAO, Z.; WANG, J.; SUN, L.; LI, B.; HU, Y.; WU, Y. (2019). An improved
- 728 chaos similarity model for hydrological forecasting. *Journal of Hydrology*, 577. DOI:
- 729 <u>10.1016/j.jhydrol.2019.123953</u>.
- T30 LIMA, G. R. T.; SCOFIELD, G. B. (2021). Feasibility study on operational use of neural
- 731 networks in a flash flood early warning system. Revista Brasileira de Recursos Hídricos,
- 732 26(7), 2-11. DOI: <u>10.1590/2318-0331.262120200152</u>.
- T33 LUPPICHINI, M.; FAVALLI, M.; ISOLA, I.; NANNIPIERI, L.; GIANNECCHINI, R.;
- 734 BINI, M. (2019). Influence of Topographic Resolution and Accuracy on Hydraulic
- 735 Channel Flow Simulations: Case Study of the Versilia River (Italy). Remote Sensing,
- 736 11(1630), 1-22. DOI: 10.3390/rs11131630.
- 737 LUPPICHINI, M.; BARSANTI, M.; GIANNECCHINI, R.; BINI, M. (2022). Deep
- 738 learning models to predict flood events in fast-flowing watersheds. Science of the Total
- 739 Environment, 813, 1-10. DOI: 10.1016/j.scitotenv.2021.151885.
- 740 MAHDAVI-MEYMAND, A.; SULISZ, W.; ZOUNEMAT-KERMANI, M. (2023).
- 741 Hybrid and integrative evolutionary Machine Learning in hydrology: a systematic review
- and meta-analysis. Archives of Computational Methods in Engineering. DOI:
- 743 10.1007/s11831-023-10017-y.
- MINE, M. R. M.; TUCCI, C. E. M. (2002). Gerenciamento da produção de energia e
- 745 controle de inundação: Foz de Areia no rio Iguaçu. Revista Brasileira de Recursos
- *Hidricos*, 7(3), 85-107. DOI: <u>10.21168/rbrh.v7n3.p85-107</u>.

- 747 MOHAMMADI, B.; LINH, N. T. T.; PHAM, Q. B.; AHMED, A. N.; VOJTEKOVÁ, J.;
- 748 GUAN, Y.; ABBA, S. I.; EL-SHAFIE, A. (2020). Adaptive neuro-fuzzy inference system
- 749 coupled with shuffled frog leaping algorithm for predicting river streamflow time series.
- 750 Hydrological Sciences Journal, 65(10), 1738-1751. DOI:
- 751 <u>10.1080/02626667.2020.1758703</u>.
- 752 MORETTIN, P. A.; SINGER, J. M. (2023). Estatística e ciência de dados. Rio de Janeiro:
- 753 LTC.
- MOSAVI, A.; GOLSHAN, M.; CHOUBIN, B.; ZIEGLER, A. D.; SIGAROODI, S. K.;
- 755 ZHANG, F.; DINEVA, A. A. (2021). Fuzzy clustering and distributed model for
- streamflow estimation in ungauged watersheds. Scientific Reports, 11(8243). DOI:
- 757 <u>10.1038/s41598-021-87691-0</u>.
- MOURA, C. N.; JAN, S.; DETZEL, D. H. M. (2022). Evaluating the long short-term
- 759 memory (LSTM) network for discharge prediction under changing climate conditions.
- *Hydrology Research*, 53(5), 657-667. DOI: <u>10.2166/nh.2022.044</u>.
- National Water and Sanitation Agency ANA. (2023). Portal HidroWeb. Brasília: ANA.
- Retrieved in 2023, from: https://www.snirh.gov.br/hidroweb/serieshistoricas.
- NIU, W.; FENG, Z. (2021). Evaluating the performances of several artificial intelligence
- methods in forecasting daily streamflow time series for sustainable water resources
- 765 management. Sustainable Cities and Society, 64, 1-12. DOI: 10.1016/j.scs.2020.102562.
- 766 RIBEIRO, V. H. A.; REYNOSO-MEZA, G.; SIQUEIRA, H. V. (2020). Multi-objective
- ensembles of echo state networks and extreme learning machines for streamflow series
- 768 forecasting. Engineering Applications of Artificial Intelligence, 95, 1-19. DOI:
- 769 10.1016/j.engappai.2020.103910.
- 770 ROCHA, P. S. M. (2012). Gestão em áreas de risco de enchentes: estudo de caso para
- *União da Vitória Paraná* (Dissertação de Mestrado) Universidade Positivo, Curitiba.
- ROY, A.; CHAKRABORTY, S. (2023). Support Vector Machine in structural reliability
- analysis: A review. Reliability Engineering and System Safety, 233. DOI:
- 774 10.1016/j.ress.2023.109126.

- SAADI, M.; OUDIN, L.; RIBSTEIN, P. (2019). Random Forest ability in regionalizing
- 776 hourly hydrological model parameters. Water, 11(1540), 1-22. DOI:
- 777 <u>10.3390/w11081540</u>.
- 778 SAMANTARAY, S.; SAHOO, P.; SAHOO, A.; SATAPATHY, D. P. (2023). Flood
- discharge prediction using improved ANFIS model combined with hybrid particle swarm
- optimization and slime mould algorithm. Environmental Science and Pollution Research,
- 781 30, 83845-83872. DOI: <u>10.1007/s11356-023-27844-y</u>.
- 782 SCHOLKOPF, B.; SMOLA, A. J.; WILLIAMSON, R. C.; BARTLETT, P. L. (2000).
- New Support Vector Algorithms. *Neural Computation*, 12, 1207–1245.
- 784 SCHOPPA, L.; DISSE, M.; BACHMAIR, S. (2020). Evaluating the performance of
- random forest for large-scale flood discharge simulation. Journal of Hydrology, 590, 1-
- 786 13. DOI: 10.1016/j.jhydrol.2020.125531.
- 787 Sociedade de Estudos Contemporâneos Comissão Regional Permanente de Prevenção
- 788 Contra Enchentes do Rio Iguaçu SEC-CORPRERI. (1999). Conhecendo e Convivendo
- 789 com Enchentes.
- 790 SHARMA, R. K.; KUMAR, S.; PADMALAL, D.; ROY, A. (2023). Streamflow
- 791 prediction using machine learning models in selected rivers of Southern India.
- 792 International Journal of River Basin Management. DOI:
- 793 10.1080/15715124.2023.2196635.
- 794 SHUKLA, R.; KUMAR, P.; VISHWAKARMA, D. K.; ALI, R.; KUMAR, R.; KURIQI,
- 795 A. (2022). Modeling of stage-discharge using back propagation ANN-, ANFIS-, and
- 796 WANN-based computing techniques. Theoretical and Applied Climatology, 147, 867-
- 797 889. DOI: 10.1007/s00704-021-03863-y.
- 798 SINGH, V. P. (2018). Hydrologic modeling: progress and future directions. *Geoscience*
- *Letters*, 5(15), 1-18. DOI: 10.1186/s40562-018-0113-z.
- 800 SNIEDER, E.; SHAKIR, R.; KHAN, U. T. (2020). A comprehensive comparison of four
- input variable selection methods for artificial neural network flow forecasting models.
- *Journal of Hydrology*, 583. DOI: 10.1016/j.jhydrol.2019.124299.
- 803 STAUDINGER, M.; STAHL, K.; SEIBERT, J.; CLARK, M. P.; TALLAKSEN, L. M.
- 804 (2011). Comparison of hydrological model structures based on recession and low flow

- simulations. Hydrology and Earth System Sciences, 15(11), 3447-3459. DOI:
- 806 10.5194/hess-15-3447-2011.
- STEFFEN, P. C.; GOMES, J. (2018). Clustering of historical floods observed on Iguaçu
- 808 River, in União da Vitória, Paraná. Revista Brasileira de Recursos Hídricos, 23(38), 1-
- 809 12. DOI: <u>10.1590/2318-0331.231820170107</u>.
- 810 TARASOVA, L.; MERZ, R.; KISS, A.; BASSO, S.; BLOSCHL, G.; MERZ, B.;
- VIGLIONE, A.; PLOTNER, S.; GUSE, B.; SCHUMANN, A.; FISCHER, S.; AHRENS,
- 812 B.; ANWAR, F.; BARDOSSY, A.; BUHLER, P.; HABERLANDT, U.; KREIBICH, H.;
- 813 KRUG, A.; LUN, D.; MULLER-THOMY, H.; PIDOTO, R.; PRIMO, C.; SEIDEL, J.;
- VOROGUSHYN, S.; WIETZKE, L. (2019). Causative classification of river flood
- events. Water, 6(4), 1-23. DOI: <u>10.1002/wat2.1353</u>.
- TOZZI, B. K. M.; FILL, H. D. O. A. (2020). Verification of the stationarity of flow series
- in the Iguaçu River basin. Revista Brasileira de Recursos Hídricos, 25(10), 1-9. DOI:
- 818 <u>10.1590/2318-0331.252020180171.</u>
- UNDUCHE, F.; HABTAMU, T.; SENBETA, D.; ZHU, E. (2018). Evaluation of four
- 820 hydrological models for operational flood forecasting in a Canadian Prairie watershed.
- 821 Hydrological Sciences Journal, 63(8), 1133-1149, DOI:
- 822 10.1080/02626667.2018.1474219.
- WANG, X.; WANG, Y.; YUAN, P.; WANG, L.; CHENG, D. (2021). An adaptive daily
- runoff forecast model using VMD-LSTM-PSO hybrid approach. *Hydrological Sciences*
- *Journal*, 66(9), 1488-1502. DOI: 10.1080/02626667.2021.1937631.
- 826 XU, T.; LIANG, F. (2021). Machine learning for hydrologic sciences: an introductory
- 827 overview. Wires Water, 8(1533), 1-29. DOI: 10.1002/wat2.1533.
- YASEEN, Z. M.; FARIS, H.; AL-ANSARI, N. (2020). Hybridized Extreme Learning
- Machine Model with Salp Swarm Algorithm: a novel predictive model for hydrological
- application. *Complexity*, p. 1-14, 2020. DOI: 10.1155/2020/8206245.
- ZAKHROUF, M.; HAMID, B.; KIM, S.; MADANI, S. (2023). Novel insights for
- 832 streamflow forecasting based on deep learning models combined the evolutionary
- 833 optimization algorithm. *Physical Geography*, 44(1), 31-54. DOI:
- 834 10.1080/02723646.2021.1943126.

ZHANG, Z.; ZHANG, Q.; SINGH, V. P. (2018). Univariate streamflow forecasting using commonly used data-driven models: literature review and case study. *Hydrological Sciences Journal*, 63(7), 1091-1111. DOI: 10.1080/02626667.2018.1469756.

