



UNIVERSIDADE FEDERAL DE CAMPINA GRANDE  
CENTRO DE TECNOLOGIA E RECURSOS NATURAIS  
UNIDADE ACADÊMICA DE ENGENHARIA CIVIL  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA CIVIL E AMBIENTAL

**CÍCERO FELLIPE DINIZ DE SANTANA**

**ÁRVORES DE DECISÃO E SIMULAÇÃO ESTOCÁSTICA PARA ANÁLISE DE  
RISCOS EM SISTEMAS DE RESERVATÓRIOS SUPERFICIAIS**

Campina Grande - PB

2022

**CÍCERO FELLIPE DINIZ DE SANTANA**

**ÁRVORES DE DECISÃO E SIMULAÇÃO ESTOCÁSTICA PARA ANÁLISE DE  
RISCOS EM SISTEMAS DE RESERVATÓRIOS SUPERFICIAIS**

Tese apresentada ao Programa de Pós-Graduação em Engenharia Civil e Ambiental, como parte de requisito à obtenção do grau de Doutor em Engenharia Civil e Ambiental na área de concentração de Recursos Hídricos.

Orientador: Dr. Wilson Fadlo Curi  
Coorientador: Dr. Valterlin da Silva Santos

Campina Grande - PB

2022

S232a Santana, Cícero Fellipe Diniz de.  
Árvores de decisão e simulação estocástica para análise de riscos em sistemas de reservatórios superficiais / Cícero Fellipe Diniz de Santana. – Campina Grande, 2022.  
163 f.: il. color.

Tese (Doutorado em Engenharia Civil e Ambiental) – Universidade Federal de Campina Grande, Centro de Tecnologia e Recursos Naturais, 2022.  
"Orientação: Prof. Dr. Wilson Fadlo Curi; Coorientação: Prof. Dr. Valterlin da Silva Santos".  
Referências.

1. Tomada de Decisão. 2. Confiabilidade. 3. Bacias Hidrográficas. 4. Aprendizado de Máquina. 5. Análise de Risco. 6. Árvores de Decisão. 7. Planejamento de Recursos Hídricos.  
I. Curi, Wilson Fadlo. II. Santos, Valterlin da Silva. III. Título.

CDU 005.53(043)

**CÍCERO FELLIPE DINIZ DE SANTANA**

**ÁRVORES DE DECISÃO E SIMULAÇÃO ESTOCÁSTICA PARA ANÁLISE DE  
RISCOS EM SISTEMAS DE RESERVATÓRIOS SUPERFICIAIS**

Data da defesa: 04 de agosto de 2022

**Banca Examinadora**

---

Orientador: Prof. Dr. Wilson Fadlo Curi  
(Universidade Federal de Campina Grande)

---

Co-orientador: Prof. Dr. Valterlin da Silva Santos  
(Universidade Federal de Campina Grande)

---

Prof<sup>a</sup>. Dr<sup>a</sup>. Andréa Carla Lima Rodrigues  
(Universidade Federal de Campina Grande)

---

Prof. Dr. Camilo Allyson Simões de Farias  
(Universidade Federal de Campina Grande)

---

Prof<sup>a</sup>. Dr<sup>a</sup>. Dayse Luna Barbosa  
(Universidade Federal de Campina Grande)

---

Prof. Dr. Alcigeimes Batista Celeste  
(Universidade Federal de Sergipe)



MINISTÉRIO DA EDUCAÇÃO  
**UNIVERSIDADE FEDERAL DE CAMPINA GRANDE**  
 POS-GRADUACAO ENGENHARIA CIVIL AMBIENTAL  
 Rua Aprigio Veloso, 882, - Bairro Universitario, Campina Grande/PB, CEP 58429-900

### REGISTRO DE PRESENÇA E ASSINATURAS

1. **ATA DA DEFESA PARA CONCESSÃO DO GRAU DE DOUTOR EM ENGENHARIA CIVIL E AMBIENTAL**

2. **ALUNO(A): CÍCERO FELLIPE DINIZ DE SANTANA / COMISSÃO EXAMINADORA: DR. WILSON FADLO CURI – PPGECA/UFMG (PRESIDENTE) - ORIENTADOR, DR. VALTERLIN DA SILVA SANTOS – UAEC/UFMG – COORIENTADOR, DR.ª ANDRÉA CARLA LIMA RODRIGUES – PPGECA/UFMG – EXAMINADORA INTERNA, DR.ª DAYSE LUNA BARBOSA – UAEC/UFMG – EXAMINADORA INTERNA, DR. CAMILO ALLYSON SIMÕES DE FARIAS - UACTA/UFMG - EXAMINADOR EXTERNO, DR. ALCIGEIMES BATISTA CELESTE - UFS - EXAMINADOR EXTERNO (PORTARIA 25/2022). / TITULO DA TESE: “MODELO DE ANÁLISE DE RISCOS UTILIZANDO ÁRVORES DE DECISÃO E SIMULAÇÃO ESTOCÁSTICA APLICADA EM SISTEMAS DE RESERVATÓRIOS SUPERFICIAIS” / ÁREA DE CONCENTRAÇÃO: RECURSOS HÍDRICOS E SANEAMENTO AMBIENTAL / HORA DE INICIO: 16:00 HORAS / FORMA DA SESSÃO: POR VÍDEO CONFERÊNCIA.**

3. **EM SESSÃO REALIZADA POR VÍDEO CONFERÊNCIA, APÓS EXPOSIÇÃO DE CERCA DE 40 MINUTOS, O(A) ALUNO(A) FOI ARGUIDO(A) ORALMENTE PELOS MEMBROS DA COMISSÃO EXAMINADORA, TENDO DEMONSTRADO SUFICIÊNCIA DE CONHECIMENTO E CAPACIDADE DE SISTEMATIZAÇÃO NO TEMA DE SUA TESE, SENDO-LHE ATRIBUÍDO O CONCEITO “EM EXIGÊNCIA”, SENDO QUE A POSSIBILIDADE DE APROVAÇÃO ESTÁ CONDICIONADA À AVALIAÇÃO DA NOVA VERSÃO DO TRABALHO FINAL, SEGUINDO PROCEDIMENTOS PREVISTOS NA RESOLUÇÃO DO PROGRAMA. O PRESIDENTE DA COMISSÃO EXAMINADORA, OUVIDOS OS DEMAIS MEMBROS, DEVERÁ FICAR RESPONSÁVEL POR ATESTAR QUE AS CORREÇÕES SOLICITADAS NA LISTA DE EXIGÊNCIAS FORAM ATENDIDAS NA VERSÃO FINAL DO TRABALHO. A COMISSÃO EXAMINADORA CUMPRINDO OS PRAZOS REGIMENTAIS, ESTABELECE UM PRAZO MÁXIMO DE 30 DIAS PARA QUE SEJAM FEITAS AS ALTERAÇÕES EXIGIDAS. APÓS O DEPÓSITO FINAL DO DOCUMENTO DE TESE, DEVIDAMENTE REVISADO E MEDIANTE ATESTADO DO ORIENTADOR, O CONCEITO "EM EXIGÊNCIA" PASSARÁ IMEDIATAMENTE PARA O DE “APROVADO”. NA FORMA REGULAMENTAR, FOI LAVRADA A PRESENTE ATA, QUE É ASSINADA POR MIM, WELLINGTON LAURENTINO DOS SANTOS, SECRETÁRIO, ALUNO E OS MEMBROS DA COMISSÃO EXAMINADORA PRESENTES.**

4. **CAMPINA GRANDE, 04 DE AGOSTO DE 2022**

5.

Documento assinado digitalmente  
 **ALCIGEIMES BATISTA CELESTE**  
 Data: 29/08/2022 20:37:32-0300  
 Verifique em <https://verificador.iti.br>

6.

7.

8. **ALCIGEIMES BATISTA CELESTE - UFS - EXAMINADOR EXTERNO**



Documento assinado eletronicamente por **WILSON FADLO CURI, PROFESSOR**, em 08/08/2022, às 14:03, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).



Documento assinado eletronicamente por **DAYSE LUNA BARBOSA, PROFESSOR DO MAGISTERIO SUPERIOR**, em 09/08/2022, às 10:15, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).

---



Documento assinado eletronicamente por **ANDREA CARLA LIMA RODRIGUES, PROFESSOR**, em 09/08/2022, às 16:23, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).

---



Documento assinado eletronicamente por **CAMILO ALLYSON SIMOES DE FARIAS, PROFESSOR DO MAGISTERIO SUPERIOR**, em 10/08/2022, às 17:01, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).

---



Documento assinado eletronicamente por **CICERO FELLIPE DINIZ DE SANTANA, PROFESSOR DO MAGISTERIO SUPERIOR**, em 17/08/2022, às 10:28, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).

---



Documento assinado eletronicamente por **WELLINGTON LAURENTINO DOS SANTOS, SECRETÁRIO (A)**, em 17/08/2022, às 10:31, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).

---



Documento assinado eletronicamente por **VALTERLIN DA SILVA SANTOS, PROFESSOR DO MAGISTERIO SUPERIOR**, em 17/08/2022, às 11:13, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).

---



A autenticidade deste documento pode ser conferida no site <https://sei.ufcg.edu.br/autenticidade>, informando o código verificador **2550386** e o código CRC **055442A3**.

---

## AGRADECIMENTOS

Ao meu Deus todo poderoso, meu Senhor e Salvador, que iluminou todos os meus passos durante toda a minha vida.

À minha amada esposa Larissa que me impulsiona a ser alguém melhor todos dias, me dando apoio e proteção para alcançar meus objetivos, como pesquisador, profissional e ser humano. Também ao meu filho Elias, que mesmo tão pequeno é para mim a fonte de coragem para enfrentar os desafios da vida, sendo uma motivação e um orgulho.

A minha mãe que é meu maior exemplo de vida, pelo seu imenso amor, que me faz ser sempre amado em qualquer batalha que a vida me trazer.

Aos meus familiares, por sempre acreditarem na minha capacidade e que meu empenho seria recompensado.

Aos meus orientadores, professor Dr. Wilson Fadlo Curi e professor Dr. Valterlin da Silva Santos, que me ofereceram o direcionamento adequado e a expertise necessária para o desenvolvimento da pesquisa. A minha banca examinadora que com imenso carinho e as muitas horas de dedicação se debruçaram ostensivamente sobre minha tese para que eu pudesse alcançar meu real potencial.

Aos professores Dr<sup>a</sup>. Celeide Maria Belmont Sabino Meira, professora Dr<sup>a</sup>. Ruth Silveira do Nascimento e professor Dr. Rui de Oliveira por terem sido fonte de inspiração para minha vida em todos esses anos de academia.

A todos os colegas do Laboratório de Conservação em Energia Elétrica da UEPB, que fizeram parte dessa conquista, pela disponibilidade e dedicação em aprender e disseminar conhecimento.

Aos meus amigos do CCTA-UFCG Campus Pombal que no momento em que vi a morte, Deus os enviou para que pudessem me salvar e assim ter a chance de continuar lutando pelos meus dias. Aos meus amigos de trabalho e estrada, pelo companheirismo e auxílio durante esta fase importante da minha vida. Aos meus amigos da pós-graduação, por partilharmos dos mesmos objetivos na jornada acadêmica.

À Universidade Federal de Campina Grande por todo o investimento educacional, me fornecendo uma educação de excelente qualidade.

Enfim, a todos os que propiciaram e ajudaram na minha titulação de doutor, pós-graduado em Engenharia Civil e Ambiental.

A Deus que me manteve vivo debaixo do seu imenso amor, para viver meus dias por quem amo. Obrigado por Larissa, Elias e toda a nossa família. “Antes eu te conhecia só por ouvir falar, mas agora eu te vejo com os meus próprios olhos.” Jó 42:5

## RESUMO

Um dos principais desafios enfrentados na gestão dos recursos hídricos tem sido a dificuldade na mensuração dos riscos que os sistemas de reservatórios estão sujeitos, diante das várias possibilidades de cenários hidrológicos. A análise dos riscos como ferramenta de apoio às tomadas de decisão tem feito uso de metodologias que contribuem para a melhor alocação da água para múltiplos usos e objetivos, com destaque para a modelagem matemático-computacional, incluindo aspectos técnicos-operacionais e legais, importantes na definição de estratégias eficientes para o adequado uso da água, minimizando as falhas do sistema. Este trabalho teve como objetivo aplicar o modelo de árvore de decisão e simulação estocástica para a análise de risco de um sistema de reservatórios superficiais. A metodologia baseou-se na geração de dados sintéticos de precipitação e vazão, que, posteriormente, foram utilizados em simulações de balanço hídrico, sendo então calculados os indicadores de risco e aplicado o modelo de árvore de decisão, utilizando a confiabilidade do sistema como variável dependente. A pesquisa foi aplicada à região do alto curso do Rio Piranhas-PB, com destaque para o sistema de reservatórios Engenheiro Ávidos – São Gonçalo. Os resultados obtidos possibilitaram efetuar um diagnóstico da operação e identificação dos pontos críticos para o não atendimento às vazões requeridas pelas demandas do sistema. Foi possível estabelecer uma relação entre as confiabilidades e a alocação das vazões, bem como a precipitação e o volume dos reservatórios, indicando padrões nas condições hidrológicas que levaram aos piores e melhores percentuais de confiabilidade. As árvores de decisão obtiveram desempenho e robustez satisfatórios na previsão do risco para as demandas e cenários de operação, importante para identificação das melhores tomadas de decisões no planejamento dos recursos hídricos.

**Palavras-Chave:** Confiabilidade, Bacias hidrográficas, Aprendizado de Máquina.

## ABSTRACT

Some of the main challenges faced in the water resources management have been the difficulty in measuring the risks that reservoir systems are subject, given the various possibilities of hydrological scenarios. Risk analysis as a tool to support decision-making has used methodologies that contribute to a better allocation of water for multiple uses and objectives, with emphasis on mathematical-computational modeling, including technical-operational and legal aspects, which are important in defining efficient strategies for the proper use of water while minimizing system failures. This work aimed to apply the decision tree model and stochastic simulation for the risk analysis of a surface reservoir system. The methodology was based on the use of synthetic precipitation and flow data generation, which were later used in water balance simulations, where the risk indicators were calculated and the decision tree model was applied, using the reliability of the system as a dependent variable. The research was applied to the region of the upper course of the Piranhas-PB River, with emphasis on the Engenheiro Ávidos – São Gonçalo reservoir system. The obtained results made it possible to make a diagnosis of the operation and the identification of critical points for not meeting the required flows by the demands of the system. It was possible to establish a relationship between the reliability and the allocation of flows, as well as the precipitation and the volume of the reservoirs, indicating patterns in the hydrological conditions that led to the worst and best percentages of reliability. The decision trees achieved satisfactory performance and robustness in predicting the risk for the demands and operating scenarios, which is important for identifying the best decision-making in the planning of water resources.

**Keywords:** Reliability, Watersheds, Modeling.

## LISTA DE TABELAS

<b>Tabela 1</b>	– Resumo das equações dos modelos de desempenho aplicadas a série temporal. ...	42
<b>Tabela 2</b>	– Classes de situação de risco para as confiabilidades médias .....	46
<b>Tabela 3</b>	– Principais características geográficas da Bacia Hidrográficas do Rio Piranhas. ....	51
<b>Tabela 4</b>	– Características fisiográficas das unidades de planejamento estudadas.....	51
<b>Tabela 5</b>	– Disponibilidade hídrica e garantias da região do Alto Piranhas. ....	54
<b>Tabela 6</b>	– Características volumétricas dos reservatórios .....	57
<b>Tabela 7</b>	– Lâminas de evaporação utilizadas na modelagem do sistema.....	57
<b>Tabela 8</b>	– Vazões e fontes hídricas dos componentes de demanda do sistema.....	58
<b>Tabela 9</b>	– Variáveis independentes utilizados para a construção das árvores.....	62
<b>Tabela 10</b>	– Estatística descritiva das confiabilidades para as demandas do CO1 .....	64
<b>Tabela 11</b>	– Desempenho da árvore de decisão para o abastecimento urbano do CO1. ....	65
<b>Tabela 12</b>	– Níveis de importância das variáveis para abastecimento urbano do CO1.....	67
<b>Tabela 13</b>	– Regras de operação para as classes de risco do abastecimento urbano do CO1..	67
<b>Tabela 14</b>	– Desempenho da árvore de decisão para o abastecimento urbano do CO1. ....	69
<b>Tabela 15</b>	– Níveis de importância das variáveis para abastecimento urbano do CO1.....	70
<b>Tabela 16</b>	– Regras de operação para as classes de risco para a irrigação do CO1.....	70
<b>Tabela 17</b>	- Estatística descritiva das confiabilidades para as demandas do CO2 .....	73
<b>Tabela 18</b>	– Desempenho da árvore de decisão para o abastecimento urbano do CO2. ....	76
<b>Tabela 19</b>	– Níveis de importância das variáveis para abastecimento urbano do CO2.....	76
<b>Tabela 20</b>	– Regras de operação para as classes de risco do abastecimento urbano do CO2..	78
<b>Tabela 21</b>	– Desempenho da árvore de decisão para o abastecimento urbano do CO2. ....	79
<b>Tabela 22</b>	– Níveis de importância das variáveis para abastecimento urbano do CO2.....	81
<b>Tabela 23</b>	– Regras de operação para as classes de risco para a irrigação do CO2.....	81
<b>Tabela 24</b>	– Estatística descritiva das confiabilidades para as demandas do CO3.....	84
<b>Tabela 25</b>	– Desempenho da árvore de decisão para o abastecimento urbano do CO3. ....	86
<b>Tabela 26</b>	– Níveis de importância das variáveis para abastecimento urbano do CO3.....	86
<b>Tabela 27</b>	– Regras de operação para as classes de risco do abastecimento urbano do CO2..	88
<b>Tabela 28</b>	– Desempenho da árvore de decisão para o abastecimento urbano do CO2. ....	89
<b>Tabela 29</b>	– Níveis de importância das variáveis para abastecimento urbano do CO3.....	91
<b>Tabela 30</b>	– Regras de operação para as classes de risco para a irrigação do CO3.....	91
<b>Tabela 31</b>	– Estatística descritiva das confiabilidades para as demandas do CO4.....	93

<b>Tabela 32</b> – Desempenho da árvore de decisão para o abastecimento urbano do CO4. ....	96
<b>Tabela 33</b> – Níveis de importância das variáveis para abastecimento urbano do CO4.....	96
<b>Tabela 34</b> – Regras de operação para as classes de risco do abastecimento urbano do CO4..	98
<b>Tabela 35</b> – Desempenho da árvore de decisão para o abastecimento urbano do CO4. ....	100
<b>Tabela 36</b> – Níveis de importância das variáveis para abastecimento urbano do CO4.....	102
<b>Tabela 37</b> – Regras de operação para as classes de risco para a irrigação do CO4.....	103
<b>Tabela 38</b> – Estatística descritiva das confiabilidades para as demandas do CO5.....	105
<b>Tabela 39</b> – Desempenho da árvore de decisão para o abastecimento urbano do CO5. ....	107
<b>Tabela 40</b> – Níveis de importância das variáveis para abastecimento urbano do CO5.....	107
<b>Tabela 41</b> – Diagnóstico do sistema de acordo com as <i>Cer</i> do abastecimento urbano no CO5.....	110
<b>Tabela 42</b> – Desempenho da árvore de decisão para o abastecimento urbano do CO5. ....	111
<b>Tabela 43</b> – Níveis de importância das variáveis para a irrigação do CO5.....	113
<b>Tabela 44</b> – Diagnóstico do sistema de acordo com as <i>Cer</i> da irrigação no CO5.....	114

## LISTA DE ILUSTRAÇÕES

<b>Figura 1</b> – Exemplo de modelo de árvore classificatória. ....	309
<b>Figura 2</b> – Representação gráfica de $p_i$ com relação a $Ent(p_i)$ .....	30
<b>Figura 3</b> – Fluxograma das etapas de aplicação do RAMSS.....	38
<b>Figura 4</b> – Fluxograma de aplicação do modelo RAMSS.....	39
<b>Figura 5</b> – Estrutura de visualização das folhas do modelo .....	48
<b>Figura 6</b> – Exemplo de folha da árvore de decisão.....	49
<b>Figura 7</b> – Regiões do Alto Piranhas e do Médio Piranhas Paraibano.....	52
<b>Figura 8</b> – Localização da sub-bacia do Alto Curso do Rio Piranhas.....	53
<b>Figura 9</b> – Estrutura topológica do sistema Eng. Ávidos-São Gonçalo. ....	55
<b>Figura 10</b> – Estruturação dos cenários operacionais de simulação. ....	59
<b>Figura 11</b> – Estruturação do SEASG para o CO1, CO2 e CO3 no AcquaNet. ....	60
<b>Figura 12</b> – Estruturação do SEASG para o CO4 e CO5 no AcquaNet.....	61
<b>Figura 13</b> – Percentuais de simulações conforme a classificação de risco para o CO1. ....	64
<b>Figura 14</b> – Árvore de decisão para as confiabilidades médias no abastecimento urbano no CO1.....	66
<b>Figura 15</b> – Árvore de decisão para as confiabilidades médias para a irrigação no CO1. ....	71
<b>Figura 16</b> – Percentuais de simulações conforme a classificação de risco para o CO2. ....	75
<b>Figura 17</b> – Árvore de decisão para as confiabilidades médias no abastecimento urbano no CO2.....	77
<b>Figura 18</b> – Árvore de decisão para as confiabilidades médias para a irrigação no CO2. ....	80
<b>Figura 19</b> – Percentuais de simulações conforme a classificação de risco para o CO3. ....	85
<b>Figura 20</b> – Árvore de decisão para as confiabilidades médias no abastecimento urbano no CO3.....	87
<b>Figura 21</b> – Árvore de decisão para as confiabilidades médias para a irrigação no CO3. ....	90
<b>Figura 22</b> – Percentuais de simulações conforme a classificação de risco para o CO4. ....	95
<b>Figura 23</b> – Árvore de decisão para as confiabilidades médias no abastecimento urbano no CO4.....	97
<b>Figura 24</b> – Árvore de decisão para as confiabilidades médias no abastecimento urbano no CO4.....	101
<b>Figura 25</b> – Percentuais de simulações conforme a classificação de risco para o CO5. ....	106

**Figura 26** – Árvore de decisão para as confiabilidades médias no abastecimento urbano no CO5.....108

**Figura 27** – Árvore de decisão para as confiabilidades médias para a irrigação no CO5. .... 112

## LISTA DE ABREVIATURAS E SIGLAS

ACF – Função de Autocorrelação

AESA – Agência Executiva de Gestão das Águas do Estado da Paraíba

AR – Modelo Autorregressivo

ARIMA – *Autoregressive Integrated Moving Average*

ARMA – *Autoregressive and Moving Average*

BHRPPA – Bacia Hidrográfica do Rio Piancó Piranhas-Açu

C – Indicador de Confiabilidade de Hashimoto

CBHAP – Comitê da Bacia Hidrográfica do Alto Piranhas

CDF – Função de Densidade Acumulada

*Cer* – Classe do estado de risco

EHS – Eventos Hidrológicos Sintéticos

CO – Cenário Operacional

DF – Teste de *Dickey-Fuller*

DNOCS – Departamento Nacional de Obras Contra as Secas

GOTA – Grupo de Otimização Total da Água

INMET – Instituto Nacional de Meteorologia

KPSS – Teste de *Kwiatkowski-Phillips-Schmidt-Shin*

LB – Teste de *Ljung-Box*

MA – Modelo de Médias Móveis

MAE – *Mean Absolute Error*

MASE – *Mean Absolute Standard Error*

ME – *Mean Error*

MF – Método dos Fragmentos

PACF – Função de Autocorrelação Parcial

PDRHP – Plano Diretor de Recursos Hídricos do Rio Piancó

PERHPB – Plano Estadual de Recursos Hídricos do Estado da Paraíba

PISF – Projeto de Integração do Rio São Francisco

*Pma* – Precipitações médias anuais

PRHPPA – Plano de Recursos Hídricos do Rio Piancó-Piranhas-Açu

*Qms* – Vazões médias sintéticas

R – Indicador de Resiliência de Hashimoto

RAMSS – Risk Analysis Model by Stochastic Simulation  
Rco – Risco do cenário operacional  
REA – Reservatório Eng. Ávidos  
RLM – Modelo de regressão linear múltipla  
RMSE – *Root Mean Standard Error*  
RN – Modelo de Redes Neurais  
RSG – Reservatório São Gonçalo  
S – Índice de Sustentabilidade de Hashimoto  
SARIMA – *Seasonal Autoregressive Integrated Moving Average*  
SE – *Standard Error*  
SEASG – Sistema de reservatórios Engenheiro Ávidos e São Gonçalo  
SH – Série Histórica  
SMAP - *Soil Moisture Accounting Procedure*  
TF – Modelo Thomas-Fiering  
UPHAP – Unidade de Planejamento Hidrológico do Alto Curso do Rio Piranhas  
UPHMP – Unidade de Planejamento Hidrológico do Médio Curso do Rio Piranhas  
V – Indicador de Vulnerabilidade de Hashimoto  
Vmr – Volumes médios das simulados

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>16</b>
1.1	OBJETIVO GERAL .....	20
1.2	OBJETIVOS ESPECÍFICOS .....	20
<b>2</b>	<b>REFERENCIAL TEÓRICO .....</b>	<b>21</b>
2.1	IMPORTÂNCIA E CENÁRIO DOS RECURSOS HÍDRICOS.....	21
2.2	OPERAÇÃO DE RESERVATÓRIOS.....	22
2.3	ANÁLISE DE RISCOS .....	25
2.4	ÁRVORES DE DECISÃO .....	29
<b>3</b>	<b>METODOLOGIA .....</b>	<b>37</b>
3.1	CLASSIFICAÇÃO DA PESQUISA .....	37
3.2	ABORDAGEM METODOLÓGICA E INOVAÇÃO CIENTÍFICA .....	38
3.3	ETAPAS DO RAMSS .....	40
<b>3.3.1</b>	<b>Etapa I – Preparação dos dados iniciais de precipitação .....</b>	<b>40</b>
<b>3.3.2</b>	<b>Etapa II – Análise estatística e geração sintética.....</b>	<b>41</b>
<b>3.3.3</b>	<b>Etapa III – Geração sintética das séries de precipitação.....</b>	<b>42</b>
<b>3.3.4</b>	<b>Etapa IV – Ajustes de correlação entre reservatórios .....</b>	<b>43</b>
<b>3.3.5</b>	<b>Etapa V – Modelagem hidrológica das vazões .....</b>	<b>44</b>
<b>3.3.6</b>	<b>Etapa VI – Simulação de balanço hídrico do sistema.....</b>	<b>45</b>
<b>3.3.7</b>	<b>Etapa VII – Cálculo dos indicadores de desempenho.....</b>	<b>45</b>
<b>3.3.8</b>	<b>Etapa VIII - Modelagem das árvores de decisão.....</b>	<b>46</b>
3.4	ESTUDO DE CASO.....	50
<b>3.4.1</b>	<b>Bacia Hidrográfica do Rio Piancó-Piranhas-Açu .....</b>	<b>50</b>
<b>3.4.2</b>	<b>Região do Alto Curso do Rio Piranhas.....</b>	<b>53</b>
<b>3.4.3</b>	<b>Sistema de reservatórios Engenheiro Ávidos – São Gonçalo .....</b>	<b>55</b>
<b>3.4.4</b>	<b>Elaboração dos cenários operacionais.....</b>	<b>58</b>
<b>4</b>	<b>RESULTADOS.....</b>	<b>64</b>
4.1	CENÁRIO OPERACIONAL 1 .....	64
<b>4.1.1</b>	<b>Árvore de decisão para demandas abastecimento urbano do CO1.....</b>	<b>65</b>
<b>4.1.2</b>	<b>Árvore de decisão para irrigação do CO1 .....</b>	<b>70</b>
4.2	CENÁRIO OPERACIONAL 2 .....	74
<b>4.2.1</b>	<b>Árvore de decisão para abastecimento urbano do CO2.....</b>	<b>75</b>
<b>4.2.2</b>	<b>Árvore de decisão para irrigação do CO2 .....</b>	<b>80</b>

4.3	CENÁRIO OPERACIONAL 3 .....	84
4.3.1	Árvore de decisão para o abastecimento urbano do CO3 .....	85
4.3.2	Árvore de decisão para a irrigação do CO3 .....	90
4.4	CENÁRIO OPERACIONAL 4 .....	94
4.4.1	Árvore de decisão para o abastecimento urbano do CO4 .....	95
4.4.2	Árvore de decisão para a irrigação do CO4 .....	100
4.5	CENÁRIO OPERACIONAL 5 .....	104
4.5.1	Árvore de decisão para o abastecimento urbano do CO5 .....	105
4.5.2	Árvore de decisão para a irrigação do CO5 .....	111
5	<b>DISCUSSÃO</b> .....	117
5.1	ANÁLISE DOS CENÁRIOS .....	117
5.2	ÁRVORES DE DECISÃO PARA MODELAGEM DO RISCO .....	122
5.3	AVALIAÇÃO DA PROPOSTA METODOLÓGICA .....	124
6	<b>CONCLUSÃO</b> .....	1255
7	<b>RECOMENDAÇÕES PARA TRABALHOS FUTUROS</b> .....	126
	<b>REFERÊNCIAS</b> .....	128
	<b>ANEXOS</b> .....	136

## 1 INTRODUÇÃO

Devido ao crescimento populacional e as alterações climáticas em todo o mundo, diversas regiões têm sofrido com a escassez hídrica, atingindo negativamente a saúde pública, a economia e o modo de vida das populações, o que limita o desenvolvimento e que, possivelmente, motivarão confrontos bélicos severos futuramente.

O Brasil, por seu potencial de disponibilidade hídrica para consumo humano, se destaca internacionalmente, o que provoca a contínua preocupação e atualização legal de técnicas do manejo de seus recursos hídricos. A região Nordeste merece destaque nesse sentido, pois apesar da situação hídrica do país, possui diversos problemas crônicos relacionados a falta de água (SILVEIRA; SILVA, 2019).

O cenário na região é, por vezes, alarmante, dado o aumento das temperaturas médias, da evapotranspiração, seguido pela redução das precipitações e da oferta de água para alguns fins, a exemplo da agricultura, sendo necessária a racionalização do uso da água, afetando diretamente a produtividade de tais atividades (SAMPAIO *et al.*, 2018).

Entre os anos de 2010 a 2016, o Nordeste brasileiro enfrentou uma das maiores secas já registradas. Segundo uma análise realizada pela Confederação Nacional de Municípios (CNM) no ano de 2013, a seca na Região Nordeste afligiu 38% da população da região. As políticas públicas emergenciais do Governo Federal buscaram, de forma remediativa, amenizar, mas não solucionar as questões pertinentes à seca. Algumas medidas já anteriormente implementadas, como auxílios financeiros à população, abrangem o atendimento de cerca 1,5 milhão de habitantes do sertão, que tem contribuído para evitar o grande êxodo rural já ocorrido em períodos históricos, como nas décadas de 1970 e 1980 (CNM, 2013; SILVA *et al.*, 2017).

Segundo Ahmadi, Haddad e Loáiciga (2015), as variações características de precipitação e temperatura podem afetar as vazões dos rios, tanto em sua intensidade como na duração. O monitoramento das chuvas e o sistema integrado de informações do país, importantes para o estudo da representação matemática e computacional dos fenômenos hidrológicos, ainda carece de estruturação e consistência na coleta dos dados. Por isso, a integração de várias técnicas de análise tem contribuído para uma solução nesse sentido.

Apesar da evolução tecnológica no planejamento, as ferramentas mais adotadas para interpretação dos sistemas e solução dos problemas de gestão acabam sendo aquelas baseadas em processos determinísticos, apresentando limitações quanto as mais variadas influências as quais os recursos hídricos estão sujeitos (SUN *et al.*, 2018).

A literatura aborda a necessidade de consideração das incertezas e dos processos estocásticos, tanto no escoamento superficial quanto no desempenho dos sistemas de reservatórios, destacando sua importância e que não devem ser negligenciadas na gestão dos recursos hídricos, podendo ser um problema recorrente em análises de sistema de abastecimento (NOURANI *et al.*, 2020).

Os estudos ligados a alocação de água para prevenção de escassez hídrica vêm sendo cada vez mais desenvolvidos (CHENG *et al.*, 2017). Em grande parte são considerados os níveis de confiabilidade do atendimento às demandas em longo prazo, através das análises de cenários de aportes de água em reservatórios, insumos sazonais, mudanças nas prioridades de alocação, incluindo a otimização e aplicação de análise estocástica de vazões utilizando simulação, entre outros (KLEMEŠ, 1969; WURBS, 2005; YAMOUT; HATFIELD; ROMEIJN, 2007; HIGGINS; ARCHER; HAJKOWICZ, 2008; KUO; HSU, 2011; BORYCZKO; TCHÓRZEWSKA-CIEŚLAK, 2012; GIULIANI; GALELLI; SONCINI-SESSA, 2014; ZHANG *et al.*, 2015; CHENG *et al.*, 2016; KHALIFA, 2018).

Para ajudar a solucionar esses problemas, o uso da análise de riscos, por sua vez, viabiliza a avaliação dos sistemas de recursos hídricos e das regras de operação dos reservatórios, assim como permite uma melhor compreensão da complexidade do problema de compatibilização das ofertas e demandas de água entre os usos múltiplos, presentes nas bacias hidrográficas, bem como auxilia nas tomadas de decisão e na identificação do desempenho e das falhas nos sistemas de distribuição (SUN *et al.*, 2018; CASTRO; MARQUES, 2019; SUGAHARA *et al.*, 2020).

Como forma de melhorar essas análises, o aprendizado de máquina, que é uma das áreas da ciência que utiliza a modelagem matemático-computacional, tem sido utilizado para a construção de modelos (regressão, redes neurais, árvores de decisão etc), visando a melhoria da análise em diversos sistemas, inclusive nos recursos hídricos, através da previsão de variáveis, classificação de fenômenos ou estados, agrupamentos de elementos, reconhecimento de padrões, entre outras formas (ZHANG *et al.*, 2018; FAN *et al.*, 2022; ZHU *et al.*, 2022).

Dentre os modelos mais estudados está a árvore de decisão, um modelo computacional que possibilita o aprendizado de máquina supervisionado, com foco na classificação ou regressão para um conjunto de dados (OKADA; NEVES; SHITSUKA, 2019). São modelos que tem como característica a construção de classificadores para a previsão da classe de um conjunto de dados desconhecidos, com base nos valores da base de dados usada para o treinamento do modelo, além da capacidade de ter seus resultados representados de forma

visual, o que facilita a interpretação e compreensão do comportamento dos variáveis (JÚNIOR, 2018).

É possível que a construção de modelos, como as árvores de decisão, com finalidade de adequação da oferta e da demanda de água, sem depender da existência de grandes bases de dados históricos, considerando situações hidro climáticas desfavoráveis e utilizando análise de risco para o compreensão de problemas na operação dos sistemas de recursos hídricos, beneficie o planejamento dos sistemas de reservatórios quanto a identificação das falhas no atendimento as demandas, facilitando o entendimento sobre a influência dos aspectos hidrológicos no risco.

Este trabalho portanto, teve como proposta a aplicação das árvores de decisão visando a melhor compreensão e análise dos riscos para o diagnóstico da operação dos sistemas de reservatórios superficiais. Foram utilizadas, como variáveis para o modelo, os resultados de simulações realizadas com base em dados de precipitação e de vazão gerados sinteticamente.

Os resultados das simulações foram utilizados posteriormente no cálculo do indicador de confiabilidade de Hashimoto, Stedinger e Loucks (1982), que juntamente com os dados sintéticos serviu como variável dependente para modelagem das árvores de decisão e, com isso, indicar as principais faixas de valores das vazões e precipitações relacionadas com o indicador.

A área de estudo escolhida para a aplicação do modelo foi o sistema de reservatórios Engenheiro Ávidos – São Gonçalo, localizado na Bacia Hidrográfica do Rio Piancó-Piranhas-Açu, mais especificamente a Unidade de Planejamento Hidrológico do Alto Curso do Rio Piranhas – UPHAP.

O sistema foi escolhido devido ao histórico recorrente de crises em seus reservatórios causadas pela escassez de água, característica do semiárido nordestino brasileiro. Outro quesito importante foi a influência hidrológica de alguns reservatórios relevantes e interdependentes para as bacias hidrográficas situadas no Estado da Paraíba, sendo estes reservatórios de grande importância para gestão dos recursos hídricos em nível estadual e federal.

Além disso, a UPHAP possui uma quantidade suficiente de informações que são indispensáveis para a aplicação de técnicas e métodos de engenharia para uso em pesquisas científicas, visto que alguns trabalhos acadêmicos já se debruçaram sobre esta região, a exemplo das produções científicas como os trabalhos de Neto (2018), Nepomuceno (2018) e Neto (2020), que subsidiam estudos relevantes sobre os aspectos ambiental, social, industrial, agropecuário técnico e econômicos para locais semelhantes a UPHAP.

Assim como os trabalhos mencionados, esta pesquisa também faz parte dos estudos desenvolvidos pelo grupo GOTA – Grupo de Otimização Total da Água – da UFCG – Universidade Federal de Campina Grande, que têm trabalhado no desenvolvimento de modelos

e softwares, com o intuito de prover ferramentas para planejamento e gestão de sistemas de recursos hídricos.

Paralelamente a criação de instrumentos de gestão (sistemas de informação, modelos de otimização, criação de indicadores e índices, prover análise multicriterial e multidecisor), o GOTA também agrega estudos que levam em consideração as variabilidades climatológicas e estudos relacionados as análises do atendimento das demandas por meio da análise de riscos e de diversas outras técnicas.

O desenvolvimento do modelo teve o intuito de obter o panorama dos riscos na gestão dos sistemas de reservatórios da região em estudo, para facilitar a definição de estratégias e políticas a serem executadas. A obtenção do panorama baseou-se na previsão do estado de risco e dos cenários com mudanças nas regras de operação, com base nos padrões e comportamentos de fatores hidrológicos e de distribuição, envolvendo os processos estocásticos, categorizando e qualificando a probabilidade de ocorrência de eventos adversos, bem como a severidade dos eventos hidrológicos.

A pesquisa visou possibilitar aos planejadores e gestores melhorar o entendimento sobre os pontos críticos de controle do sistema de reservatórios, assim como as decisões na operação, com foco na melhor alocação da água, conciliados com a diminuição dos conflitos nos sistemas de reservatórios. Para isso, teve por finalidade identificar o risco do sistema para diferentes cenários operacionais, através da simulação da alocação de água em atendimento aos requerimentos das demandas com seus múltiplos usos.

Essa pesquisa também buscou viabilizar o desenvolvimento de novos trabalhos científicos no âmbito da análise de risco na gestão dos recursos hídricos, devido ao seu caráter exploratório, potencial de aplicação prática e a abordagem estocástica. Com isso, leva a construção de novas perspectivas quanto a avaliação da operação e análise de riscos dos sistemas de reservatórios. Ainda há a possibilidade de diversas variações e modificações no processo de modelagem do risco.

## 1.1 OBJETIVO GERAL

Aplicar o modelo de árvore de decisão e simulação estocástica para a análise de risco dos sistemas de reservatórios superficiais, tendo como área de estudo a região do alto curso do Rio Piranhas-PB, com destaque para o sistema de reservatórios Engenheiro Ávidos – São Gonçalo.

## 1.2 OBJETIVOS ESPECÍFICOS

- I. Avaliar através da geração sintética de dados e da simulação, a influência das variações das precipitações e das vazões sob o risco no sistema de estudo;
- II. Elaborar e analisar os cenários hidrológicos e operacionais gerados através do indicador de confiabilidade do sistema;
- III. Identificar os pontos críticos de risco do sistema de recursos hídricos estudado através das árvores de decisão;
- IV. Propor, por meio dos resultados do diagnóstico da operação do sistema de acordo com o risco, estratégias de operação para o planejamento dos recursos hídricos;
- V. Analisar o desempenho da proposta metodológica de utilização das árvores de decisão na análise de risco nos sistemas de reservatórios superficiais.

## 2 REFERENCIAL TEÓRICO

### 2.1 IMPORTÂNCIA E CENÁRIO DOS RECURSOS HÍDRICOS

Desde a Conferência Mundial das Nações Unidas em 1992, o uso dos recursos hídricos possui princípios fundamentais, que estabelecem a água como um recurso vulnerável, essencial para a manutenção da vida, do desenvolvimento e do meio ambiente e possui valor econômico para todos os seus usos e deve ser considerada como um bem econômico. Atualmente o seu uso está presente nas mais variadas atividades do ser humano e, ao longo da história, o domínio sobre os conhecimentos necessários à sua utilização foi fundamental para a evolução da sociedade e das civilizações (MORGADO, 2019).

A transposição do Rio São Francisco para o atendimento às demandas hídricas da região Nordeste, em operação desde o ano de 2018, já se configura como a maior obra de transferência de água do país. A transposição do Rio São Francisco é uma alternativa técnica que desde as épocas coloniais, vem sendo vista como impulsionadora do desenvolvimento regional e geradora de empregos nos vários setores como a agricultura, pecuária, além de contribuir para a geração de energia e suporte a mineração (VIEIRA, 2011).

As várias iniciativas para tentar solucionar os problemas gerados pela falta de água não estão somente presentes no Brasil, mas no mundo inteiro. Segundo Rivero (2018), aparentemente, os volumes quantificados de água doce no planeta apresentam-se satisfatórios, o que tem pouquíssima significância, devido ao fato da distribuição das reservas d'água ser preocupantemente heterogênea.

Com a escassez de água cada vez mais frequente no contexto global, é necessário em alguns casos e locais, que sejam implementadas medidas para dissolução de conflitos, mesmo que por meio de restrições do uso. Os problemas acarretados pela diminuição da oferta hídrica levam inevitavelmente a redução do desenvolvimento econômico e principalmente a degradação da qualidade de vida da população (PULIDO-VELAZQUEZ; CABRERA MARCET; GARRIDO COLMENERO, 2014).

A Organização das Nações Unidas – ONU – em seu Relatório Mundial sobre o Desenvolvimento dos Recursos Hídricos de 2019, afirma que a alta demanda proveniente da produção agrícola, industrial e energética aumentará significativamente nas próximas décadas. Dentre estes setores, o agrícola é o maior consumidor, responsável por 70% desse consumo, que nos últimos 20 anos tem apresentado um crescimento acelerado. Esse repentino crescimento,

advém do maior número de áreas irrigadas, da interiorização da agricultura nos continentes e do aumento exacerbado da produção pecuária (ONU, 2019).

Segundo a ONU o problema da água é preocupante, pois a busca por água de boa qualidade tem se tornado cada vez mais difícil nos últimos anos. De acordo com a organização, quase 900 milhões de pessoas vivem sem acesso à água potável. Além desse valor, calcula-se que 40% da população mundial (2,6 bilhões de habitantes) estão desprovidas de serviços de saneamento básico. Diante deste cenário, viu-se a importância das nações unidas reconhecer, dentre as metas de desenvolvimento do milênio, a água e o saneamento como direitos humanos (PICCOLI *et al.*, 2016; ONU, 2019).

## 2.2 OPERAÇÃO DE RESERVATÓRIOS

As atividades humanas em bacias hidrográficas (rurais ou urbanas) provocam, em longo prazo, alterações expressivas na dinâmica da água, podendo reduzir a disponibilidade dos mananciais. Esse processo, que ocorre naturalmente, pode ser potencializado pela combinação de desequilíbrio dos aportes pluviométricos e manejo inadequado dos solos. Uma série de outros problemas advindos dessa cadeia de atividades mal planejadas podem levar a escassez de água em algumas sub-bacias (VIEIRA, 2011).

Devido a importância do correto manejo das bacias hidrográficas, a operação dos reservatórios deve partir do adequado planejamento e da gestão eficiente dos recursos hídricos sem estes princípios fundamentais, os sistemas de abastecimento estão fadados a serem deteriorados seguidos pelo agravamento dos diversos conflitos ligados ao uso da água nas bacias hidrográficas.

No processo de distribuição dos recursos hídricos, a situação ideal seria aquela em que os usuários recebem plenamente estes recursos conforme suas necessidades. Entretanto, a escassez faz com que a distribuição seja direcionada para os setores prioritários, que possuem melhor retorno social, levando em consideração o atendimento aos objetivos comuns de todos os usuários.

Reservatórios de grande porte desempenham um papel essencial na gestão de recursos hídricos para irrigação agrícola, abastecimento e controle de enchentes. Segundo Zhang *et al.* (2018), reservatórios e barragens são infraestruturas vitais construídas pelo homem que desempenham papéis essenciais no controle de inundações, geração de energia, abastecimento

de água, navegação e outras funções. A realização dessas funções exige operação eficiente do reservatório e os controles efetivos sobre a vazão de um reservatório ou barragem. É necessário que sejam utilizados sistemas robustos de operação de reservatórios sob fluxo normal e em condições extremas.

Os estudos sobre as operações dos reservatórios têm sido realizados de maneira ostensiva pela comunidade acadêmica principalmente na resolução de problemas existentes durante as secas (SANTOS, 2011; VIEIRA, 2011; CHANG; WANG; TSAI, 2016; VAROUCHAKIS; PALOGOS; KARATZAS, 2016; JAISWAL, 2018; RONG *et al.*, 2019).

Wu *et al.* (2018) apresentaram um estudo com uma estrutura útil para examinar as influências das regras de operação do reservatório durante períodos de seca nas correlações multi escala entre secas hidrológicas e secas meteorológicas. Os resultados indicam que as operações do reservatório exerceram uma influência significativa (impacto positivo) na evolução da seca hidrológica de curto prazo (1 e 3 meses) reduzindo a duração da seca e aliviando sua magnitude e alterou a correlação linear entre as secas. Os impactos da operação do reservatório na condição de seca na relação entre secas hidrológicas e secas meteorológicas eram claramente diferentes daqueles sob regras normais. Este estudo de Wu *et al.* (2018), forneceu mais informações para os formuladores de políticas sobre o gerenciamento de reservatórios durante uma seca.

Ao longo da última década, as técnicas de Inteligência Artificial (IA) tornaram-se cada vez mais populares no campo de previsões, planejamento de operação de reservatórios e abordagens de programação. O uso dos modelos tem ajudado no desenvolvimento das regras de operação de um reservatório a partir das bases de dados históricos já existentes. Esses modelos demonstram capacidade em simular condições desfavoráveis de para o sistema, ajudando definição da operação que ajudam a minimizar o problema (ZHANG *et al.*, 2018).

Segundo Yassin *et al.* (2019), os reservatórios afetam significativamente os regimes de escoamento em sistemas de bacias hidrográficas, alterando a magnitude e o tempo de escoamentos. A não representação desses efeitos limita o desempenho de modelos hidrológicos nas muitas bacias altamente regulamentadas em todo o mundo e limita a aplicabilidade de tais modelos para investigar o futuro dos sistemas de bacias hidrográficas através da análise de cenários (por exemplo, cenários de clima, uso da terra ou regulação do reservatório mudanças). Ainda segundo os autores, barragens e reservatórios alteram os regimes de fluxo natural em rios tanto em termos de magnitude quanto de tempo de fluxo. Assim, para rios que contêm grandes ou pequenas barragens e reservatórios, os regimes de fluxo são uma combinação de condições naturais e gerenciadas. Várias correntes de pesquisa dentro da comunidade acadêmica que

estudam a modelagem da operação dos reservatórios, gerenciam essa mistura de fluxos naturais e administrados de forma diferente.

Archfield *et al.* (2015) compararam três famílias de modelos para operação de reservatórios que podem ser usados em escalas continentais: Modelos de Captação (MC), Modelos Globais de Segurança Hídrica (MGSH) e Modelos de Superfície terrestre (MS). Os MC geralmente ignoram a gestão da água e se concentram em captações de cabeceiras não gerenciadas. Os MGSH focam em questões de gestão da água em grande escala e usados em simulações de escoamento, que são dificultadas pela falta de dados relacionados com a gestão e a operação. Os MS terrestre tradicionalmente se concentram em fornecer condições de contorno mais baixas para modelos atmosféricos, mas estão sendo cada vez mais usados para aplicações hidrológicas, os quais são conhecidos como Modelos Hidrológicos Superficiais (MHS).

Uma quarta família de modelos, que são relevantes para o trabalho aqui apresentado, são os modelos de gestão da água (MGA). Modeladores de água que sabem como a água é gerenciada dentro de suas bacias de interesse geralmente usam MGA que contêm representações muito detalhadas das decisões de gestão, mas muitas vezes consideram os processos de fluxo natural de uma forma muito mais rudimentar do que os MC.

Saab *et al.* (2022) realizaram um estudo que revisou estudos anteriores que usaram métodos de otimização, simulação e Inteligência Artificial (IA) para encontrar políticas operacionais ótimas para sistemas de reservatórios nos últimos 20 anos. De fato, políticas operacionais bem-sucedidas não podem ser obtidas sem previsões precisas das principais variáveis hidrológicas num sistema de reservatórios, que são entradas e evaporação. O estudo se concentrou em fornecer uma visão geral das aplicações de modelos baseados em IA para prever afluências e evaporação do reservatório. Segundo os autores, nos últimos anos, observou-se um aumento significativo no tipo e número de métodos desenvolvidos que podem ser usados para modelar e prever variáveis hidrológicas, incluindo métodos populares baseados em dados.

Atualmente, a literatura conta com grande número de estudos que envolvem métodos onde são requeridos muitos dados meteorológicos observados e portanto, muito mais propensos a erros. A aplicação desses métodos nem sempre é realizada de forma fácil pois não é possível ter todos os dados disponíveis de todos os lugares estudados. Existe também o problema da utilização dos métodos que envolvem equações empíricas, pois estas não são tão eficientes quanto o necessário devido à complexidade não linear das variáveis observados para as regiões estudadas (YANG *et al.*, 2019).

Conforme Saab *et al.* (2022), durante as últimas décadas, muitos métodos modernos têm sido desenvolvidos para prever as variáveis hidrológicas de um reservatório como os modelos de IA. Tais métodos têm recebido muito interesse de hidrólogos por lidarem com padrões não estáticos, dinâmicos e estocásticos de dados hidrológicos. Os modelos baseados em IA são caracterizados por terem a capacidade de lidar com uma grande quantidade de dados. Especificamente, no campo da previsão dos principais parâmetros em um sistema de reservatório, existem muitos métodos de IA que têm sido usados, como Redes Neurais Artificiais (RNA), Máquina de Vetores de Suporte (MVS), Lógica Fuzzy (LF) entre outros.

É possível, portanto, afirmar que a otimização e a simulação dos sistemas de reservatórios contribuem para garantia de sua correta operação e uso ideal dos recursos hídricos, contudo é importante que essas técnicas sejam realizadas através de modelos que ofereçam precisão e eficácia nos resultados. A operação dos reservatórios deve ser realizada através de uma visão abrangente dos métodos e modelos aplicados, bem como através do fornecimento de informações necessárias para modelagem, expandindo o horizonte de pesquisa e aumentando o potencial dos modelos.

### 2.3 ANÁLISE DE RISCOS

A quantificação e qualificação dos riscos funcionam de forma eficiente como instrumento da política ambiental, que busca fazer com que programas, planos ou políticas sejam considerados, adotando medidas que eliminem ou os reduzam a níveis toleráveis. O intuito é de, por meio do conhecimento científico de engenharia, contribuir com o planejamento urbano das bacias hidrográficas e com a melhoria da gestão dos recursos hídricos com vistas a preservação ambiental dos rios.

A quantidade de recursos hídricos em todo o mundo, utilizados para o desenvolvimento humano, seja na produção de produtos ou no consumo vem aumentando nos últimos séculos. Entretanto, a falta de cuidados com o meio ambiente, falta de tratamento dos efluentes poluidores e a prevenção e condução correta dos poluentes, ocasionou a diminuição da oferta de água potável para satisfazer diversos usos (LEONETI; DO PRADO; DE OLIVEIRA, 2011).

Os riscos associados a oferta desequilibrada e a carência da água são determinantes para evolução da sociedade, como fator condicionante para o desenvolvimento das regiões que apresentam déficit hídrico (PEIXINHO, 2010).

A solução mais eficaz para o combate a estes problemas é a correta gestão dos recursos hídricos, que atua na formulação de princípios e diretrizes para a estruturação de sistemas gerenciais e à tomada de decisão. Os objetivos da gestão dos recursos hídricos, nesse sentido, são de promover o uso, controle e segurança no atendimento à população (PEIXINHO, 2010; CHIODI; SARCINELLE; UEZU, 2013).

O fato é que a realidade brasileira atualmente apresenta deficiências no que diz respeito ao saneamento e aos recursos hídricos, desde a perspectiva local até o panorama nacional, afetando todos os espectros da população. O planejamento urbano e o modelo de crescimento econômico adotado no país levam ao surgimento de grupos e segmentos sociais afligidos pelos problemas ambientais, favorecendo a proliferação de doenças e parasitas, que atingem as populações mais pobres, as quais vivem em condições sanitárias inadequadas (LISBOA; HELLER; SILVEIRA, 2013).

Assim para que os problemas de âmbito ambiental da gestão dos recursos hídricos possam ser debatidos e encontradas vias para a melhoria da qualidade de vida das comunidades, é necessário que haja instrumentos de gestão que auxiliem na construção do conhecimento em torno da problemática (PEIXINHO, 2010; LISBOA; HELLER; SILVEIRA, 2013).

Para auxiliar na tomada de decisões sobre o assunto, os estudos de análise de riscos se fazem necessários e têm sido utilizados em várias aplicações no âmbito dos recursos hídricos. Li e Huang (2013) afirmam que por décadas, a escassez de água, o abastecimento não confiável, a má qualidade da água e os desastres de inundação têm sido considerados os principais problemas discutidos em escala de bacias hidrográficas. Ainda segundo esses, a complexidade dos sistemas faz com que os planejadores considerem a análise aprofundada dos riscos na tomada de decisões, por meio da implementação da modelagem e da programação (LI; HUANG, 2013).

Nesse contexto, a análise de riscos necessita da capacidade de identificar as vulnerabilidades, os riscos ambientais das atividades humanas atuantes na bacia e auxiliar na identificação das alternativas e soluções mais eficientes. O procedimento de análise deve seguir uma metodologia clara, que forneça resultados objetivos e práticos, composta por um conjunto de etapas adaptáveis às peculiaridades do problema, para qualquer cenário avaliado.

O desenvolvimento e a evolução das ferramentas computacionais promoveram o aumento da velocidade e da capacidade de processamento de informações, que possibilitaram o aprimoramento dos métodos de análise de riscos. Um exemplo é o uso da simulação como uma etapa fundamental da análise, que favorece o estudo de cenários e a obtenção de resultados

de variáveis qualitativas e quantitativas, servindo como parâmetros de avaliação para diferentes processos de decisão (CHENG *et al.*, 2017).

Um das principais ferramentas de análise de riscos na gestão de sistemas de recursos hídricos têm sido os indicadores de desempenho, chamados de confiabilidade, resiliência e vulnerabilidade, propostos e organizados por Hashimoto, Stedinger e Loucks (1982). A utilidade da avaliação por meio dos indicadores está relacionada com o auxílio à tomada de decisão na compreensão do comportamento dos sistemas, suas configurações, políticas operacionais e objetivos (CASTRO; MARQUES, 2019).

A Confiabilidade –  $C$  tem como intuito mensurar a frequência que um sistema possui em atender às demandas, representando a probabilidade que o sistema tem em não falhar. O seu entendimento pode ser considerado como uma mensuração de não risco, pois a alta confiabilidade do sistema indica uma menor quantidade de tempo com o sistema apresentando falha no atendimento às demandas.

De acordo com Araújo *et al.* (2020), para cada demanda de um sistema de recursos hídricos com seus respectivos níveis de atendimento desejáveis, é possível identificar os intervalos de fornecimento satisfatório e não satisfatório. A Equação 1 expressa o cálculo para obtenção da confiabilidade.

$$C = P\{X_t < X_0\} = 1 - \frac{n}{N} \quad (1)$$

Na equação 1,  $X_0$  corresponde ao valor do nível de atendimento estabelecido para a variável e  $X_t$  o valor observado no tempo  $t$ . A falha, portanto, é identificada quando  $X_t < X_0$ . O número de intervalos de tempo é representado por  $n$  (hora, dia, mês, etc.) em que  $X_t < X_0$  e  $N$  o tempo de duração ou tamanho do registro observado.

A Resiliência –  $R$ , por sua vez, possui um intuito mais relacionado a duração dos eventos de falha, diferente da verificação de ocorrências, papel da confiabilidade. A resiliência tem como característica representar numérica e percentualmente, a capacidade que o sistema possui em se recuperar de um atendimento insatisfatório, definida como o inverso do valor esperado do tempo médio de permanência em falha.

A Equação 2 apresenta a formulação numérica para mensuração da resiliência,

$$R = \frac{1}{E\{d\}} = \left[ \frac{1}{M} \sum_{j=1}^M d_j \right]^{-1} \quad (2)$$

onde  $E$  é a esperança do tempo médio de permanência em falha;  $d_j$  representa a duração da falha no tempo  $j$ , num intervalo de tempo, com  $j = 1, \dots, M$ , onde  $M$  é o número de eventos de ocorrência de falhas.

Segundo De Andrade, Curi e Curi (2009), quanto maior o período que o sistema permanece insatisfatório, menos resiliente é, sendo os sistemas mais resilientes os de rápida recuperação. A pouca resiliência é caracterizada por eventos prolongados de falhas, com lenta recuperação.

Já a Vulnerabilidade –  $V$ , dentre os demais indicadores, não está diretamente relacionada com o conceito de desvio ou falha temporal, por frequência ou duração do déficit, mas sim ao conceito de severidade ou magnitude do problema enfrentado no atendimento da demanda, como por exemplo a quantidade de água em escassez em um reservatório.

É importante destacar que o sistema com maior vulnerabilidade também possuirá o maior déficit hídrico (DE ANDRADE; CURI; CURI, 2009). A vulnerabilidade pode ser expressa conforme a Equação 3,

$$V = E\{s\} = \frac{1}{M} \sum_1^M S_j \quad (3)$$

onde  $E$  é o valor esperado (magnitude média) do déficit,  $M$  é o número total de intervalos de falha e  $S_j$  o volume do déficit para cada intervalo de falha  $j$ .

Os três indicadores em conjunto são aplicados como forma de verificar a sustentabilidade dos sistemas hídricos, auxiliando no comparativo entre diferentes políticas de gestão por meio de cenários simulados para bacias hidrográficas. É possível, por meio da combinação e análise integrada os indicadores, avaliar as ações realizadas nos componentes do sistema, identificando os mais vulneráveis e menos resilientes, com o intuito de contribuir para a elaboração das melhores políticas (SANDOVAL-SOLIS; MCKINNEY; LOUCKS, 2011; DIAO, 2016).

A união dos indicadores de desempenho descritos anteriormente, resulta no valor do índice de sustentabilidade, conforme apresentado por Loucks (1997), sendo utilizado para a avaliação dos resultados de ações realizadas pelo manejo das bacias hidrográficas.

O índice de sustentabilidade tem como principal característica sintetizar o resultado dos indicadores de Hashimoto, Stedinger e Loucks (1982), indicando uma maior capacidade adaptativa e mais sustentável quando o sistema é avaliado positivamente pelo índice.

A Equação 4 apresenta o cálculo realizado para obtenção do índice (LOUCKS, 1997; CASTRO; MARQUES, 2019).

$$S = C * R * (1 - V) \quad (4)$$

Na equação supracitada, o índice de sustentabilidade -  $S$ , é calculado através do produto entre a confiabilidade, a resiliência e a parcela complementar da vulnerabilidade.

## 2.4 ÁRVORES DE DECISÃO

As árvores de decisão podem ser utilizadas em vários contextos inclusive ambientais e são geralmente entendidas como modelos que explicam as variações em uma determinada resposta a partir da divisão (*splits*) e agrupamento de variáveis independentes preditoras em certos limites.

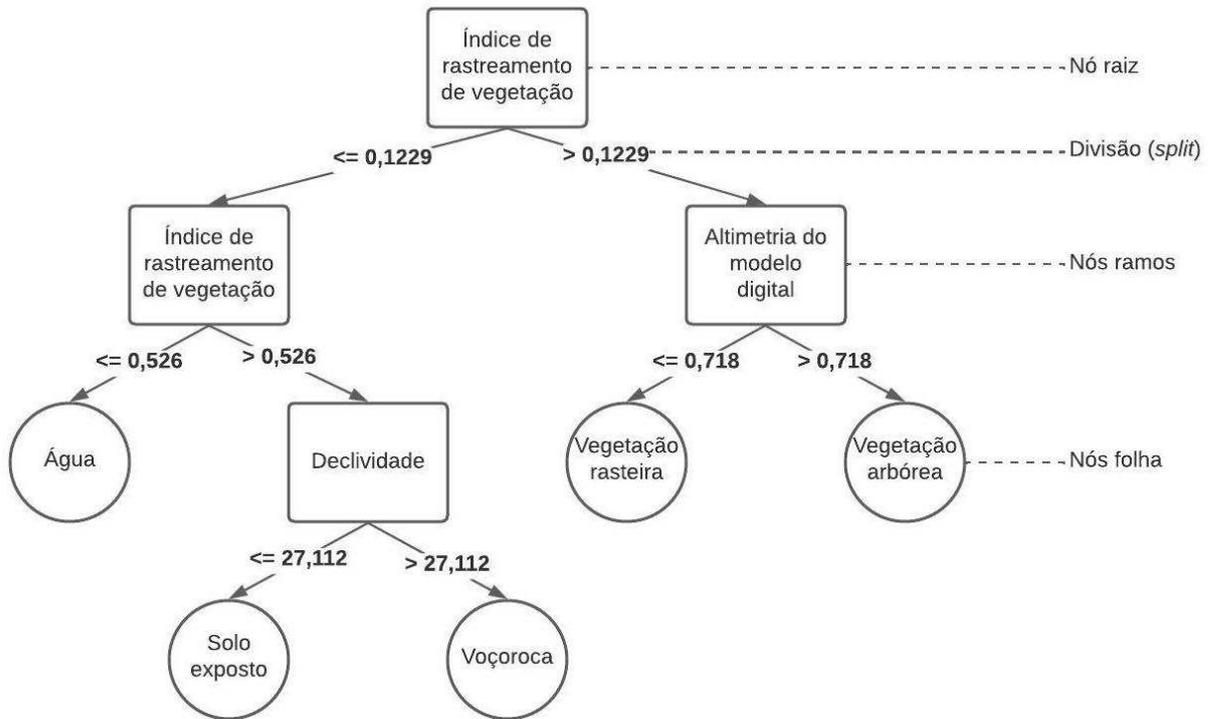
A aplicação do método resulta em uma resposta visual, que consiste em um fluxograma com formato de árvore, com estruturação do tipo *top-down*, através de ramificações condicionadas aos valores das variáveis, que são conhecidas a partir dos padrões identificados nos dados ou seja, a estrutura é montada através de um agrupamento inicial e ramificando a partir deste as demais *splits* (EVERAERT; BENNETSEN; GOETHALS, 2016).

O modelo de árvores de decisão, foi inicialmente criado como um modelo indutivo, mas que atualmente possui várias variantes. Entre estes modelos é um dos mais conhecidos pela comunidade científica e que faz parte do grupo de modelos de aprendizagem de máquina denominados como modelos classificatórios, além de outros como *K-means*, Teorema de Bayes, *Random Forest*, Redes Neurais etc.

Trata-se de uma ferramenta versátil e fácil de usar. Como o nome sugere, a árvore de decisão apresenta o conhecimento extraído usando uma estrutura semelhante a uma árvore que consiste em um nó raiz (contendo todo o conjunto de dados usado no modelo) e uma série de ramificações, que levam aos terminais folhas e que definem a relação entre as variáveis de entrada e o resultado desejado (MURPHY *et al.*, 2016). É um modelo hierárquico de suporte à decisão, composto de regras de decisão que dividem recursivamente as variáveis independentes em zonas homogêneas (LEE; LEE, 2015).

A Figura 1 ilustra um exemplo de árvore de decisão para classificação adaptado de Tedesco, Antunes e Oliani (2014), em que o modelo foi utilizado para a detecção de formação erosiva (voçoroca) por meio de classificação hierárquica tendo como base imagens de satélite.

**Figura 1** - Exemplo de modelo de árvore classificatória.



Fonte: Autor (2021)

O principal componente utilizado para geração do modelo é a entropia. Segundo Hannan e Anmala (2021), a entropia é uma variável que indica o grau de aleatoriedade em um determinado conjunto de variáveis de entrada.

A entropia  $Ent(S)$  é dada pela Equação 5,

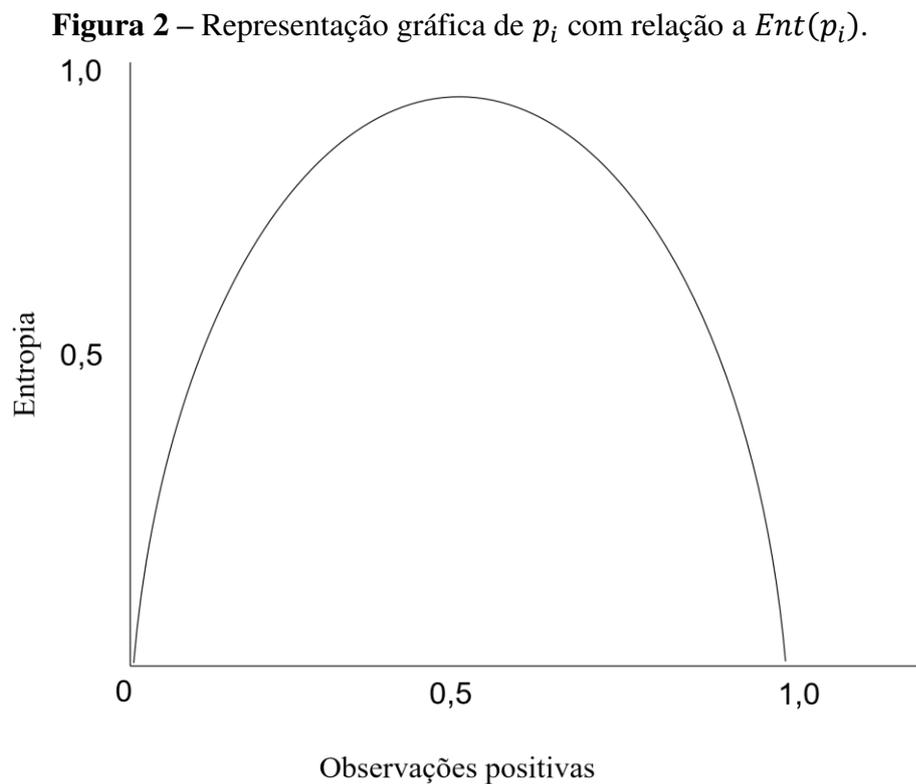
$$Ent(A) = -\sum_{i=1}^K p_i \log_2 p_i \quad (5)$$

onde  $A$  é o conjunto amostral;  $K$  é número de classes e  $p_i$  é a proporção de  $A$  que pertence à classe  $i$ , sendo  $p_i$ :

$$p_i = \frac{|A_i|}{|A|} \quad (6)$$

onde  $|A_i|$  é o número de amostras classificadas na  $i$ -ésima partição e  $|A|$  é o número total de amostras do conjunto  $A$ .

Conceitualmente, a  $Ent(A)$  geralmente é utilizada para medir a pureza ou impureza de um conjunto de dados associados a uma classificação. Um exemplo mais facilmente compreendido é a classificação de dados binários. A Figura 2 demonstra o comportamento da  $Ent(A)$  no eixo das ordenadas em relação a proporção de observações pertencentes a classe  $i$ .



No caso desse exemplo, quando o número de observações positivas é igual a zero, a entropia também será igual a zero, pois todo o conjunto não pertencerá a classe. Quando a proporção de observações é de 50% da amostra, temos entropia máxima, já que se tratam de dados binários e o maior grau de desordem possível. Assim a entropia tem seu máximo ( $\log_2 p_i$ ) quando  $p_i = p_j$ , sendo  $p_j$  a proporção de observações não pertencentes a classe  $j$ , para qualquer  $i \neq j$ . Caso  $p_i$  fosse = 1 ou seja, todas as observações pertencessem a classe  $i$ , então  $Ent(p_i) = 0$  (DETZEL; BESSA; MINE, 2013; FERNANDES, 2017).

O processo de divisão de um atributo continua até a entropia dos subconjuntos resultantes ser menor do que a entrada anterior ou conjunto de treinamento, eventualmente levando a nós de folha com entropia zero. A minimização da entropia é desejada, pois reduz o

número de regras das árvores de decisão. A redução da entropia dá viés para as árvores possuírem menos galhos. A entropia é definida como uma medida teórica da informação relacionada a “incerteza”, presente no conjunto de dados devido a várias classes (HANNAN; ANMALA, 2021).

Importante destacar que além da entropia existem outras formas de hierarquizar os parâmetros e definir as divisões dos grupos de dados como o valor do Índice de Gini. Segundo Santos e Maciel (2018) este índice criado pelo matemático italiano Conrado Gini, mede o grau de heterogeneidade dos dados, determinado para cada nó da árvore conforme a equação 7.

$$\text{Índice Gini} = 1 - \sum_{i=1}^K p_i^2 \quad (7)$$

Caso o resultado para o índice seja igual a zero, significa que o nó da árvore é considerado “puro” ou seja, os dados possuem relação totalmente pareada com a variável do nó, caso contrário, quando o resultado se aproxima de um é considerado “impuro”, o que possibilita particionar os dados conforme os graus de pureza (SANTOS; MACIEL, 2018).

As aplicações das árvores de decisão surgiram com o desenvolvimento do modelo ID3 por Quinlan (1979), baseado na dicotomização interativa, sendo posteriormente desenvolvido o modelo CART por Breiman *et al.* (1984), com a finalidade de produzir apenas árvores binárias (cada nó pode ter no máximo dois ramos). Após o CART, Quinlan (1993) desenvolveu o modelo C4.5, capaz de gerar árvores n-árias, sendo este considerado o mais avançados, porém todos atualmente servem de base para o surgimento de novos modelos variantes. Tanto o CART como o C4.5 utilizam estratégias parecidas: são recursivos, com uma abordagem *top-down* da estrutura e tem como meta a construção de árvores menores com máximo de acurácia possível (WILGES *et al.*, 2010; ANJOS *et al.*, 2017; SUQUINA, 2019).

O valor do critério de classificação, a exemplo da entropia, serve de base para o cálculo do ganho de informação, que é o atributo de entrada relacionado com as classes. A Equação 8 apresenta o ganho de informação como

$$G_{info} = Ent(A) - \sum_{j=1}^K \frac{|A_j|}{|A|} Ent(A_j) \quad (8)$$

onde  $m$  é o número de classes possíveis. Ao maximizar o ganho de informação, a entropia para um determinado parâmetro é minimizada (HANNAN; ANMALA, 2021).

O valor do ganho de informação pode ser interpretado como a capacidade que a variável independente consegue através dos dados de treinamento indicar uma determinada classificação, possuindo um efeito contrário ao que é observado para a entropia. Quanto maior a aleatoriedade em uma variável independente, menor será a o ganho de informação (FERNANDES, 2017).

Existem vários métodos para calibração e validação do modelo. A primeira é conhecida como "calibração aleatória", na qual a combinação de dados de treinamento e teste é escolhida arbitrariamente a partir de uma amostra e, em seguida, utilizada para calibração. Um segundo método é o "calibração estratificada", que envolve dividir os dados de treinamento em vários grupos distintos e depois calibrar o modelo usando cada grupo separadamente. A terceira técnica é conhecida como "validação cruzada", que consiste no particionamento do conjunto de dados em duas subamostras, de forma que uma subamostra é utilizada para a estimação dos parâmetros do modelo (calibração ou ajuste) e a outra de tamanho é empregada na validação do modelo (AGUIAR, 2013).

Entre as principais vantagens do uso das árvores de decisão estão a acurácia semelhante à de outros modelos mais complexos, uma semântica praticamente intuitiva, de fácil compreensão, robusta a ruídos, além de possuir alta compressibilidade e aplicabilidade (VELOSO *et al.*, 2021).

Simultaneamente com a representação visual dos resultados, que permite ao tomador de decisão identificar a relação e as interdependências de cada decisão e formular uma previsão apropriada, o modelo também permite abordagens computacionais capazes de capturar a relação não linear nos dados (SYACHRANI; JEONG; CHUNG, 2013; EVERAERT; BENNETSEN; GOETHALS, 2016).

Existem algumas desvantagens com relação ao uso das árvores. A maioria dos pontos negativos estão ligados ao seu crescimento e profundidade que faz com que as árvores tenham que passar por um pós-tratamento, procedimento comumente denominado como "poda".

A própria poda traz consigo benefícios e prejuízos para a definição do modelo como a possibilidade de tornar mais amigável à interpretação, positivamente visto como uma forma de facilitar sua aplicação prática, porém, se feita de forma inadequada, prejudica o desempenho do modelo, levando a obtenção de resultados falsos positivos, entre o que é previsto e o que é observado. Outro fator que dificulta o processo de poda é a determinação do quanto deve ser podado, pois uma poda branda reflete no sobreajuste dos dados e na consideração de ruídos na formação dos nós folha (FERNANDES, 2017; SUQUINA, 2019).

Além disso a definição prévia do parâmetro a ser utilizado na raiz da árvore ou o desbalanceamento exagerado dos dados pode levar a resultados diferentes para as regras e decisões, caso sejam feitas variações nesses aspectos.

Para correção desses problemas, foi criado um modelo mais diretamente ligado as árvores, denominado de *Random Forest* ou seja, Floresta Aleatória. Essa outra abordagem considera a geração de várias árvores de decisão que variam parâmetros e critérios das árvores randomicamente, em que a previsão da classificação é confirmada pelos resultados da maioria das árvores (DOTA; CUGNASCA; BARBOSA, 2015; ANJOS *et al.*, 2017).

O método favorece uma maior performance na previsão, porém releva o aspecto visual das árvores de decisão única, algo que deve ser levado em consideração quando o modelo tem a intenção de servir como suporte para a tomada de decisão por outros profissionais da gestão que não estejam preocupados com a base matemática por trás da resposta fornecida pelos modelos.

Dentre as diversas aplicações na engenharia e em outras ciências, as árvores de decisão também têm sido aplicadas para melhoria da situação dos recursos hídricos no geral. Sendo assim, é possível destacar alguns trabalhos realizados na área em suas diversas frentes de estudo.

Lee e Lee (2015) realizaram a aplicação do modelo de árvore de decisão para análise do mapeamento potencial da produtividade das águas subterrâneas, em que foram relacionados dados de produtividade de águas subterrâneas com fatores hidrogeológicos, topográficos, geológicos, de vegetação e de solo, coletados espacialmente. Os resultados indicaram um alto potencial de produtividade de águas subterrâneas nas diversas cidades usadas como objetos de estudo obtendo acurácia de 79,47% até 96,55%, indicando a utilidade das árvores para os recursos hídricos subterrâneos.

Outro estudo importante envolvendo as árvores de decisão e os recursos hídricos foi realizado por DOTA, Cugnasca e Barbosa (2015), que comparou algoritmos de árvore de decisão na classificação da qualidade da água contaminada pelo solo. Os dados utilizados para aplicação dos algoritmos foram coletados a partir de uma amostra de água contaminada do solo, possibilitando a criação de modelos de classificação para identificar a qualidade da água. A classificação proposta pelos modelos indicou forte relação estatística entre as classes que qualificam a amostra e os valores que descreveram cada classe. O estudo demonstrou a versatilidade das Árvores de decisão na integração entre o aprendizado de máquina e a análise laboratorial.

Também relacionado aos recursos hídricos, Syachrani, Jeong e Chung (2013) desenvolveram modelo de deterioração para tubulações de esgotamento sanitário. O modelo buscou permitir que os gestores pudessem identificar e formular decisões no gerenciamento do esgotamento com relação a deterioração das tubulações sendo este um problema complexo de infraestrutura civil que depende muito da qualidade dos dados de entrada e da técnica computacional utilizada. O desempenho da árvore de decisão superou outras técnicas como a regressão e as redes neurais geradas com o mesmo conjunto de dados do estudo.

Hannan e Anmala (2021) utilizaram árvores de decisão para classificação em previsão de coliformes fecais em córregos da bacia hidrográfica do Upper Green River, em Kentucky nos Estados Unidos, mostrando a possibilidade de uso das árvores no monitoramento da qualidade da água. Neste estudo foram utilizados vários tipos de árvores, como CART, ID3, Floresta Aleatória, entre outros. Os vários modelos utilizados conseguiram realizar classificação e a previsão com êxito, sendo as árvores de decisão tradicionais as que obtiveram melhor precisão.

Com relação ao abastecimento de água para consumo humano, Murphy *et al.* (2016) utilizaram as árvores de decisão para investigar potencial de aplicação de técnicas de aprendizado de máquina para identificação de fatores importantes sobre alertas de água potável em pequenos sistemas de água no Canadá. O estudo evidenciou a relação entre o nível de instrução dos Funcionários das concessionárias, a localização geográfica dos sistemas, o de fonte de captação e o nível de tratamento da água com os alertas de água potável, sendo modelo capaz de prever corretamente em 79% dos casos. Os autores destacaram facilidade de uso e a alta capacidade de suporte das árvores para os gestores e agências reguladoras de água.

O outro estudo realizado por Lu e Ma (2020), também mostrou a importância das árvores de decisão para a previsão de variáveis. O estudo foi realizado visando o desenvolvimento de um modelo híbrido de árvore de decisão para prever a qualidade da água em curto prazo em rios. Estudo considerou a utilização da Floresta randômica e do *Extreme Gradient Boosting (XGBoost)*, técnicas avançadas de eliminação de ruído. Pesquisa foi aplicada além disso a pesquisa foi aplicada na bacia do Rio Tualatin, um dos rios mais poluídos do mundo. Ângulos os modelos foram usados para prever valores de indicadores de qualidade como temperatura oxigênio dissolvido valor de PH e turbidez, sendo comparados entre si e apresentando bom desempenho na previsão.

Já com relação operação de reservatórios, Zhao e Cai (2020) desenvolveram regras de operação através da combinação das árvores de decisão com cadeias de Markov, com o intuito de regular o escoamento de água para o atendimento as demandas dos sistemas de reservatórios.

o modelo desenvolvido levou em consideração diversos fatores como as afluências, volumes de reservatório e a dinâmica de operação, possibilitando a criação de módulos de operação. O estudo considerou um total de 50 reservatórios da Bacia de Upper Colorado River nos Estados Unidos, sendo o modelo validado com 11 reservatórios da mesma região. Os resultados demonstraram a possibilidade de transição padronizada entre os módulos das regras de operação, melhorando operações futuras, planejamento para construção de componentes e compreensão dos fenômenos da bacia.

Por fim fica evidente a utilidade e versatilidade da aplicação das árvores de decisão, principalmente nos recursos hídricos em suas mais variadas áreas, havendo apenas a necessidade de dados suficientes para o treinamento dos modelos. A aplicação nos recursos hídricos e aliada a análise de riscos também pode ser considerada como uma área promissora, devido a quantidade de técnicas e abordagens viáveis a depender do problema a ser resolvido.

### 3 METODOLOGIA

A revisão bibliográfica feita neste trabalho mostrou a importância dos estudos envolvendo a análise de riscos e como a modelagem pode ser utilizada na melhoria da análise dos sistemas de recursos hídricos.

Assim nesta seção serão detalhados os procedimentos e formulações matemáticas bem como os métodos utilizados no desenvolvimento de um modelo de análise dos riscos de não atendimento as demandas de um sistema de reservatórios superficiais, com enfoque na previsão do risco do sistema, com base nos padrões e comportamentos de fatores hidrológicos envolvendo os processos estocásticos. Têm por finalidade, identificar o risco do sistema para os diferentes cenários hidrológicos, através da simulação da alocação de água em atendimento aos elementos de demanda com seus múltiplos usos.

#### 3.1 CLASSIFICAÇÃO DA PESQUISA

Quanto a metodologia, esta pesquisa é classificada como Método Hipotético-Dedutivo. Para Pereira *et al.* (2018), o método se inicia pela percepção da possibilidade de aprimoramento dos conhecimentos já existentes, sendo formuladas hipóteses e, pelo processo de inferência dedutiva, são testadas as predições de ocorrência de fenômenos abrangidos por estas, gerando um novo conhecimento, que passa pelo mesmo processo de avaliação de lacunas, formulação de hipóteses, falseadas ou verificadas à medida que são validadas.

No que diz respeito à natureza da pesquisa, esta é classificada como sendo do tipo pesquisa aplicada. Gil (2007 p. 27) apresenta as pesquisas com essa característica como sendo “voltadas à aquisição de conhecimentos com vistas à aplicação numa situação específica”. Pode ser entendida como aquela que tem como foco a aplicação e utilização do conhecimento, com resultados servindo como construção de um novo conhecimento utilitário.

O objetivo do estudo é classificado como exploratório, baseado na hipótese de integração de várias técnicas sem a perspectiva de resultados totalmente previstos. Para Gil (2007), a pesquisa exploratória deve tornar o problema mais explícito, adquirindo maior familiaridade ou construir hipóteses à solução, envolvendo a investigação das experiências prévias, análises e o estado da arte do tema (GIL, 2007; SILVA, 2015).

Quanto aos procedimentos é classificada como estudo de caso, com o propósito de aprofundar os conhecimentos sobre as causas e comportamento do elemento ou sistema

estudado. Segundo Pereira *et al.* (2018), o intuito não é intervir no objeto a ser estudado, mas entender de modo interpretativo sobre a compreensão do problema, de modo pragmático, apresentando a perspectiva global do objeto de estudo, de forma coerente e completa.

### 3.2 ABORDAGEM METODOLÓGICA E INOVAÇÃO CIENTÍFICA

A abordagem utilizada foi denominada de *Risk Analysis Model by Stochastic Simulation* – RAMSS – ou seja, um modelo de análise de riscos por meio dos dados obtidos através da simulação estocástica das redes de reservatórios, que tem como foco principal o apoio ao suporte à decisão na previsão do risco de não atendimento às demandas do sistema.

O RAMSS buscou trazer inovações tanto para a análise de riscos quanto para as árvores de decisão, assim como também contribuir para melhoria na gestão dos recursos hídricos, complementando e trazendo aprimoramentos à trabalhos anteriores.

As principais inovações deste trabalho são a utilização de árvores de decisão para a análise de risco em recursos hídricos, bem como a possibilidade de previsão do diagnóstico de risco, associado ao nível da confiabilidade do sistema e o fornecimento das vazões requeridas pelos elementos de demanda, juntamente com a caracterização do panorama de atendimento do sistema para as classes de risco, que são as variáveis dependentes do modelo.

Além desses fatores, a abordagem metodológica foi direcionada ao uso de dados de entrada sintéticos para as variáveis independentes, com a finalidade de investigar as diversas possibilidades de cenários de atendimento às demandas.

A partir da consideração da estocasticidade nos sistemas de recursos hídricos com a geração de séries sintéticas para simulação, é possível investigar as diversas alternativas de resposta para a situação hídrica dos sistemas, para as possibilidades de comportamento de variáveis como a precipitação.

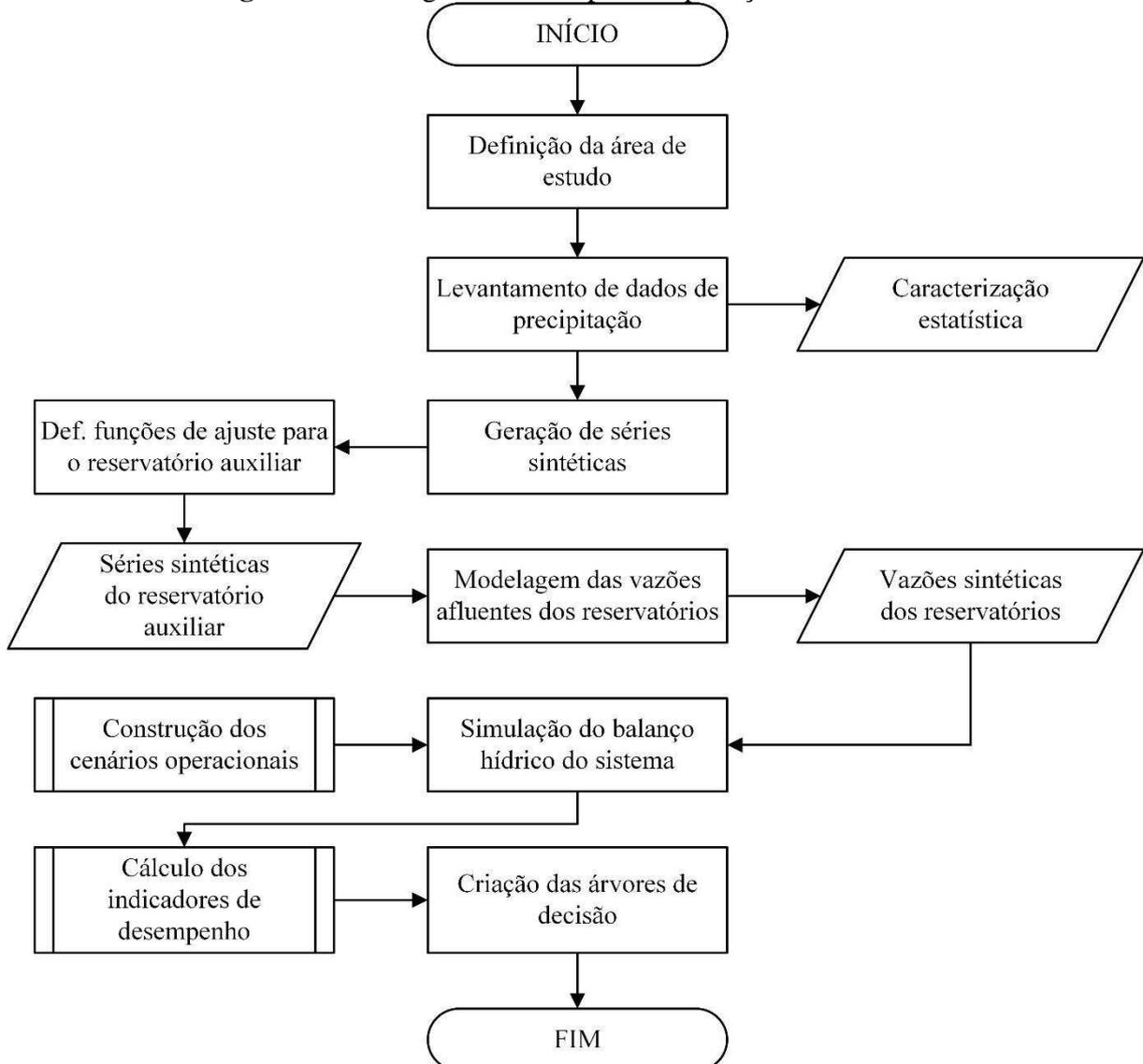
O estudo levou em consideração a construção de etapas que levassem ao atendimento dos objetivos, visando atuar de forma que a tomada de decisão realizada por parte dos gestores pudesse contemplar todos os principais cenários, de análise indispensável para o planejamento e execução das ações estratégicas para os sistemas de recursos hídricos.

O modelo proposto poderá ser utilizado para fornecer informações à geração, avaliação e sugestão de alternativas de operação das redes de reservatórios, para que os comitês de bacia e os usuários possam tomar as decisões de maneira especializada, nos projetos de intervenção e na operação dos sistemas. Devido à quantidade de dados utilizados, gerados e do

processamento das operações matemáticas, foram utilizadas ferramentas computacionais de análise de dados e linguagens de programação, tais como: LabFIT; LabSid Acquanet e *Python*.

A Figura 3 apresenta a sequência das etapas e procedimentos realizados, organizados e sintetizados, contendo em formato de fluxograma o processo metodológico do RAMSS. Os detalhes do procedimento são apresentados na próxima seção.

**Figura 3** – Fluxograma das etapas de aplicação do RAMSS.

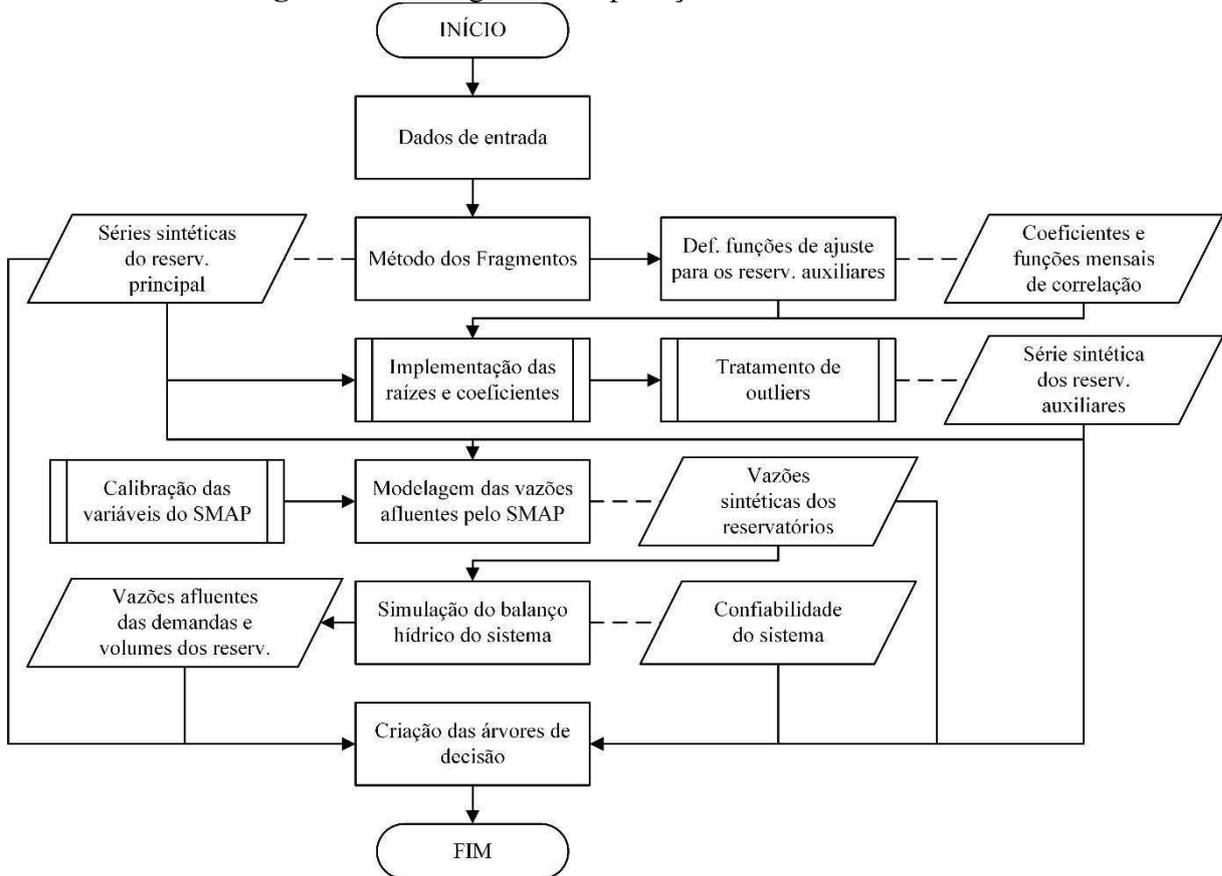


Fonte: Autor (2022).

### 3.3 ETAPAS DO RAMSS

A construção do RAMSS se deu a partir da integração de outros modelos. A Figura 4 apresenta o fluxograma. Os detalhes da aplicação dos processos apresentados pelo fluxograma estão descritos na próxima seção.

**Figura 4** – Fluxograma de aplicação do modelo RAMSS.



Fonte: Autor (2022).

#### 3.3.1 Etapa I – Preparação dos dados iniciais de precipitação

Para aplicação do RAMSS inicialmente foi definida a variável hidroambiental estocástica que foi a precipitação, devido ao caráter estocástico e a importância que possui para o entendimento do panorama hídrico dos sistemas de reservatórios. Foram feitos então o levantamento e preparação dos dados de entrada, para a construção das Séries Históricas – SH's de referência.

A escolha direcionou o levantamento das informações de bancos de dados como o Instituto Nacional de Meteorologia – INMET – e a Agência Executiva de Gestão das Águas do Estado da Paraíba – AESA – Estado em que a pesquisa foi aplicada. A coleta e tratamento dos dados resultou nas SH dos reservatórios. Os dados para cada um dos postos do INMET e AESA, foram agrupados com o intuito de obter a maior série temporal possível, favorecendo um maior nível de confiança e grau de liberdade estatística.

Após o levantamento dos dados de precipitação, foi realizado o preenchimento de falhas para as séries históricas dos postos pluviométricos escolhidos, com base no método de ponderação regional integrado a regressão linear, o estabelecimento de regressões lineares entre os postos com dados a serem preenchidos e cada um dos postos vizinhos (PRUSKI *et al.*, 2004; OLIVEIRA *et al.*, 2010).

Com o preenchimento de falhas, foram feitas análises de consistências para os postos utilizando o método da dupla massa. Posteriormente, foi aplicado o método de Thiessen (1911), para encontrar a precipitação média da bacia. Dessa forma, foram obtidas as séries temporais de dados históricos dos reservatórios estudados, em escala mensal, resultando nas SH's utilizadas nas etapas seguintes, conforme mostradas nas Tabelas A1 e A2 (Anexo).

### **3.3.2 Etapa II – Análise estatística e geração sintética**

A partir das SH's, foram calculadas as propriedades estatísticas, que serviram de base para a análise dos dados, bem como variáveis utilizadas pelos modelos aplicados no RAMSS.

Os comportamentos das SH's foram analisados através de gráficos *boxplot*, histogramas mensais e as distribuições de quartis. Além disso, foram realizados o teste de normalidade de Shapiro-Wilk e o teste não paramétrico de Mann-Whitney, para verificação da diferença estatística entre as médias das SH's e, com isso, avaliar a necessidade de utilizar uma ou mais SH's na geração sintética dos cenários de chuva.

Juntamente com a estatística descritiva, as SH's também foram decompostas para análise dos comportamentos de tendência e sazonalidade. Foram verificadas as autocorrelações e estacionariedade através de testes estatísticos como Ljung-box – LB, Kwiatkowski-Phillips-Schmidt-Shin – KPSS e Dickey-Fuller – DF (LJUNG e BOX, 1978; KWIATKOWSKI *et al.*, 1992; DICKEY e FULLER, 1979).

### 3.3.3 Etapa III – Geração sintética das séries de precipitação

Após a análise estatística foi então realizada a geração de séries sintéticas com o propósito de observar as variações hidroclimáticas possíveis no sistema de reservatórios, a partir de condições meteorológicas representadas pela precipitação, juntamente com variações no escoamento superficial, representadas pela vazão afluente aos reservatórios.

Dentre os modelos analisados para a geração sintética, o Método dos Fragmentos – MF foi o escolhido. O total de séries sintéticas geradas foi de 10 mil com o intuito de representar a estocasticidade no sistema e a maior quantidade viável de possibilidades de chuva, que ao serem implementadas nos processos subsequentes, resultaram na análise de diversos cenários distintos para verificação dos riscos associados.

Para o MF, foram gerados os fragmentos da SH do reservatório principal, por meio do cálculo das frações mensais, calculadas a partir da razão entre a precipitação de cada mês, sobre a precipitação anual.

Os valores das classes para uso dos fragmentos encontrados foram organizados em ordem crescente, resultando em vetores de limites de classes de desagregação, onde os valores da variável aleatória independente, representando uma precipitação anual hipotética, calculada repetidamente e obtida por meio da geração randomizada, é enquadrada em uma das classes de fragmentos e assim multiplicada pela fração de cada mês pertencente a classe.

Os limites foram organizados, sendo o limite inferior da primeira classe igual a zero, seguindo do primeiro limite superior, igual a média da primeira e segunda precipitação anual, que por sua vez foi o limite inferior da segunda classe, com o procedimento se repetindo até a última classe. Os fragmentos foram gerados para a SH e as precipitações sintéticas anuais foram geradas segundo os números aleatórios, a partir da função de densidade de probabilidade seguindo a distribuição Gamma, conforme realizado por Celeste, Curi e Curi (2007).

Para validação dos valores aleatórios gerados, foi realizada uma análise de independência conforme Reddy (1987), sendo elaborado um correlograma amostral, com os limites de tolerância para processos independentes para o nível de 5%, calculados segundo a equação 15,

$$\rho_{\ell}(5\%) = \frac{-1 \pm 1,645\sqrt{n-\ell-1}}{n-\ell} \quad (15)$$

onde  $\rho_\ell$  é o coeficiente de correlação correspondendo ao *lag*  $\ell$  para a amostra de tamanho  $n$ . Conforme sugerido por Reddy (1987 *apud* Celeste; Curi e Curi, 2007), o número máximo de *lags* para os quais os coeficientes devem ser calculados é de  $\frac{1}{4}$  do tamanho da amostra.

A avaliação dos modelos também foi realizada por meio das métricas de desempenho: *Mean Error* – ME; *Mean Absolute Error* – MAE; *Root Mean Standard Error* – RMSE e *Mean Absolute Percentage Error* – MAPE, a fim de verificar quais dos modelos tinham melhor ajuste entre o observado e o modelado. A Tabela 1 apresenta as equações 16, 17, 18 e 19, utilizadas para os cálculos das métricas de desempenho.

**Tabela 1** – Resumo das equações dos modelos de desempenho aplicadas a série temporal.

Métrica de desempenho	Equação
ME	$\frac{1}{n} \sum_{i=1}^n (\hat{x}_i - x_i)$ (16)
MAE	$\frac{1}{n} \sum_{i=1}^n  \hat{x}_i - x_i $ (17)
RMSE	$\sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{x}_i - x_i)^2}$ (18)
MAPE	$\frac{\sum_{i=1}^n \frac{ x_n - \hat{x}_i }{n}}{\sum_{i=1}^n \frac{ x_n - x_{n-1} }{(n-1)}}$ (19)

Fonte: Autor (2022).

O valor de  $n$  nas equações das métricas representa o tamanho da amostra. Os valores de  $\hat{x}_i$  são os dados observados para o período de estudo com  $n$  observações, enquanto  $x_i$  são os dados obtidos a partir da aplicação dos modelos. Em resumo, a avaliação é feita através da diferença entre o gerado e o observado para a série de referência.

### 3.3.4 Etapa IV – Ajustes de correlação entre reservatórios

Em posse das séries pluviométricas sintéticas do reservatório principal, foi observada a necessidade de correlacionar os valores de precipitação obtidos entre as SH's, pois ao realizar a modelagem de séries sintéticas para mais de um reservatório nos sistemas, exceto não fosse utilizada a geração multivariada, não haveria a possibilidade manter a similaridade temporal e intensidade de precipitação mensal entre os reservatórios.

Tendo em vista a praticidade na aplicação e o uso da metodologia em sistemas de reservatórios com diferentes estruturas topológicas, foram encontradas as equações de correlação ajustadas para cada mês da série. A escolha da melhor função de correlação se deu

a partir do cruzamento dos dados históricos mensais dos reservatórios, que foram aplicados ao *software* LabFIT, tendo como principal utilidade a varredura e análise dos dados cruzados para a determinação das equações de melhor aproximação do comportamento entre os meses, escolhidas a partir do banco de equações presente no *software*.

O LabFIT listou as equações com os valores dos coeficientes de determinação, o coeficiente Qui-quadrado, o coeficiente Qui-quadrado crítico e o grau de liberdade para cada mês, que estavam mais adequados para os conjuntos de dados. Assim, as equações utilizadas levaram em consideração a obtenção de dados não negativos, com correlação de precipitações proporcional e o mínimo do surgimento de *outliers*. A Tabela A3 (Anexo) apresenta as equações que melhor se ajustaram aos dados de precipitação para cada mês obtidas pelo LabFIT, com seus respectivos parâmetros, coeficientes de determinação e o número de iterações realizadas pelo *software*. As variáveis X e Y da Tabela A3 são os dados históricos dos reservatórios agrupados por mês.

Dessa forma as séries sintéticas produzidas a partir dos ajustes de correlação ainda necessitaram passar pelo tratamento de *outliers*, através da técnica de Winsorização, que consiste em definir todos os valores discrepantes para um percentual especificado dos dados, sendo o percentual de corte utilizado os valores em 1% da amostra, retirando assim os valores extremos de precipitação que poderiam causar erros na obtenção das vazões geradas a partir dos dados de precipitação. Segundo Duarte, Girão e Paulo (2017), estudos mostram que a Winsorização apresenta uma robustez considerável se comparada a grande parte dos métodos, além da fácil aplicação.

### **3.3.5 Etapa V – Modelagem hidrológica das vazões**

Após esses procedimentos foi realizada a modelagem das vazões afluentes aos reservatórios. Foi utilizado o modelo hidrológico SMAP com o objetivo de determinar as vazões mensais de entrada a serem aplicadas na simulação do sistema de estudo. Além dos dados pluviométricos, foram utilizados no SMAP dados gerais da área de estudo (Tabela A5) como: a área de drenagem -  $Ad$ ; taxa de umidade -  $Tu$ ; escoamento de base -  $Eb$ ; capacidade de campo -  $Cc$  e Abstração inicial -  $Ai$ , além da utilização de parâmetros específicos para a região em que se encontra a área de estudo. Outro ponto importante para aplicação do modelo foi a calibração, realizada através da definição dos parâmetros a partir do posto de referência.

Os parâmetros utilizados para a aplicação do modelo foram: Capacidade de saturação do solo - *Sat*; Coeficiente de Recarga do Aquífero - *Crec*; Taxa de geração de escoamento superficial - *Pes*; Taxa de depleção do nível da água do aquífero subterrâneo - *k* (Tabela A6).

De posse dos dados de entrada necessários ao modelo SMAP, foram inseridas as séries pluviométricas mensais dos reservatórios, o que correspondeu a geração das 10 mil séries de vazões, cada uma contendo 10 anos em intervalos mensais.

### **3.3.6 Etapa VI – Simulação de balanço hídrico do sistema**

Através da reunião de dados e informações sobre a estrutura topológica da área de estudo, foi utilizado o *software LabSID Acquanet*, para realização da simulação do balanço hídrico. Antes da execução das simulações, foram primeiramente definidos e elaborados os cenários operacionais – CO do sistema. A elaboração dos cenários contemplou as variações das situações ocorridas em pontos estratégicos da topologia, a partir das condições de funcionamento do sistema de reservatórios estudado.

Tendo sido selecionados os cenários operacionais, foram então criados os arquivos base para a simulação no *software*, sendo estes previamente configurados para manterem suas características hidrológicas idênticas em todas as séries sintéticas de vazão implementadas, onde cada arquivo de cada cenário teria as vazões naturais diferentes para os reservatórios.

Para cada simulação foram definidos alguns parâmetros dos reservatórios como: volume inicial em  $\text{hm}^3$ ; volumes máximos e mínimo; taxa de evaporação em  $\text{m}^3/\text{mês}$ ; a tabela com os dados da curva cota-área-volume; as vazões naturais em  $\text{m}^3/\text{s}$  (dados sintéticos). Além desses elementos, as vazões e prioridades de atendimento das demandas urbanas e rurais também foram definidas, sendo as demandas urbanas de prioridade máxima, seguida das demandas rurais.

Com a definição dos cenários operacionais, foram executadas as 10 mil simulações quantitativas para os cenários, fornecendo os dados dos balanços hídricos dos reservatórios e as curvas de vazão fornecida para as demandas.

### **3.3.7 Etapa VII – Cálculo dos indicadores de desempenho**

Após a obtenção dos dados de cada cenário de simulação, foram calculados os valores dos indicadores de Confiabilidade –  $C$ , Resiliência –  $R$  e Vulnerabilidade –  $V$ . Os valores dos indicadores foram obtidos para cada demanda do sistema simulado.

Também devido a elevada quantidade de resultados das simulações, os indicadores foram avaliados através dos valores médios para cada grupo de dados, sendo também avaliadas as estatísticas descritivas e a distribuição dos dados para as simulações de cada CO.

As correlações existentes entre os dados foram analisadas através de gráficos de dispersão, matrizes de correlação e testes estatísticos não paramétricos como *Mann-Whitney* e o teste de correlação de *Spearman*, com a finalidade de identificar padrões e semelhanças entre os valores de  $C$ ,  $R$  e  $V$ , com vistas a estimação indireta dos indicadores por meio da modelagem matemática, tornando desnecessária a associação dos dados climáticos para todos os indicadores.

Apesar da análise conjunta dos indicadores ser importante para a análise de riscos nos recursos hídricos, o uso dos parâmetros  $C$ ,  $R$  e  $V$  para a avaliação comparativa entre os eventos sintéticos de cada CO, poderia ocasionar na categorização equivocada de cenários com aspectos hidrológicos distintos, em uma mesma classe ou grupo.

Tal observação levou a escolha do indicador  $C$  como base para avaliação da associação entre o risco e as características hidrológicas das simulações sintéticas, devido ao baixo nível de complexidade exigido à sua interpretação e devido ao caráter exploratório do estudo ao ser feita a aplicação das árvores de decisão com os indicadores, evitando assim a perda no desempenho dos modelos.

### **3.3.8 Etapa VIII - Modelagem das árvores de decisão**

Após obtidos todos os valores de  $C$ ,  $R$  e  $V$  para as demandas de cada simulação, se iniciou o processo de modelagem das árvores de decisão para classificação dos riscos, de acordo com o comportamento das variáveis explanatórias que caracterizaram a situação hidrológica dos cenários simulados.

Para cada cenário foram criadas duas árvores de decisão, sendo uma para avaliar o efeito das vazões afluentes às demandas de abastecimento urbano e outra para avaliar o efeito das vazões afluentes às demandas de irrigação.

Assim, foram geradas cinco classes com faixas de porcentagem de  $C$ , em que para cada faixa, as árvores de decisão buscaram definir o padrão de comportamento das demandas,

sendo as faixas escolhidas visando o auxílio aos gestores e decisores, na operação dos sistemas de recursos hídricos (Tabela 2).

**Tabela 2** – Classes de situação de risco ( $Cer$ ) para as confiabilidades médias

Uso	Situação	Confiabilidade média anual
Abastecimento urbano	Ótimo	Igual a 100%
	Bom	Entre 95% e 99,99%
	Aceitável	Entre 70% e 94,99%
	Ruim	Entre 50% e 69,99%
	Péssimo	Abaixo de 50%
Irrigação	Ótimo	Igual a 100%
	Bom	Entre 90% e 99,99%
	Aceitável	Entre 70% e 89,99%
	Ruim	Entre 50% e 69,99%
	Péssimo	Abaixo de 50%

Fonte: Autor (2022).

A diferença entre as classes de risco ocorreu apenas na classe “Bom”. Para ambos os usos, a média de  $C$  anual da classe “Péssimo” foi definida como sendo abaixo de 50%, haja vista que neste caso metade do sistema já estaria em falha, representando um cenário de crise hídrica.

Por não haverem critérios pré estabelecidos e obrigatórios atualmente na gestão dos recursos hídricos, as faixas de confiabilidade para as classes foram definidas visando a melhor distribuição possível dos grupos de simulações classificadas, para possibilitar existência de dados suficientes para o treinamento das árvores. Vale salientar que as faixas de porcentagem para as classes também podem ser flexibilizadas de acordo com as necessidades de análise para diferentes sistemas.

Após esse procedimento, são criados os bancos de dados com as variáveis dependentes e independentes. Os conjuntos de dados são compostos por:

- $Cer_{(t)}$  – Classe da situação do risco (variável dependente discreta);
- $P_{(t)}$  – Precipitação média no ano  $t$ ;
- Vazões médias afluentes das demandas no ano  $t$ ;
- $Vfin_{(t-1)}$  – Volume final simulado no ano  $t-1$ .

Estas variáveis foram escolhidas tendo em vista a análise da relação entre riscos com o comportamento da contribuição hídrica ocasionada pela pluviometria e a alocação do sistema, sendo estas variáveis importantes para a execução da simulação.

Com os dados já tratados, é então feita a separação dos dados das simulações em 75% para treinamento dos modelos e 25% para teste em cada um dos CO, sendo estes escolhidos randomicamente.

As árvores foram construídas em estrutura *top-down* a partir da definição do nó raiz, sendo feito o particionamento dos dados em subconjuntos. Para isso, foram calculados os valores de entropia para assim verificar a homogeneidade das amostras, sendo a entropia igual a zero quando a amostra é completamente homogênea e igual a um quando a amostra é dividida igualmente.

Primeiramente foi calculada a entropia da variável dependente (*Cer*). Logo após o cálculo, os conjuntos de dados foram divididos pelas diferentes variáveis independentes e calculadas as entropias correlacionadas as classes da *Cer*. Assim, foi calculado o ganho de informação a partir da subtração da entropia da variável dependente pela entropia das divisões das amostras correlacionadas. A variável que proporciona um maior ganho de informação foi aquela definida como nó de decisão.

Estabelecido o nó de decisão, os conjuntos de dados foram divididos, e o cálculo das entropias foi realizado recursivamente para as ramificações criadas, até o limite de homogeneidade mínima ou com entropia zero, até que todos os dados fossem classificados.

Dessa forma, foram geradas as árvores de decisão, levando em consideração os critérios de entropia e o ganho de informação para cada conjuntos de simulações. Os estimadores utilizados e ajustados para a modelagem das árvores foram:

- a profundidade máxima das ramificações da árvore;
- número máximo de variáveis avaliadas por *split*;
- o número mínimo de amostras por *split*;
- o número mínimo de amostras por folha.

A profundidade máxima das ramificações e o número máximo de variáveis avaliadas por *split* estabelecidos para execução do modelo foram baseadas no número de elementos de demanda de cada CO, sendo um nível de profundidade para cada variável considerada ou seja, em um CO possuindo 8 variáveis, a profundidade máxima seria 8 níveis de classificação. A

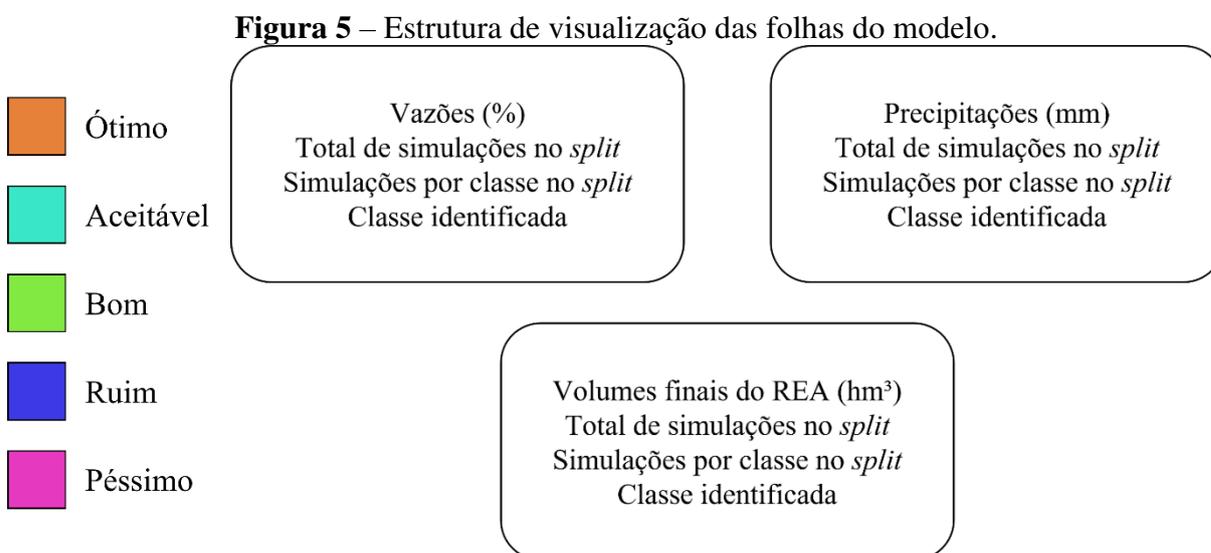
dependem da entropia das variáveis, que pode ser baixa ou nenhuma, nem todas as variáveis estariam presentes na árvore como critério de classificação do risco, pois a avaliação do valor da variável não interferiria no risco.

O número mínimo de amostras por *split* e o número mínimo de amostras por folhas foi definido considerando a representatividade mínima de 1.000 simulações, correspondendo a 1% do total de anos simulados em cada CO, haja vista que a classificação foi realizada sem que as quantidades de amostras por classes fossem balanceadas, com o intuito de preservar as características sintéticas iniciais.

As árvores foram avaliadas por meio da verificação dos coeficientes de acurácia, precisão e revogação para os modelos. Os valores dos estimadores foram otimizados através da validação cruzada da matriz de parâmetros estimados. Também foram feitos testes *T-Student* com a finalidade de avaliar as árvores de decisão com profundidades diferentes, comparando a acurácia das árvores com profundidades variando de 1 até 30 níveis, sendo verificada a diferença significativa entre o valor otimizado e a média dos coeficientes.

Para as árvores de decisão escolhidas foram realizadas validações cruzadas, selecionando grupos aleatórios de treino e teste a partir da amostra, com 10 partições por rodada de validação, verificando assim a acurácia das aplicações, sendo estas analisadas a partir da verificação da normalidade entre os coeficientes da validação.

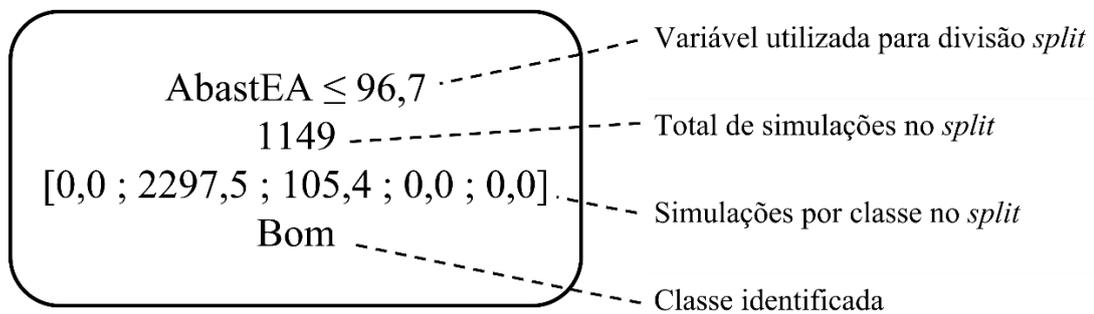
Na visualização das árvores de decisão, as folhas das ramificações com os subconjuntos de dados foram representadas apresentando a estrutura mostrada na Figura 5 juntamente com a coloração para identificação das *Cer*.



Fonte: Autor (2022).

A Figura 6 traz um exemplo de folha da árvore de decisão. A primeira informação trazida na folha é o intervalo de divisão da variável observada, em porcentagem de vazão fornecida, em hm<sup>3</sup> de volume de reservatório ou em milímetros de precipitação. Logo abaixo é mostrado o total de simulações agrupadas no subconjunto após a definição do intervalo. Em seguida é mostrada a quantidade de simulações em cada classe após o *split* do intervalo, ordenados da classe “Ótimo” até a “Péssimo”. Por fim é mostrada a classe correspondente para o subconjunto.

**Figura 6** – Exemplo de folha da árvore de decisão.



Fonte: Autor (2022).

### 3.4 ESTUDO DE CASO

#### 3.4.1 Bacia Hidrográfica do Rio Piancó-Piranhas-Açu

Foi selecionada a Bacia Hidrográfica do Rio Piancó-Piranhas-Açu – BHRPPA – na Paraíba, como principal região de estudo, mais especificamente a Unidade de Planejamento Hidrológico do Alto Curso do Rio Piranhas – UPHAP – e a Unidade de Planejamento Hidrológico do Médio Curso do Rio Piranhas Paraibano – UPHMP – sendo dada maior evidência ao sistema que incorpora os reservatórios Engenheiro Ávidos – REA – e São Gonçalo – RSG – com contribuições do Rio Piancó, ao longo do percurso que deságua para entrada do Rio Açu.

Os motivos da escolha da UPHAP como área de estudo se dão por esta ser controlada por reservatórios que possuem histórico recorrente de crises causadas pela escassez de água característica do semiárido nordestino brasileiro. Além disso, a UPHAP possui uma quantidade suficiente de informações que são indispensáveis para a aplicação de técnicas e métodos de

engenharia para uso em pesquisas científicas, visto que alguns trabalhos acadêmicos já se debruçaram sobre esta UPH, a exemplo das produções científicas publicadas pelo grupo GOTA-UFCG, que subsidiam estudos relevantes sobre os aspectos ambientais, sociais, industriais, agropecuários, técnicos e econômicos para locais semelhantes a UPHAP.

As fontes de dados utilizadas para levantamento das informações técnicas sobre a área de estudo foram provenientes do Plano Diretor de Recursos Hídricos da Bacia do Rio Piancó - PDRHP e Alto Piranhas do Estado da Paraíba (SCIENTEC, 1997) e do Plano de Recursos Hídricos da Bacia do Rio Piancó-Piranhas-Açu – PRHPPA (ANA, 2016), juntamente com pesquisas de grande respaldo científico, com destaque para os vários trabalhos já desenvolvidos pelo grupo de pesquisas GOTA da UFCG, com destaque para os trabalhos de Neto (2018), Nepomuceno (2018) e Neto (2020).

Outro quesito importante é que a região possui a influência hidrológica de alguns reservatórios relevantes e interdependentes para as bacias hidrográficas situadas no Estado da Paraíba, sendo estes reservatórios de grande importância para gestão dos recursos hídricos em nível estadual e federal.

Segundo a ANA, a BHRPPA faz parte da Região Hidrográfica do Nordeste Oriental – RN localizada no extremo oeste do Estado da Paraíba, entre as latitudes 6° 36' 47'' e 7° 22' 56'' Sul e entre as longitudes 37° 48' 15'' e 38° 38' 15'' Oeste no Sertão Paraibano. Os tipos climáticos segundo a classificação de Köppen é A (clima tropical) e B (clima árido), sendo que o clima árido predomina na bacia, mas em porções das UPH's Piancó, Alto Piranhas e Peixe ocorre o clima Tropical.

Na região semiárida do Nordeste brasileiro, a BHRPPA possui área de drenagem de 43.683 km<sup>2</sup>, está parcialmente inserida nos Estados da Paraíba (60%) e do Rio Grande do Norte (40%) e ocupa cerca de 15% do território da Região Hidrográfica Atlântico Nordeste Oriental, com chuvas concentradas em poucos meses do ano e um padrão de forte variabilidade interanual, caracterizado pela alternância entre anos de pluviosidade acima da média, regular e anos consecutivos de valores abaixo da média, que resultam em secas prolongadas e baixa disponibilidade hídrica.

As áreas das sub-bacias hidrograficamente pertencentes ao Rio Piranhas estão localizadas e quantificadas de acordo com a Tabela 3. O sistema de estudo está localizado no sertão paraibano, com precipitações médias anuais é de 854,6 mm para o RSG e 869,5 mm para o REA (AESAs, 2020), os quais se concentram, cerca de 80%, entre os meses de janeiro a abril.

**Tabela 3** – Principais características geográficas da Bacia Hidrográficas do Rio Piranhas.

Área total (km <sup>2</sup> )	Sub-bacias e regiões	Área (km <sup>2</sup> )	Latitude	Longitude
26.047,49	Rio do Peixe	3.420,84	6°20'47" / 7°03'53" S	37°57'52" / 38°46'48" O
	Rio Piancó	9.242,75	6°43'52" / 7°50'28" S	37°26'56" / 38°42'56" O
	Rio Espinharas	2.891,60	6°41'18" / 7°21'51" S	36°43'41" / 37°33'50" O
	Rio Seridó	3.442,36	6°16'57" / 7°03'56" S	36°13'12" / 37°15'16" O
	Alto Piranhas	2.588,45	6°37'18" / 7°22'56" S	37°48'11" / 38°41'14" O
	Médio Piranhas	4.461,48	6°01'38" / 7°00'90" S	37°09'25" / 38°01'44" O

Fonte: Paraíba (1996).

A pluviometria da bacia na qual os reservatórios estão inseridos possui períodos de chuvas concentradas em poucos meses do ano, com padrão sazonal interanual consecutivo, iniciando o período médio de precipitações a partir do mês de dezembro, com ápice do índice pluviométrico no mês de fevereiro, transitando para o período de estiagem durante o mês de maio.

O período de estiagem segue a partir de junho e vai até novembro. A pluviosidade ao longo dos anos apresenta uma variabilidade distribuída de forma que os anos com pluviosidade acima da média são recorrentemente seguidos por ciclos de seca, o que faz com a disponibilidade hídrica para as demandas seja afetada (NEPOMUCENO, 2018; NETO, 2018).

A BHRPPA está subdividida em 7 sub-bacias: Bacia do Alto, Médio e Baixo Piranhas, Bacia do Rio do Peixe, Bacia do Espinharas, Bacia do Piancó e Bacia do Seridó. Os principais critérios escolhidos para divisão foram a hidrografia, presença de reservatórios de grande porte e unidades de gestão adotadas pelos Estados. A UPHAP possui uma área de 2.562 km<sup>2</sup> que corresponde a 5,9% da BHRPPA, abrangendo 19 municípios. Já a UPHMP possui uma área de 2.894 km<sup>2</sup> que corresponde a 6,4%, abrangendo 24 municípios. A caracterização das UPH's no que se refere às suas áreas, percentual que ocupam na bacia, número de municípios abrangidos e sedes municipais, encontra-se na Tabela 4.

**Tabela 4** – Características fisiográficas das unidades de planejamento estudadas.

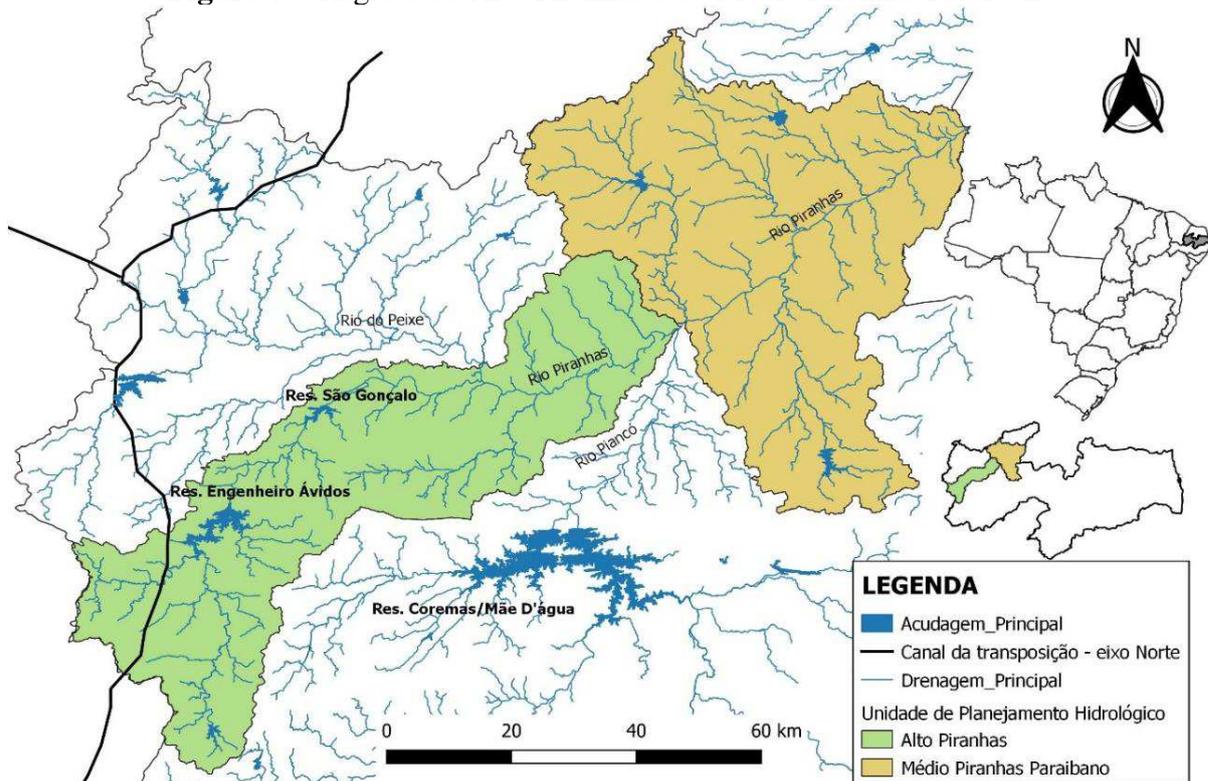
UPH	Área (km <sup>2</sup> )	Área (%)	Nº de Municípios	Nº de Sedes
Alto Piranhas	2562	5,9	19	7
Médio Piranhas Paraíba	2894	6,6	24	11

Fonte: ANA, 2016

Os principais rios da BHRPPA, os rios Piancó-Piranhas-Açu, possuem uma extensão de 400 km, nascendo na Serra de Piancó-PB e desaguando no município de Macau no Estado do Rio Grande do Norte (RN). O Rio Piranhas apresenta suas nascentes na Serra da Arara no município de Bonito de Santa Fé, recebendo contribuições significativas de quatro cursos d'água na sua margem esquerda: Riacho do Juá, Riacho da Caiçara, Riacho Cajazeiras, Riacho

Grande. Na sua margem direita, recebe seis contribuições, quais sejam, Riacho do Domingos, Riacho São Domingos, Riacho Mutuca, Riacho Logradouro, Riacho Catolé, Riacho Bonfim. A Figura 7 apresenta as regiões das sub-bacias hidrográficas do Alto Piranhas e do Médio Piranhas Paraibano.

**Figura 7** – Regiões do Alto Piranhas e do Médio Piranhas Paraibano.

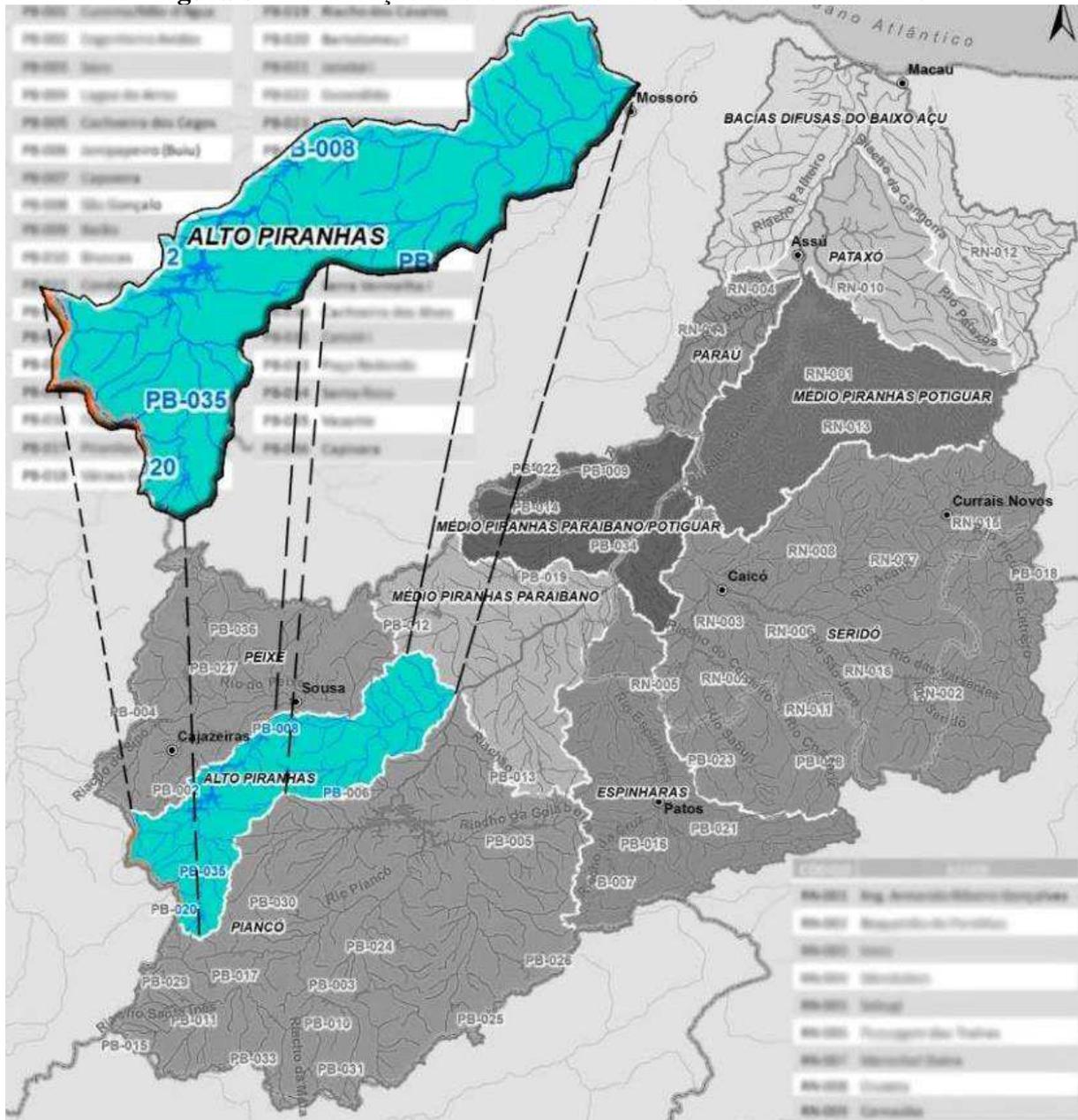


Fonte: Adaptado de ANA (2016).

### 3.4.2 Região do Alto Curso do Rio Piranhas

A região do Alto Piranhas (Figura 8) está inserida dentro da BHRPPA, localizada no Estado da Paraíba na área extrema Oeste, fazendo fronteira com a bacia do Rio Piancó ao Sul e Leste, com a bacia do Rio do Peixe ao Norte, e com a bacia do Médio Curso do Rio Piranhas a Nordeste. O Rio Piranhas tem sua nascente na Serra da Arara em Bonito de Santa Fé-PB e tem seu talvegue na região do Alto Curso, com extensão de 178 km, da nascente até o exutório da bacia de Pombal-PB.

**Figura 8** – Localização da sub-bacia do Alto Curso do Rio Piranhas.



Fonte: Neto (2020).

Segundo Neto (2020), os indicadores fisiográficos apresentados no Plano Estadual de Recursos Hídricos da Paraíba – PERHPB – mostram que a bacia possui um aspecto alongado, implicando em uma baixa susceptibilidade a enchentes. Entretanto, a UPHAP se apresenta com bastante declividade, possuindo inclinações médias da ordem de 9,1 m/km (0,91%). Com relação a densidade de drenagem, a bacia do Alto Piranhas possui 1,4 km de riachos para cada 1 km<sup>2</sup> de área, considerada em eficiência de drenagem razoável conforme a média.

De acordo com Paraíba (1996), a classificação climatológica para a bacia conforme a classificação de Köppen, indica que a área de captação é considerada de clima semiárido quente

– classificado como *Bsh* – e quente e úmido com chuvas concentradas entre verão e outono – Aw. As precipitações anuais médias são de 900 mm/ano aproximadamente, podendo alcançar a máxima de até 1.400 mm/ano em anos mais chuvosos e a mínima de 600 mm/ano em anos de estiagem.

A Tabela 5 apresenta a disponibilidade hídrica dos respectivos reservatórios considerando as curvas de garantia dos açudes da bacia do Rio Piranhas.

**Tabela 5** – Disponibilidade hídrica e garantias da região do Alto Piranhas.

Açude	Disponibilidade (l/s)		
	100%	95%	90%
São José	21,08	53,84	80,36
Bartolomeu I	16,92	41,72	57,91
Eng. Ávidos	1450	1960	2470
São Gonçalo	300	550	695
<i>Subtotal</i>	1788,00	2605,56	3303,27

Fonte: Paraíba (1996).

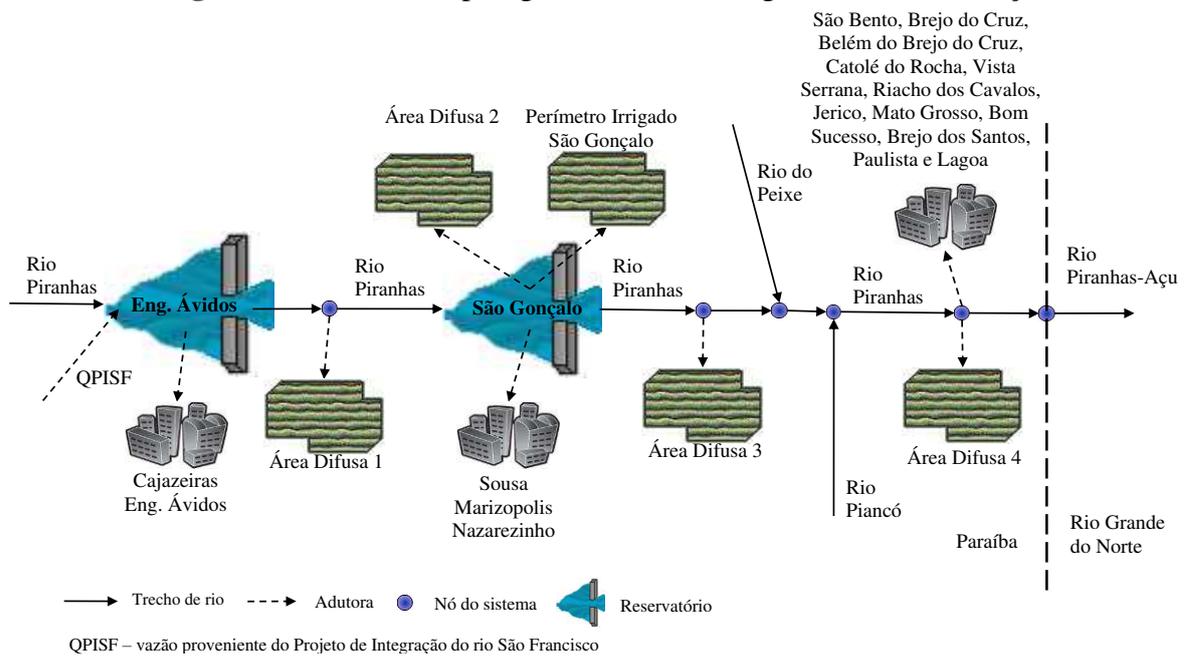
Segundo ANA (2016), a UPHAP contém 72 reservatórios superficiais, que têm como principal finalidade a regularização das vazões em períodos de estiagem. Contudo, dentre estes 72 mananciais, Engenheiro Ávidos, São Gonçalo e Bartolomeu são os que se destacam, cujas capacidades são, nessa ordem, 293 hm<sup>3</sup>, 40 hm<sup>3</sup> e 17 hm<sup>3</sup>.

De acordo com o PERHPB, nos seus aspectos gerais, a situação dos recursos hídricos da UPHAP é de demandas e ofertas equilibradas, com o devido planejamento e gerenciamento. Os órgãos governamentais, todavia, apontam para a necessidade de estudos envolvendo a otimização do uso hídrico e socioeconômico dos reservatórios existentes.

### 3.4.3 Sistema de reservatórios Engenheiro Ávidos – São Gonçalo

O sistema de recursos hídricos estudado é formado por dois reservatórios localizados na BHRPPA. Os reservatórios foram o reservatório REA e o RSG, ambos integrados em série e têm como propósito o uso múltiplo das águas. A Figura 9 apresenta a estrutura topológica que descreve o SEASG, com suas principais demandas, afluentes, conexões e distribuições dos componentes de gerenciamento.

**Figura 9** – Estrutura topológica do sistema Eng. Ávidos-São Gonçalo.



Fonte: Adaptado de Enéas (2018).

Seguindo o escoamento do Rio Piranhas, a estrutura SEASG se inicia com a entrada natural de água pelo Rio Piranhas no REA, junto com a entrada da vazão advinda do Rio São Francisco. Após a passagem pelo reservatório, ela é acumulada e distribuída para o abastecimento urbano de Cajazeiras-PB e da área irrigada difusa atendida a jusante. A vazão a jusante do REA segue para entrada no RSG, para que dele sejam atendidas as demandas de abastecimento urbano das cidades de Souza-PB, Marizópolis-PB e Nazarezinho-PB, como também as demandas agrícolas, áreas difusas e o perímetro irrigado do São Gonçalo.

Essa vazão remanescente segue pelo Rio Piranhas para atendimento da terceira área agrícola difusa e mais à frente recebe contribuições, do Rio do Peixe e do Rio Piancó, este último tendo a montante a contribuição do Sistema Coremas-Mãe d'Água. Logo após estas entradas, o Rio Piranhas distribui a água para uma quarta área agrícola difusa, bem como os municípios paraibanos de: São Bento; Brejo do Cruz; Belém do Brejo do Cruz; Catolé do Rocha; Vista Serrana; Riacho dos Cavalos; Jericó; Mato Grosso; Bom Sucesso; Brejo dos Santos; Paulista e Lagoa. Por fim, o Rio Piranhas desagua na bacia do Rio Piranhas-Açu, inserida no Estado do Rio Grande do Norte.

Ambos os reservatórios são responsáveis pelo fornecimento de água para duas importantes cidades do interior da Paraíba, Cajazeiras e Sousa, além de abastecer outros centros urbanos menores de distritos circundantes, correspondendo ao atendimento de aproximadamente 107.923 habitantes (IBGE, 2020). Segundo Reis (2019), os reservatórios

também suprem demandas hídricas agrícolas nas proximidades do São Gonçalo, com área total de 3.139 ha, como também outras demandas difusas no percurso do Rio Piranhas e no entorno do reservatório (NETO, 2018).

Segundo Nepomuceno (2018), os dois reservatórios estudados são de responsabilidade Federal (ambos construídos pelo DNOCS). A dominialidade dos trechos dos rios que escoam na região em questão é estabelecida pela Resolução ANA nº 399/2004, que especifica a inserção nos limites do Estado da Paraíba, correspondendo ao trecho superior do Rio Piranhas-Açu, a montante da confluência com o Rio Piancó. Desta forma, todos os rios nesta sub-bacia, inclusive o próprio curso superior, possuem dominialidade paraibana.

O REA é o principal reservatório da bacia, localizado no município de Cajazeiras, no Sertão da Paraíba, construído em 1936, com capacidade de armazenamento máxima de 293 milhões de m<sup>3</sup>, tendo como principal função o abastecimento urbano do Município de Cajazeiras. Também construído em 1936, o segundo reservatório é o RSG, à 27 km de distância do REA, no Município de Sousa, abastecendo o município e o perímetro irrigado da região com aproximadamente 5.000 ha, com capacidade máxima de armazenamento de 40,58 milhões de m<sup>3</sup> (ANA, 2018a). A Tabela 6 apresenta as principais características volumétricas para os reservatórios em estudo, com os valores médios calculados a partir das medições dos últimos 27 anos.

**Tabela 6 – Características volumétricas dos reservatórios**

Reservatórios	Volume (hm <sup>3</sup> )	
	Capacidade <sup>1</sup>	Morto <sup>1</sup>
São Gonçalo	40,58	2,04
Eng. Ávidos	293,62	6,72

Fonte: ANA<sup>1</sup> (2018a); AESA<sup>2</sup> (2021)

Para realização da modelagem do sistema, foram utilizadas curvas de comportamento para a evaporação dos reservatórios, consideradas como valores anuais fixos, sendo para este estudo apenas os valores de precipitação considerados a partir da perspectiva estocástica. A Tabela 7 apresenta os valores das evaporações em milímetros para os reservatórios, utilizados no modelo.

**Tabela 7 – Lâminas de evaporação utilizadas na modelagem do sistema.**

Reservatórios	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
São Gonçalo	182	127	106	105	134	150	189	231	247	259	241	229
Eng. Ávidos	173	122	102	102	129	145	184	224	243	252	231	221

Fonte: ANA (2017).

O sistema possui um conjunto de demandas a serem atendidas por um dos reservatórios do sistema por vazões específicas correspondentes. As vazões solicitadas pelas demandas do SEASG, estão apresentadas na Tabela 8.

**Tabela 8 – Vazões e fontes hídricas dos componentes de demanda do sistema.**

<b>Setor</b>	<b>Demanda (m<sup>3</sup>/s)</b>	<b>Fonte Hídrica</b>
Abastecimento urbano de Cajazeiras	0,19 <sup>1</sup>	Res. Eng. Ávidos
Abastecimento urbano de Sousa, Marizópolis e Nazarezinho	0,23 <sup>1</sup>	Res. São Gonçalo
Abastecimento urbano de São Bento, Brejo do Cruz, Belém do Brejo do Cruz, Catolé do Rocha, Vista Serrana, Riacho dos Cavalos, Jerico, Mato Grosso, Bom Sucesso, Brejo dos Santos, Paulista e Lagoa	0,31 <sup>2</sup>	Rio Piranhas
Perímetro Irrigado de São Gonçalo	1,4 <sup>1</sup>	Res. São Gonçalo
Área Difusa 1	0,15 <sup>1</sup>	Rio Piranhas
Área Difusa 2	0,16 <sup>1</sup>	Res. São Gonçalo
Área Difusa 3	0,17 <sup>1</sup>	Rio Piranhas
Área Difusa 4	1,6 <sup>2</sup>	Rio Piranhas

Fonte: <sup>1</sup>ANA (2018b), <sup>2</sup>ANA (2019)

### 3.4.4 Elaboração dos cenários operacionais

Com o objetivo de definir os principais cenários operacionais de simulação, a elaboração dos cenários do sistema contemplou as variações de situações ocorridas em pontos estratégicos da topologia, a partir das condições de funcionamento do SEASG, sendo os nós de estudo:

- A vazão exógena do Projeto de Integração do Rio São Francisco - PISF no reservatório Engenheiro Ávidos;
- Demandas após a confluência com o rio Piancó até a divisa dos Estado da Paraíba e do Rio Grande do Norte;
- Saída para a vazão efluente do Rio Piranhas desaguando no Estado do Rio Grande do Norte.

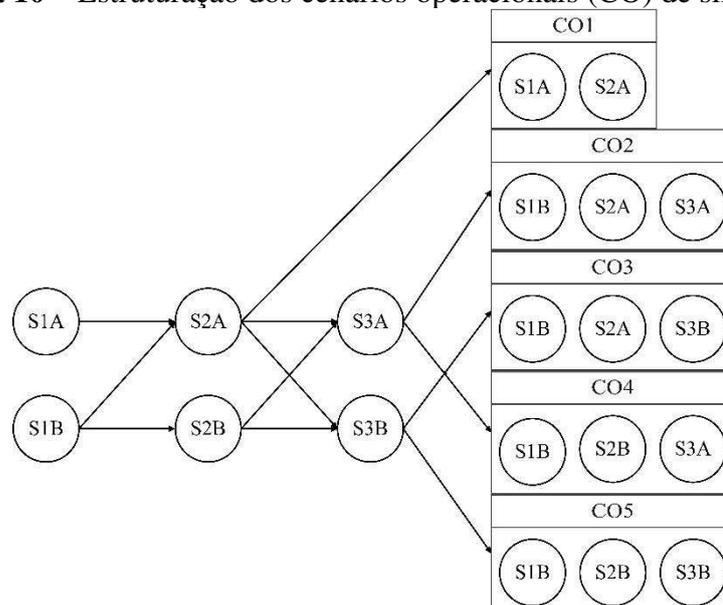
A interação entre as variações situacionais do sistema anteriormente citados, resultou na seleção de 5 possíveis cenários de operação e alocação de água, divididos em três itens:

- I. Situação relacionada com a vazão de entrada do PISF no REA.
  - Situação S1A – Não haverá vazão exógena do PISF no REA ou no sistema;

- Situação S1B – A vazão exógena do PISF no REA será de 2,7 m<sup>3</sup>/s como preconizado pelo PRHPPA (ANA, 2017), dos quais 1,0 m<sup>3</sup>/s será destinado para atendimento das demandas do Estado da Paraíba e 1,7 m<sup>3</sup>/s para atendimento das demandas do Estado do Rio Grande do Norte;
- II. Situação relacionada ao atendimento das demandas no rio Piranhas após a confluência com o rio Piancó até a divisa dos Estado da Paraíba e do Rio Grande do Norte.
- Situação S2A – As demandas serão de responsabilidade do SEASG;
  - Situação S2B – As demandas serão de responsabilidade pelo sistema Coremas/Mãe D'água como preconizado na Resolução conjunta ANA/IGARN/AESA N° 65/2019 (ANA, 2019);
- III. Situação relacionada com a vazão do Rio Piranhas na divisa dos Estados da Paraíba e do Rio Grande do Norte.
- Situação S3A – Será atendida integralmente como preconizado nas situações S1B;
  - Situação S3B – Será simulada com a mesma prioridade de atendimento das demandas de abastecimento urbano.

A Figura 10 apresenta as combinações utilizadas para modelagem dos cenários operacionais e que foram simulados.

**Figura 10** – Estruturação dos cenários operacionais (CO) de simulação.



Fonte: Autor (2022).

A natureza do estudo da interação entre essas situações pontuadas, teve como finalidade a identificação e análise do *Best-Case Scenario*, do *Worst-Case Scenario* e de seus intermediários ou seja, tanto os cenários com o melhor panorama de atendimento às demandas e menor risco potencial, quanto os de pior panorama de atendimento às demandas e maior risco potencial, observando as principais características, causas e efeitos, com relação aos riscos hídricos, aos quais o SEASG está submetido.

As combinações das situações modeladas levaram em conta a ocorrência de uma ou mais situações que puderam fornecer resultados diferentes conforme fora elaborado o cenário para o sistema.

Tendo sido selecionados os cenários operacionais, foram então criados os arquivos base para a simulação no AcquaNet, sendo estes previamente configurados para manterem as mesmas características topológicas para todos as séries sintéticas de vazão implementadas nos arquivos, onde cada arquivo de cada cenário teria as vazões naturais dos REA e RSG diferentes.

Para ambos os reservatórios foram definidos alguns parâmetros como: volume inicial em  $\text{hm}^3$ ; volumes máximos e mínimos; taxa de evaporação em  $\text{m}^3/\text{mês}$ ; a tabela com os dados da curva cota-área-volume; as vazões naturais em  $\text{m}^3/\text{s}$  (dados sintéticos). Além desses elementos, as vazões e prioridades de atendimento das demandas urbanas e rurais também foram definidas, sendo as demandas urbanas de maior prioridade, seguida das demandas rurais.

Para o Cenário Operacional 1 – CO1 – que levou em consideração o não fornecimento da vazão pelo PISF, sendo desconsiderado o atendimento das demandas no rio Piranhas após a confluência com o rio Piancó até a divisa dos Estados da Paraíba e do Rio Grande do Norte, a representação foi estruturada conforme apresentado na Figura 11. Apesar do não fornecido pelo PISF, sua afluência no sistema foi representada por um nó de passagem, que nesse caso apresentou vazão natural zero.

O Cenário Operacional 2 – CO2 – possui como principal diferença a entrada da vazão fornecida pelo PISF também representada por um nó de passagem com vazão natural igual a  $2,7 \text{ m}^3/\text{s}$ . Contudo, além do atendimento as demandas urbanas e rurais do SEASG, buscou-se o atendimento prioritário da vazão do Rio Piranhas na divisa dos Estados da Paraíba e do Rio Grande do Norte, representada por um elemento de demanda, também conforme a Figura 11, com prioridade máxima e capacidade mínima do link que interliga o nó de passagem após o RSG, correspondente a  $1,7 \text{ m}^3/\text{s}$ , conforme S2A. Este valor também foi definido para a demanda, obrigando o algoritmo de otimização *out-of-kilter*, utilizado pelo *software*, a entregar água para o elemento.

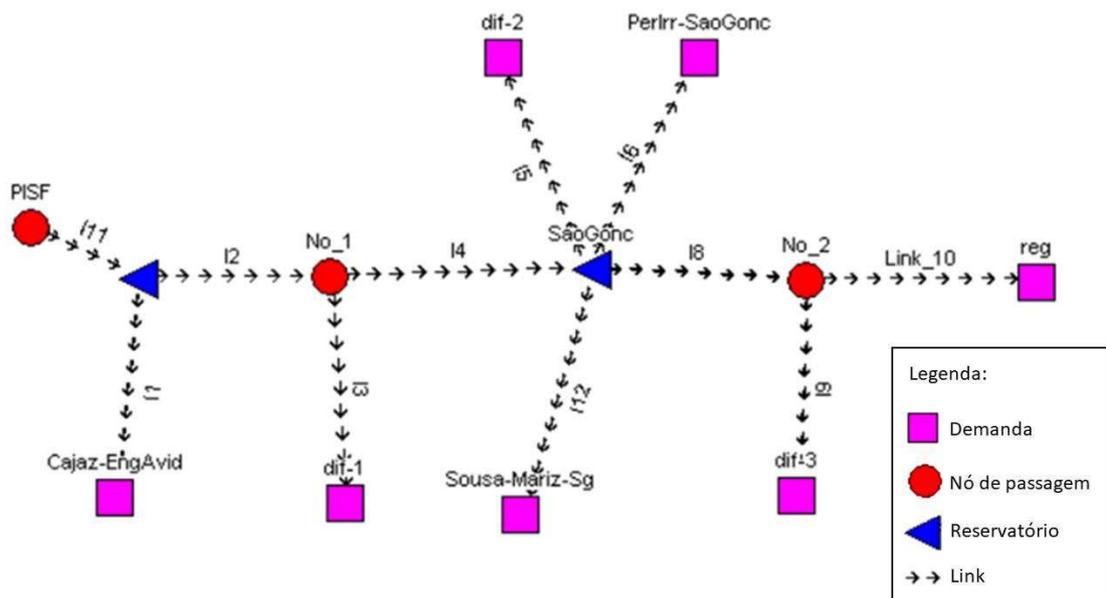
Já o Cenário Operacional 3 – CO3 – também representado pela Figura 11, possui a estrutura topológica representada de forma idêntica aos cenários anteriormente detalhados, porém com o valor da capacidade mínima do link de ligação da vazão regularizada igual a zero, com a finalidade de torná-lo uma demanda. Com isso o sistema será simulado de modo a fornecer água para os elementos conforme suas prioridades sem a necessidade de atendimento máximo quando os volumes de reservação não forem suficientes.

O Cenário Operacional 4 – CO4 – representado na Figura 12, diferentemente dos anteriores, possui uma estrutura topológica diferente, onde os nós de passagem a jusante do RSG possuem a saída de dois links, ligados com duas demandas urbanas e rurais, estas compreendendo as demandas urbanas que são definidas no sistema a depender da ocorrência da situação S3A, correspondendo ao aumento da soma de vazão demandada pelos reservatórios, simultaneamente apoiados pela vazão fornecida do PISF. Outro fator importante para a definição do CO4 foi o estabelecimento da S2A, também abordada pelo CO2.

O Cenário Operacional 5 – CO5 – possui também mesma estrutura topológica apresentada pela Figura 12, seguindo os mesmos aspectos do CO4, possuindo a presença das demandas que primeiramente tem atendimento conforme a S3A, contudo obedecendo a S2B, com características de atendimento ao elemento da vazão regularizada tal como o CO3, considerando a ocorrência da S2B.

Nas imagens de tela do AcquaNet, apresentadas nas figuras 11 e 12, estão ilustradas as estruturas do sistema para a simulação dos cenários operacionais.

**Figura 11** – Estruturação do SEASG para o CO1, CO2 e CO3 no AcquaNet.



Fonte: Autor (2021)



estabelecer os melhores diagnósticos da situação do risco de acordo com as regras de operação do sistema observando as ramificações que levaram aos nós folha com o maior número de simulações classificadas em cada uma das classes.

## 4 RESULTADOS

Nesta seção estão apresentados os resultados da previsão do diagnóstico das regras de operação a partir das principais características do sistema identificados para as simulações dos diferentes cenários operacionais.

### 4.1 CENÁRIO OPERACIONAL 1

Neste cenário, a simulação não considerou a entrada de vazão advinda do PISF no SEASG, sem atendimento das demandas após a confluência com o Rio Piancó e sem regras de operação para a vazão regularizada do envio ao Rio Grande do Norte.

A análise do indicador  $C$  obtido para cada uma das demandas do CO1, pode ser realizada através da Tabela 10 e da Figura A9, contendo dados estatísticos descritivos para todas as demandas e os histogramas de distribuição das simulações para a confiabilidade.

**Tabela 10** – Estatísticas descritivas das confiabilidades de atendimento às demandas do CO1

	<i>AbastEA</i>	<i>AbastSG</i>	<i>Dif1</i>	<i>Dif2</i>	<i>Dif3</i>	<i>PISG</i>
$\bar{X}$	63,20	70,10	57,19	60,60	59,16	60,26
<b>Med</b>	63,33	70,83	56,67	60,83	59,17	60,00
<b>Std</b>	17,64	15,62	19,66	19,20	19,76	19,44
<b>Mín</b>	5,83	19,17	5,83	10,83	8,33	8,33
<b>Máx</b>	100,00	100,00	100,00	100,00	100,00	100,00
<b>q1</b>	50,83	59,17	42,50	45,83	44,17	45,83
<b>q3</b>	75,83	81,67	71,67	75,00	74,17	75,00

Fonte: Autor (2022)

O valor médio de  $C$  para a demanda urbana do RSG foi maior, fato que pode ser explicado pela maior quantidade de elementos atendidos pelo RSG, sendo menos sensível a diminuição da porcentagem do indicador.

O valor do desvio padrão se manteve entre 15,62% e 19,76% para todas as demandas, com destaque para a demanda urbana do RSG, que teve o menor valor. Tal fato sugere que o RSG opere em melhores condições de atendimento mesmo com variações climatológicas, uma vez que na simulação do SEASG, o RSG possui entrada de água por meio da precipitação e da vazão fornecida pelo REA, este sem nenhuma entrada além da precipitação, indicando maior facilidade de manutenção do atendimento.

Outros indícios de que o RSG consiga suportar melhor os efeitos causados pelas várias mudanças dos eventos sintéticos, são os valores mínimos que se mantiveram maiores tanto para

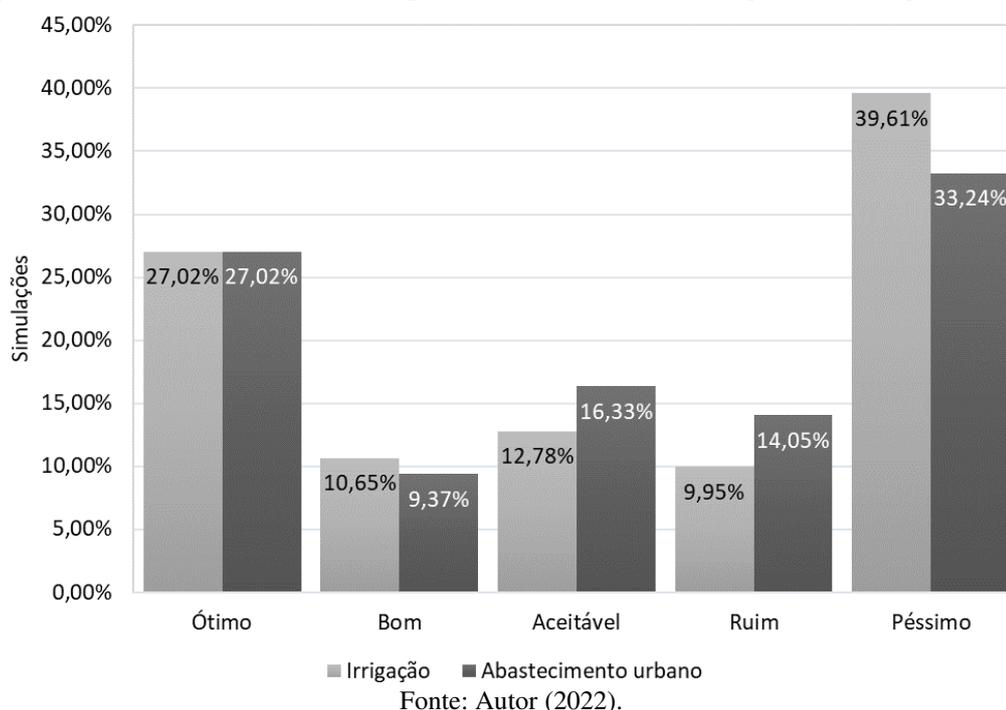
a demanda urbana do RSG quanto para as demais demandas rurais atendidas por ele. Mesmo no pior evento sintético, o SEASG conseguiu manter o atendimento um pouco melhor para as demandas atendidas pelo RSG.

É possível observar que existe uma semelhança com relação aos dados do indicador *C* para as demandas rurais. A área difusa 1 apresentou os piores resultados com relação ao número de simulações sintéticas com confiabilidade próxima aos valores da amplitude interquartil, além de também possuir um maior número de simulações com valor de *C* igual a zero.

#### 4.1.1 Árvore de decisão para demandas abastecimento urbano do CO1

A Figura 13 apresenta o percentual médio de *C* das simulações para as demandas de irrigação e no abastecimento urbano para o CO1, classificadas de acordo com cada uma das classes de risco estabelecidas.

**Figura 13** – Percentuais de simulações conforme a classificação de risco para o CO1.



A classe “Péssimo” apresentou um maior percentual de simulações, sendo as diferenças entre as classes “Aceitável” e “Ruim” de 3,55% e 4,1% respectivamente, além de que a classe “Péssimo” obteve maior quantidade de simulações para a irrigação, com 6,37% de simulações a mais que para o abastecimento urbano.

A acurácia que representa a capacidade de classificação dos modelos para o abastecimento urbano do CO1 obteve 96,11%, representando o um bom desempenho na representação das características das simulações para cada classe.

Os resultados para os modelos gerados para as confiabilidades de abastecimento urbano obtiveram precisão e *recall* conforme os valores da Tabela 11.

**Tabela 11** – Desempenho da árvore de decisão para o abastecimento urbano do CO1.

Classes	Precisão	<i>Recall</i>
Ótimo	0,98	1,00
Bom	0,99	0,89
Aceitável	0,94	0,92
Ruim	0,85	0,98
Péssimo	1,00	0,96

Fonte: Autor (2022).

A precisão representa o número de simulações previstas corretamente como pertencentes a uma classe em relação ao número total de simulações previstas. O resultado indicou o melhor desempenho na classificação da classe “Péssimo”, seguidas das classes “Bom” e “Ótimo”, que sugere um comportamento mais bem definido quando o sistema está em situações extremas.

O *recall* é a fração de simulações classificadas corretamente previstas pelo modelo, em relação ao número total de simulações realmente pertencentes a classe observada do conjunto de dados.

É possível identificar, assim como já evidenciado na Figura 13 que as classes “Ótimo” e “Péssimo” foram as que possuíram a maior quantidade de simulações classificadas. Apesar do alto nível de acurácia obtido, as classes “Bom” “Aceitável” e “Ruim”, possuíram falhas de previsão em 3,78% do total de simulações testadas, sendo a classes “Péssimo” aquela com o melhor desempenho de previsão.

A árvore de decisão para classificação do risco no abastecimento urbano para o CO1 está apresentada na Figura 14.



O teste de normalidade de Shapiro-Wilk das acurácias calculadas na validação cruzada, para a árvore de decisão da confiabilidade para o abastecimento urbano, resultou no p-valor de 0,959, mantendo a hipótese nula de que os dados apresentam distribuição normal, assim sendo possível confirmar a robustez dos modelos em detrimento da variação das variáveis de treino.

Os níveis de importância das variáveis independentes para a construção do modelo da árvore de decisão considerando o abastecimento urbano estão apresentados na Tabela 12.

**Tabela 12** – Níveis de importância das variáveis para o abastecimento urbano do CO1.

Variável independente	Nível de importância
<i>Dif1</i>	0,1%
<i>Dif2</i>	0,0%
<i>Dif3</i>	16,7%
<i>PISG</i>	0,4%
<i>AbastEA</i>	44,4%
<i>AbastSG</i>	37,7%
<i>P</i>	0,2%
<i>VfimEA</i>	0,5%

Fonte: Autor (2022).

As ramificações com o maior número de simulações de cada situação de risco para o abastecimento urbano resultaram nos diagnósticos de operação mais prováveis de acordo com as *Cer*, que foram organizadas e apresentadas na Tabela 13.

**Tabela 13** – Diagnóstico do sistema de acordo com as *Cer* do abastecimento urbano no CO1.

Classe	Descrição	Operação
Ótimo	Confiabilidade média anual igual a 100%	$P > 636,1 \text{ mm}$ $VfimEA > 47,99\%$ da capacidade de armazenamento $AbastEA > 0,175 \text{ m}^3/\text{s}$ $Dif3 > 0,165 \text{ m}^3/\text{s}$
Bom	Confiabilidade média anual entre 95% e 99,99%	$AbastEA > 0,175 \text{ m}^3/\text{s}$ $AbastSG > 0,225 \text{ m}^3/\text{s}$ $0 \text{ m}^3/\text{s} < Dif3 \leq 0,165 \text{ m}^3/\text{s}$ $PISG > 1,285 \text{ m}^3/\text{s}$
Aceitável	Confiabilidade média anual entre 70% e 94,99%	$0,125 \text{ m}^3/\text{s} < AbastEA \leq 0,175 \text{ m}^3/\text{s}$ $0,205 \text{ m}^3/\text{s} < AbastSG \leq 0,225 \text{ m}^3/\text{s}$
Ruim	Confiabilidade média anual entre 50% e 69,99%	$0,095 \text{ m}^3/\text{s} < AbastEA \leq 0,115 \text{ m}^3/\text{s}$ $0,115 \text{ m}^3/\text{s} < AbastSG \leq 0,225 \text{ m}^3/\text{s}$ $Dif1 > 0,055 \text{ m}^3/\text{s}$ $0,706 \text{ m}^3/\text{s} < PISG \leq 1,126 \text{ m}^3/\text{s}$
Péssimo	Confiabilidade média anual abaixo de 50%	$0\% < VfimEA \leq 3,95\%$ $0 \text{ m}^3/\text{s} < AbastEA \leq 0,175 \text{ m}^3/\text{s}$ $0 \text{ m}^3/\text{s} < AbastSG \leq 0,115 \text{ m}^3/\text{s}$

Fonte: Autor (2022).

A estruturação da árvore apresentou as demandas *AbastEA*, *AbastSG* e *Dif3* como as mais importantes para a classificação do risco, seguido por *VfimEA*, *PISG*, *P* e *Dif1* respectivamente. A variável de vazão *Dif2* não apresentou importância quanto a definição das classes, sendo o valor da vazão sem influência para o risco de não atendimento do abastecimento urbano.

Para a classe “Ótimo”, a variável com maior ganho de informação e que primeiro dividiu as amostras para o subconjunto da classe foi *AbastEA*, sendo as simulações com valores de *AbastEA* menores ou iguais a 92,1% da vazão requerida consideradas fora da classe, seguido da vazão *Dif3* que dividiu o subconjunto da amostra, sendo por sua vez desconsiderado “Ótimo” quando os valores de *Dif3* foram menores ou iguais a 97,1% da vazão requerida.

A modelagem da árvore para a classe indica que o desempenho dos valores de *C* para o abastecimento urbano, possui como característica mais comum a alocação de água, chuvas e volume finais do REA no ano anterior, ocorrendo nos intervalos apresentados na Tabela 13. É importante salientar que as demandas que não tiveram seus padrões de comportamento definidos, devem ser atendidas conforme o requerido, para que com isso o sistema possua percentuais de *C* conforme as classes.

A partir do subconjunto com  $AbastEA > 0,175 \text{ m}^3/\text{s}$  e  $Dif3 > 0,165 \text{ m}^3/\text{s}$ , as simulações já passaram a ser classificadas “Ótimo”, mesmo com variações para as demais variáveis, sendo as simulações com *VfimEA* a partir de 74,5 hm<sup>3</sup> consideradas como “Ótimo”.

Os padrões de alocação observados para a classe “Ótimo” obtiveram a maior quantidade de simulações da classe, sendo representada por 35,04% da amostra classificada como “Ótimo”, acima da média de 8,33%.

Outra classe que também mereceu destaque foi a classe “Bom”, sendo *AbastEA* e *Dif3* as variáveis com maior ganho de informação, contudo, a classe considera o subconjunto da amostra com as vazões da *Dif3* menores ou iguais a 92,1% da vazão requerida, com o subconjunto da amostra possuindo vazões de *AbastSG* acima de 97,8% da vazão requerida.

O diagnóstico da alocação do sistema mediante o risco são uma importante informação para a tomada de decisão pois ajuda os gestores na operação com fornecimento de água reduzida em demandas estratégicas para manter a média de *C* para o abastecimento acima de 95%.

A partir do subconjunto com  $AbastEA > 0,175 \text{ m}^3/\text{s}$  e  $Dif3 \leq 0,165 \text{ m}^3/\text{s}$ , as simulações já passaram a ser classificadas “Bom”, sendo as vazões de *AbastSG*  $> 0,225 \text{ m}^3/\text{s}$  e *PISG*  $> 1,285 \text{ m}^3/\text{s}$  as os padrões de alocação da ramificação com maior quantidade de simulações para

a classe, sendo representada por 81,79% da amostra classificada como “Bom”, acima da média de 50% da representatividade das respostas para a classe.

Os intervalos das variáveis utilizadas para determinação da classe “Péssimo” consideraram as variáveis de demanda *AbastEA* e *AbastSG* como as de maior ganho de informação, sendo o subconjunto da classe definido a partir das vazões abaixo de 92,1% e 97,8% respectivamente.

Como a classe “Péssimo” representa um cenário com metade do sistema com falha de atendimento, os resultados mostraram que a média de *C* para o abastecimento nas situações de crise hídrica depende prioritariamente das demandas urbanas, mesmo com redução do fornecimento para as demandas de irrigação.

A partir do subconjunto com  $AbastEA \leq 0,175 \text{ m}^3/\text{s}$  e  $AbastSG \leq 0,115 \text{ m}^3/\text{s}$ , as simulações já passaram a ser classificadas “Péssimo”, sendo  $VfimEA \leq 3,95 \%$  a ramificação com maior quantidade de simulações para a classe, com 95,5% da representatividade das respostas para a classe.

#### 4.1.2 Árvore de decisão para irrigação do CO1

A acurácia para a irrigação do CO1 obteve 96,81%, representando um bom desempenho na representação das características das simulações para cada classe.

Os resultados para os modelos gerados para as confiabilidades das demandas de irrigação obtiveram precisão e *recall*, conforme os valores da Tabela 14.

**Tabela 14** – Desempenho da árvore de decisão para a irrigação no CO1.

Classes	Precisão	<i>Recall</i>
Ótimo	1,00	1,00
Bom	0,97	0,99
Aceitável	0,97	0,87
Ruim	0,80	0,96
Péssimo	1,00	0,97

Fonte: Autor (2022).

Os resultados para a previsão indicaram o melhor desempenho na classificação da classe “Ótimo” e “Péssimo”, seguidas das classes “Bom” e “Aceitável”, sendo a classe “Ruim” aquela que possuiu a menor precisão, o que sugere a menor homogeneidade das amostras além de ser o menor dos conjuntos de simulações das classes.

Para o *recall*, a classe “Ótimo” obteve o melhor resultado, mostrando que esta é a classe mais importante para calibração da previsão, apesar de não haver diferença significativa se comparado ao das demais classes.

A árvore de decisão para classificação do risco na irrigação para o CO1 está apresentada na Figura 15.

O teste de normalidade de Shapiro-Wilk das acurácias calculadas na validação cruzada, para a árvore de decisão da confiabilidade para a irrigação do CO1, resultou no p-valor de 0,968, mantendo a hipótese nula com distribuição normal dos resultados de acurácia.

Os níveis de importância das variáveis independentes da árvore de decisão para a irrigação estão apresentados na Tabela 15.

**Tabela 15** – Níveis de importância das variáveis para a irrigação no CO1.

Variável independente	Nível de importância
<i>Dif1</i>	11,3%
<i>Dif2</i>	0,0%
<i>Dif3</i>	76,3%
<i>PISG</i>	12,2%
<i>AbastEA</i>	0,0%
<i>AbastSG</i>	0,0%
<i>P</i>	0,1%
<i>VfimEA</i>	0,1%

Fonte: Autor (2022).

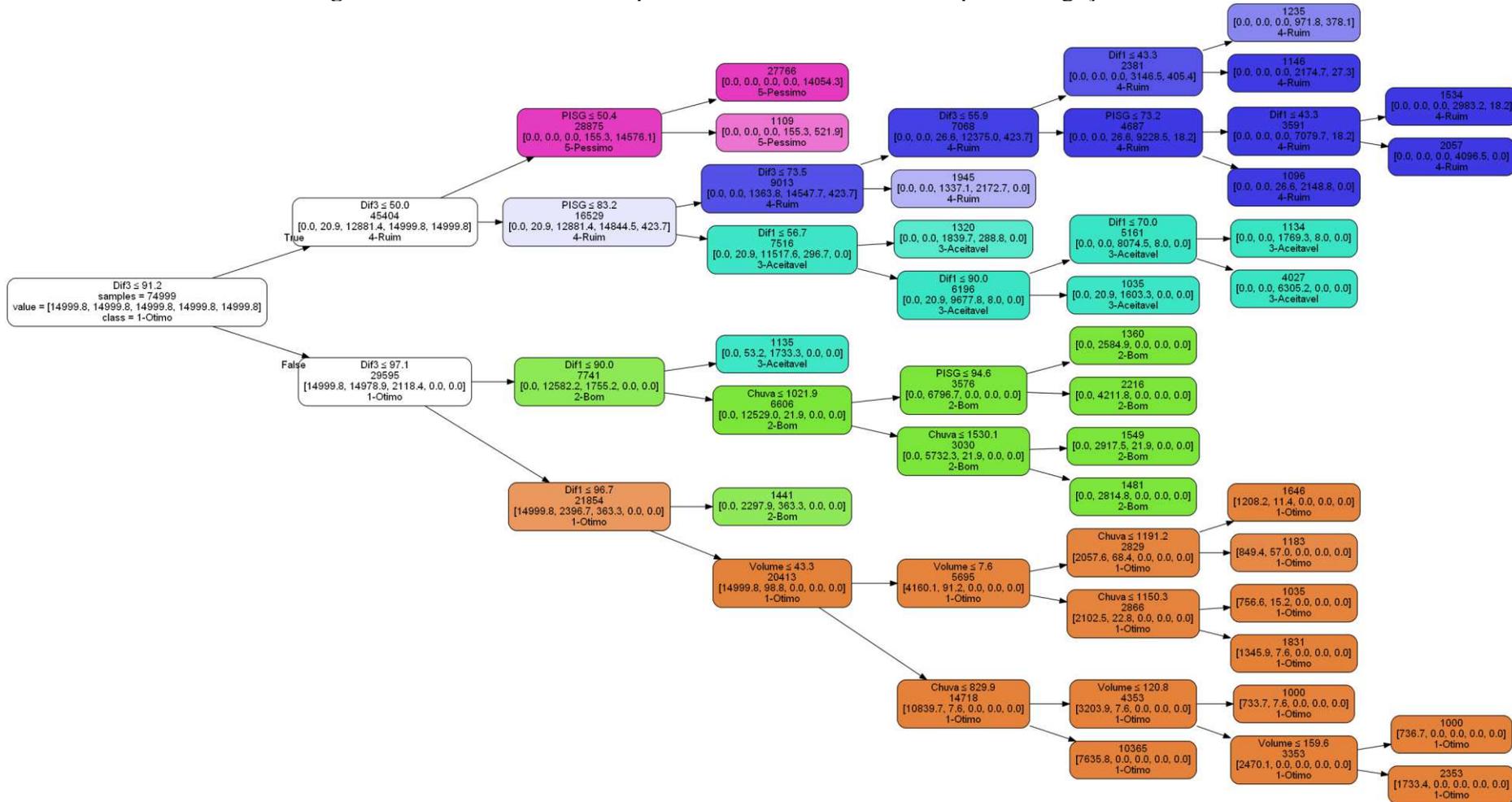
Os diagnósticos de operação mais prováveis de acordo com as *Cer* estão apresentadas na Tabela 16.

**Tabela 16** – Diagnóstico do sistema de acordo com as *Cer* para a irrigação no CO1.

Classe	Descrição	Operação
Ótimo	Confiabilidade média anual igual a 100%	$P > 829,9$ mm $VfimEA > 14,75\%$ da capacidade de armazenamento $Dif1 > 0,145$ m <sup>3</sup> /s $Dif3 > 0,165$ m <sup>3</sup> /s
Bom	Confiabilidade média anual entre 95% e 99,99%	$Dif1 > 0,135$ m <sup>3</sup> /s $0,155$ m <sup>3</sup> /s < $Dif3 \leq 0,165$ m <sup>3</sup> /s $PISG > 1,324$ m <sup>3</sup> /s
Aceitável	Confiabilidade média anual entre 70% e 94,99%	$0,105$ m <sup>3</sup> /s < $Dif1 \leq 0,135$ m <sup>3</sup> /s $0,085$ m <sup>3</sup> /s < $Dif3 \leq 0,155$ m <sup>3</sup> /s $PISG > 1,165$ m <sup>3</sup> /s
Ruim	Confiabilidade média anual entre 50% e 69,99%	$Dif1 > 0,065$ m <sup>3</sup> /s $0,085$ m <sup>3</sup> /s < $Dif3 \leq 0,125$ m <sup>3</sup> /s $0$ m <sup>3</sup> /s < $PISG \leq 1,165$ m <sup>3</sup> /s
Péssimo	Confiabilidade média anual abaixo de 50%	$0$ m <sup>3</sup> /s < $Dif3 \leq 0,085$ m <sup>3</sup> /s $0$ m <sup>3</sup> /s < $PISG \leq 0,706$ m <sup>3</sup> /s

Fonte: Autor (2022).

Figura 15 – Árvore de decisão para as confiabilidades médias para a irrigação no CO1.



Fonte: Autor (2022).

A estruturação da árvore apresentou as demandas *Dif3*, *PISG* e *Dif1* como as mais importantes para a classificação do risco, seguido por *VfimEA* e *P* respectivamente. As variáveis de vazão *Dif2*, *AbastEA* e *AbastSG* não apresentaram importância quanto a definição das classes, sendo o valor da vazão sem influência para o risco de não atendimento da irrigação, sendo possível enviar a vazão requerida para as demandas.

Para a classe “Ótimo”, a variável com maior ganho de informação e que primeiro dividiu a amostra para o subconjunto da classe foi *Dif3*, sendo as simulações com valores menores ou iguais a 97,1% da vazão requerida consideradas fora da classe, seguido da vazão *Dif1* que dividiu o subconjunto da amostra, sendo por sua vez desconsiderado “Ótimo” quando os valores de *Dif1* foram maiores que 96,7% da vazão requerida.

A modelagem da árvore no geral indicou que a quantidade de simulações por *split* se manteve acima do mínimo estabelecido, proporcionando uma árvore com menos ramificações e com menos nós folha, visto que o número de variáveis independentes com níveis de importância de maior relevância diminuiu se comparado com a árvore para o abastecimento urbano.

A partir do subconjunto com  $Dif1 > 0,145 \text{ m}^3/\text{s}$  e  $Dif3 > 0,165 \text{ m}^3/\text{s}$ , as simulações já passaram a ser classificadas “Ótimo”, mesmo com ramificações envolvendo outras variáveis, sendo *P* e *VfimEA*, sendo as simulações com *VfimEA* a partir de 43,3 hm<sup>3</sup> consideradas dentro da classe. A precipitação por sua vez levou a divisão do maior subconjunto para a classe, sendo a situação “Ótimo” considerada para valores de  $P > 829,9 \text{ mm}$ .

Os padrões de alocação do sistema de acordo com o risco para a classe “Ótimo” obtiveram a maior quantidade de simulações, sendo representada por 50,77% da amostra classificada como “Ótimo”, acima da média de 12,5%.

Para a classe “Bom”, a árvore apresentou uma maior influência das chuvas e do *PISG* na definição da situação com maior número de simulações classificadas para a classe, sendo considerado primeiramente a *Dif3* maior que 91,2% e menor ou igual a 97,1% da vazão requerida, seguida da *Dif1* maior que 90% da vazão requerida, com *PISG* operando com vazão menor ou igual a 97,6% da vazão requerida. Outra situação possível é a operação com *Dif3* acima de 97,1% e *Dif1* menor ou igual a 96,7% da vazão requerida.

Os padrões de alocação da ramificação com maior quantidade de simulações para a classe apontadas na Tabela 17 apresentaram 27,5% da amostra classificada como “Bom”, acima da média de 20% da representatividade das respostas para a classe.

Entre as possíveis situações de padrões de comportamento das variáveis para a classe “Bom”, a situação dos valores de *C* para a irrigação mostrou que os nós folhas à diferentes

situações de operação possuem possibilidades de ocorrência semelhantes, com o mínimo e máximo de simulações classificadas por resposta de 16,9% e 27,5%.

A quantidade de demandas para o uso na irrigação no CO1 é duas vezes mais que as demandas para abastecimento urbano, contudo, apenas as demandas *Dif3* e *PISG* foram consideradas para definição da classe “Péssimo”, sendo a vazão da *Dif3* menor ou igual a 50% e o *PISG* menor ou igual a 50,4% da vazão requerida.

A partir do subconjunto com *Dif3* menor ou igual a 50,0%, as simulações já passaram a ser classificadas “Péssimo”, sendo *PISG* abaixo de 50,4% a ramificação com maior quantidade de simulações para a classe, com 94,3% da representatividade das respostas para a classe.

## 4.2 CENÁRIO OPERACIONAL 2

Neste cenário a simulação foi considerada a entrada da vazão do PISF no SEASG, sem o atendimento das demandas após a confluência com o Rio Piancó e com atendimento prioritário da vazão enviada ao Rio Grande do Norte.

O fornecimento de água adicional ocasionado pelo PISF é o maior destaque deste cenário operacional, devido a presença dos seus efeitos serem observadas pelas demandas para o abastecimento urbano que foram atendidas com *C* acima de 95%, em todas as simulações realizadas.

A análise do indicador *C* obtido para cada uma das demandas do CO2, pode ser realizada através da Tabela 17, contendo dados estatísticos descritivos para todas as demandas, juntamente com a figura A9, em que estão apresentados os histogramas de distribuição do indicador *C* para as demandas rurais e urbanas do CO2.

**Tabela 17** – Estatística descritiva das confiabilidades para as demandas do CO2

	<i>AbastEA</i>	<i>AbastSG</i>	<i>Dif1</i>	<i>Dif2</i>	<i>Dif3</i>	<i>PISG</i>
$\bar{X}$	100,00	100,00	100,00	70,79	70,20	71,33
<b>Med</b>	100,00	100,00	100,00	71,67	71,67	72,50
<b>Std</b>	0,00	0,00	0,00	18,87	19,16	18,58
<b>Mín</b>	100,00	100,00	100,00	15,83	14,17	15,83
<b>Máx</b>	100,00	100,00	100,00	100,00	100,00	100,00
<b>q1</b>	100,00	100,00	100,00	56,67	55,83	57,50
<b>q3</b>	100,00	100,00	100,00	85,83	86,67	85,83

Fonte: Autor (2021)

O valor médio de  $C$  para as demandas urbanas, junto com os valores da área difusa 1 obtiveram o valor máximo para todos os eventos simulados do sistema, fato que pode ser explicado pelo acréscimo de vazão enviada pelo PISF, sem haver uma demanda proporcionalmente correspondente no SEASG.

As diferenças de valor do  $C$  são significativas, estando apenas as demandas rurais do RSG sem o atendimento plenamente atendido, sugerindo que no CO2, a proximidade da demanda do fornecimento de água, juntamente com as prioridades de atendimento levaram ao desequilíbrio na oferta de água do sistema, onde uma parcela do sistema passou a não ser atendida em detrimento de uma outra.

O valor do desvio padrão para as demandas que não apresentaram atendimento pleno, se manteve entre 18% e 20%, sendo considerados baixos para a quantidade de eventos simulados. Tal fato sugere que apesar do percentual abaixo de 100% para as demandas rurais após o RSG, o reservatório ainda foi capaz de manter uma certa estabilidade em detrimento das variações hidroclimáticas das simulações, também observado no CO1.

Assim como no CO1, os valores mínimos das demandas rurais após o RSG, possuem porcentagens entre 14% e 16%, faixa de atendimento mantido mesmo para os eventos sintéticos mais severos, sendo estes valores maiores que os mínimos encontrados no CO1 para essas mesmas demandas rurais, significando que mesmo nos cenários mais severos, o sistema pode ser considerado mais bem atendido nas demandas desfavorecidas.

Outro destaque pode ser feito com relação a *Dif1*, que apesar de ser uma demanda rural, ainda sim apresentou um comportamento diferente das demais demandas de mesma categoria e de mesma prioridade, o que sugere que a topologia do SEASG é uma característica considerada relevante para o atendimento na simulação, fator que pode ser melhor administrado em se tratando de uma situação real.

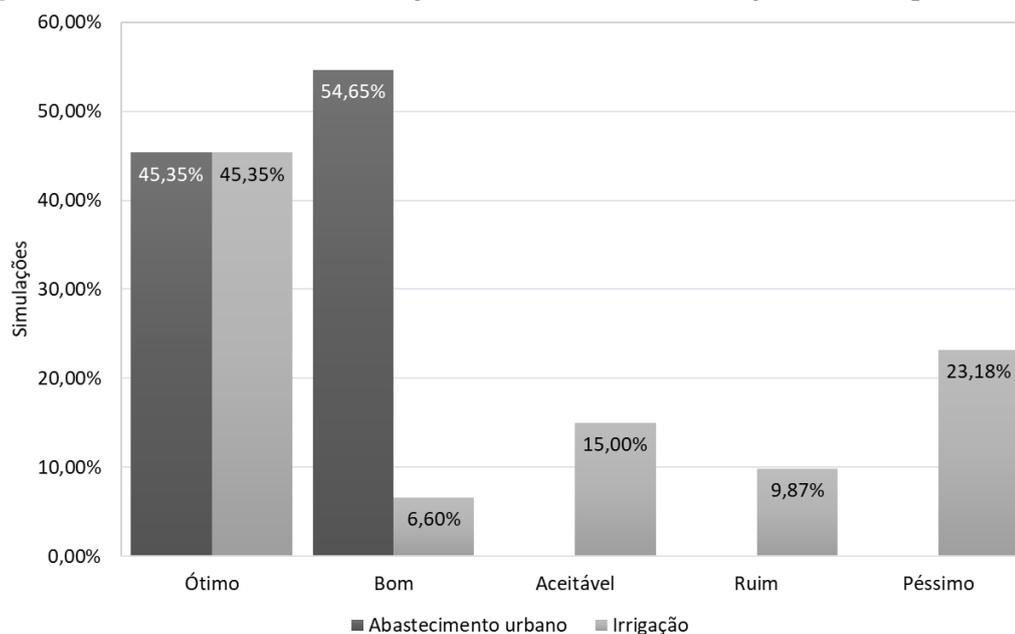
#### **4.2.1 Árvore de decisão para abastecimento urbano do CO2**

O alto valor de  $C$  para o abastecimento urbano do CO2 apresentou como principal desvantagem a impossibilidade de análise de situações com menores  $C$ , com simulações classificadas “Aceitável”, “Ruim” e “Péssimo”, que sob a perspectiva da modelagem das árvores e da definição do diagnóstico de operação limita a possibilidade de manobras para as demandas e favorecer a tomada de decisão em situações de crise hídrica.

É importante salientar que a vazão disponibilizada pelo PISF foi considerada constante durante as simulações, mas que na realidade funciona com menos regularidade.

A Figura 16 apresenta o percentual médio de *C* para as simulações para os usos na irrigação e no abastecimento urbano para o CO<sub>2</sub> respectivamente, classificadas de acordo com cada uma das classes de risco estabelecidas.

**Figura 16** – Percentuais de simulações conforme a classificação de risco para o CO<sub>2</sub>.



Fonte: Autor (2022).

As classificações para o abastecimento urbano comparadas com a irrigação apresentaram diferenças para o valor das classificações “Bom”, “Aceitável”, “Ruim” e “Péssimo”. Apesar do CO<sub>2</sub> ser considerado um cenário com maior disponibilidade de água, a irrigação ainda apresentou casos com classificações com pior desempenho. A classe “Bom” se destacou com uma considerável diferença, indicando que o uso da irrigação não apresentou equilíbrio na alocação de água com envio prioritário de água desconsiderando a manutenção de *C* de ambos os usos.

A acurácia para o abastecimento urbano do CO<sub>2</sub> obteve 99,8%, indicando o um bom desempenho na representação das características das simulações para cada classe, favorecido pela menor quantidade de classes identificadas.

Os resultados para os modelos gerados para as confiabilidades de abastecimento urbano do CO<sub>2</sub> obtiveram precisão e *recall*, conforme os valores da Tabela 18.

**Tabela 18** – Desempenho da árvore de decisão para o abastecimento urbano no CO2.

Classes	Precisão	Recall
Ótimo	1,00	1,00
Bom	1,00	0,99

Fonte: Autor (2022).

O resultado obtido para a precisão indicou o melhor desempenho nas classificações das classes “Ótimo” e “Bom”, significando que as classificações previstas pelo modelo foram plenamente identificadas conforme as classes observadas nas simulações.

As classes “Ótimo” e “Bom” obtiveram valores de *recall* semelhantes com valores máximos, sendo a classe “Bom” pouco menor e apresentando, porém com a identificação das classes observadas quase em sua totalidade.

Mesmo com melhor atendimento das demandas de abastecimento urbano correspondendo aos altos valores de *C*, as classes identificadas para a irrigação apresentaram classificação semelhante aos percentuais mostrados no CO1.

A árvore de decisão para classificação do risco no abastecimento urbano para o CO2 está apresentada na Figura 17.

O teste de normalidade de Shapiro-Wilk das acurácias calculadas na validação cruzada, para a árvore de decisão da confiabilidade para o abastecimento urbano, resultou no p-valor de 0,991, mantendo a hipótese nula de que os dados apresentam distribuição normal, assim sendo possível observar a robustez dos modelos em detrimento da variação das variáveis de treino.

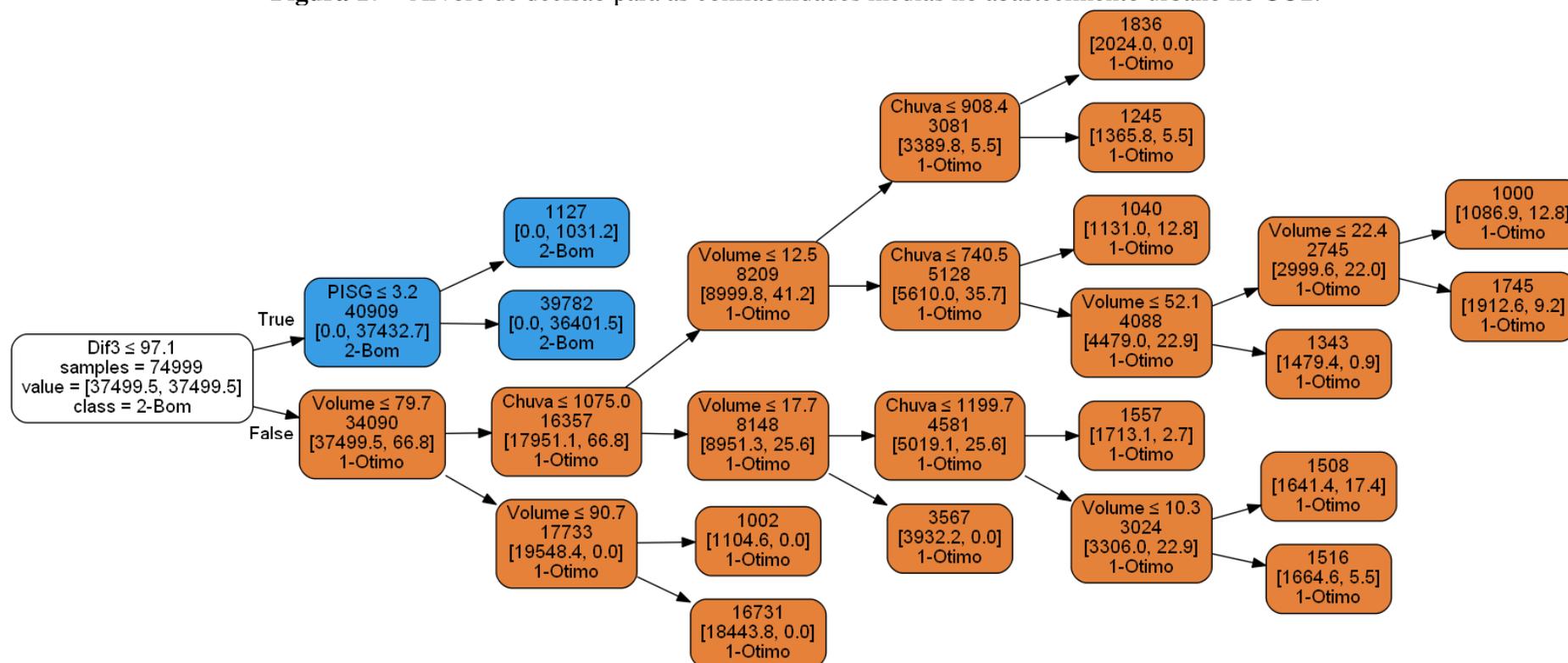
Os níveis de importância das variáveis independentes da árvore de decisão para o abastecimento urbano estão apresentados na Tabela 19.

**Tabela 19** – Níveis de importância das variáveis para o abastecimento urbano no CO2.

Variável independente	Nível de importância
<i>Dif1</i>	0,0%
<i>Dif2</i>	0,0%
<i>Dif3</i>	99,8%
<i>PISG</i>	0,1%
<i>AbastEA</i>	0,0%
<i>AbastSG</i>	0,0%
<i>P</i>	0,0%
<i>VfimEA</i>	0,2%

Fonte: Autor (2022).

**Figura 17** – Árvore de decisão para as confiabilidades médias no abastecimento urbano no CO2.



Fonte: Autor (2022).

Os níveis de importância verificados para o abastecimento urbano do CO<sub>2</sub> mostraram que com exceção do volume, a definição das classes dependeu predominantemente da vazão da *Dif3*, umas das vazões de irrigação atendida diretamente pelo RSG, apesar das principais confiabilidades afetadas são demandas de abastecimento, o que indica que quando o sistema começa a apresentar falhas no atendimento, a *Dif3* é a primeira demanda a ser restringida.

As ramificações com o maior número de simulações de cada situação de risco para o abastecimento urbano do CO<sub>2</sub> resultaram nos diagnósticos de operação mais prováveis de acordo com as *Cer*, que foram organizadas e apresentadas na Tabela 20.

**Tabela 20** - Diagnóstico do sistema de acordo com as *Cer* do abastecimento urbano no CO<sub>2</sub>.

Situação	Descrição	Operação
Ótimo	Confiabilidade média anual igual a 100%	$V_{fimEA} > 27,15\%$ da capacidade de armazenamento $Dif3 > 0,165 \text{ m}^3/\text{s}$
Bom	Confiabilidade média anual entre 95% e 99,99%	$0 \text{ m}^3/\text{s} < Dif3 \leq 0,165 \text{ m}^3/\text{s}$ $PISG > 0,045 \text{ m}^3/\text{s}$

Fonte: Autor (2022).

Os padrões de alocação para a árvore de decisão para o abastecimento urbano do CO<sub>2</sub> apresentaram ramificações para apenas duas classes que já são definidas a partir do primeiro nível de ramificação, sendo o subconjunto da classe “Ótimo” formado com a vazão de *Dif3* acima de 97,1% da vazão requerida. Para a classe, o subconjunto possui como principal característica a influência do *VfimEA*, com valores maiores que 79,7 hm<sup>3</sup>.

O diagnóstico dos padrões de alocação para a classe “Ótimo” apresentado na Tabela XX obtiveram a maior quantidade de simulações para a classe, sendo representada por 47,98% da amostra classificada como “Ótimo”, acima da média de 8,33%. Contudo, as chuvas também apresentaram influencia dentro do subconjunto amostral da classe, fazendo parte dos padrões de alocação com a segunda maior quantidade de simulações classificadas, com 10,46% da amostra. Nessa ramificação o valor do volume quando entre 17,7 hm<sup>3</sup> e 79,7 hm<sup>3</sup>, as chuvas devem ser maiores que 1075 mm.

A influência do *VfimEA* no subconjunto da classe “Ótimo” sugere que seja realizada a manutenção do volume para o cenário CO<sub>2</sub>, que está diretamente associado ao fornecimento de vazão do PISF, com envio da vazão constante para o REA, sendo o fornecimento para a *Dif3* também dependente da disponibilidade hídrica do REA. Assim como para os demais cenários, a modelagem da árvore não leva em consideração a vazões sem níveis de importância para a classificação, porém é importante que na prática as demandas sejam atendidas conforme as vazões requeridas, para que os percentuais de *C* estejam dentro dos intervalos das classes.

Também foi possível identificar o subconjunto da classe “Bom” para o abastecimento, sendo definição do subconjunto a partir da vazão de *Dif3* menor ou igual a 97,1% da vazão requerida, sendo o subconjunto já considerado na classe ‘Bom’ com a divisão do subconjunto da amostra acima do 3,2% da vazão requerida pelo PISF, o que indicou que quando o sistema apresenta perdas de desempenho no atendimento, a demanda que sofre com mais variações de vazão fornecida é o *PISG*.

O padrão de alocação do sistema para a classe foi originado da ramificação com maior quantidade de simulações, sendo representado por 97,24% da amostra classificada como “Bom”.

#### 4.2.2 Árvore de decisão para irrigação do CO2

A acurácia para a irrigação do CO2 obteve 99,7%, representando o um bom desempenho na representação das características das simulações para cada classe. Os resultados para os modelos gerados para as confiabilidades de abastecimento urbano obtiveram precisão e *recall*, conforme os valores da tabela 21.

**Tabela 21** – Desempenho da árvore de decisão para a irrigação no CO2.

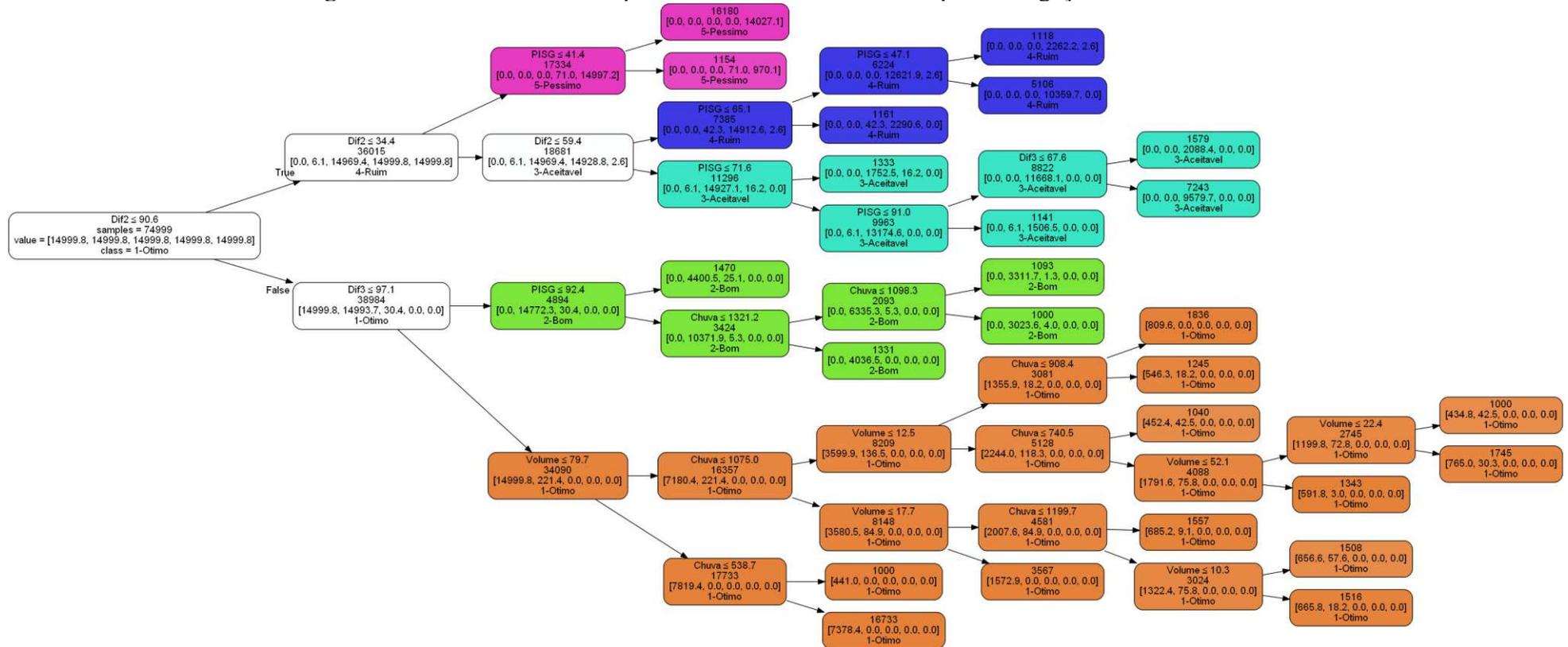
Classes	Precisão	<i>Recall</i>
Ótimo	1,00	1,00
Bom	1,00	0,99
Aceitável	1,00	1,00
Ruim	1,00	0,99
Péssimo	1,00	1,00

Fonte: Autor (2022).

Os resultados para a previsão indicaram o melhor desempenho na classificação para todas as classes, com valores máximos de precisão, significando que as classificações observadas foram plenamente previstas pelo modelo.

Para o *recall*, as classes prejudicadas foram as a classe “Ótimo” obteve o melhor resultado, mostrando que esta é a classe mais importante para calibração da previsão, apesar de não haver diferença significativa se comparado ao das demais classes. A árvore de decisão para classificação do risco na irrigação para o CO2 está apresentada na Figura 18.

Figura 18 – Árvore de decisão para as confiabilidades médias para a irrigação no CO2.



Fonte: Autor (2022).

O teste de normalidade de Shapiro-Wilk das acurácias calculadas na validação cruzada, para a árvore de decisão da confiabilidade para a irrigação do CO<sub>2</sub>, resultou no p-valor de 1,00, mantendo a hipótese nula com distribuição normal dos resultados de acurácia.

Os níveis de importância das variáveis independentes da árvore de decisão para a irrigação estão apresentados na Tabela 22.

**Tabela 22** – Níveis de importância das variáveis para a irrigação no CO<sub>2</sub>.

Variável independente	Nível de importância
<i>Dif1</i>	0,0%
<i>Dif2</i>	82,9%
<i>Dif3</i>	16,5%
<i>PISG</i>	0,3%
<i>AbastEA</i>	0,0%
<i>AbastSG</i>	0,0%
<i>P</i>	0,0%
<i>VfimEA</i>	0,2%

Fonte: Autor (2022).

Os diagnósticos de operação mais prováveis de acordo com as *Cer* estão apresentadas na Tabela 23.

**Tabela 23** – Diagnóstico do sistema de acordo com as *Cer* para a irrigação do CO<sub>2</sub>.

Situação	Descrição	Operação
Ótimo	Confiabilidade média anual igual a 100%	$P > 538,7$ mm $VfimEA > 27,15\%$ da capacidade de armazenamento $Dif2 > 0,145$ m <sup>3</sup> /s $Dif3 > 0,166$ m <sup>3</sup> /s
Bom	Confiabilidade média anual entre 95% e 99,99%	$Dif2 > 0,145$ m <sup>3</sup> /s $0$ m <sup>3</sup> /s < $Dif3 \leq 0,166$ m <sup>3</sup> /s $PISG > 1,294$ m <sup>3</sup> /s
Aceitável	Confiabilidade média anual entre 70% e 94,99%	$0,095$ m <sup>3</sup> /s < $Dif2 \leq 0,145$ m <sup>3</sup> /s $0,115$ m <sup>3</sup> /s < $Dif3 \leq 0,166$ m <sup>3</sup> /s $1,01$ m <sup>3</sup> /s < $PISG \leq 1,274$ m <sup>3</sup> /s
Ruim	Confiabilidade média anual entre 50% e 69,99%	$0,055$ m <sup>3</sup> /s < $Dif2 \leq 0,085$ m <sup>3</sup> /s $0,69$ m <sup>3</sup> /s < $PISG \leq 0,902$ m <sup>3</sup> /s
Péssimo	Confiabilidade média anual abaixo de 50%	$0$ m <sup>3</sup> /s < $Dif2 \leq 0,055$ m <sup>3</sup> /s $Dif3 > 0,005$ m <sup>3</sup> /s $0$ m <sup>3</sup> /s < $PISG \leq 0,58$ m <sup>3</sup> /s

Fonte: Autor (2022).

A estruturação da árvore apresentou *Dif2* e *Dif3* como as variáveis mais importantes para a classificação do risco, seguido por *PISG* e *VfimEA* respectivamente. As variáveis de vazão *Dif1*, *AbastEA* e *AbastSG* não apresentaram importância quanto a definição das classes,

sendo o valor da vazão sem influência para o risco de não atendimento do abastecimento urbano, sendo possível enviar a vazão requerida para as demandas.

Para a classe “Ótimo”, a variável com maior ganho de informação e que primeiro dividiu a amostra para o subconjunto da classe foi *Dif2*, sendo as simulações com valores maiores que 90,6% da vazão requerida consideradas no subconjunto da classe, seguido da vazão *Dif3* que dividiu o subconjunto amostral, sendo por sua vez considerado “Ótimo” quando os valores de *Dif3* foram maiores que 97,1% da vazão requerida.

A partir do subconjunto com  $Dif2 > 0,145 \text{ m}^3/\text{s}$  e  $Dif3 > 0,165 \text{ m}^3/\text{s}$ , as simulações já passaram a ser classificadas “Ótimo”, mesmo com ramificações envolvendo outras variáveis, sendo *P* e *VfimEA*, sendo as simulações com *VfimEA* a partir de 79,7 hm<sup>3</sup> consideradas dentro da classe. A precipitação por sua vez levou a divisão do maior subconjunto para a classe, sendo a situação “Ótimo” considerada para valores de  $P > 538,7 \text{ mm}$ .

Os padrões de alocação para a classe “Ótimo” obtiveram a maior quantidade de simulações para a classe, com representatividade de 49,08% da amostra classificada como “Ótimo”, acima da média de 8,3%.

Apesar do conjunto de variáveis estabelecidos para as ramificações da classe “Ótimo” do CO2 serem semelhantes ao conjunto visto no CO1, os valores da precipitação necessário para a divisão do maior subconjunto da classe foi menor para o CO2, abaixo da média das precipitações sintéticas utilizadas que foi de 995,5 mm, assim como o limite de vazão da *Dif2* também foi menor em 6,1%. Estes dados apontam para a mesma caracterização dos cenários, apesar da pequena diferença das características observadas, o que sugere que independente a disponibilidade de água no sistema, que primeiro serão afetadas as demandas do RSG.

Os nós folha encontrados para a classe “Bom” apresentaram quantidades de simulações semelhantes com média de 25,0% e desvio padrão de 4,39%, significando que todas as ramificações da árvore para a classe forneceram probabilidades de ocorrência semelhantes e que levaram aos percentuais de *C* acima de 95% e abaixo de 100%. O subconjunto com a classificação foi obtido a partir das simulações com vazão de *Dif2* acima de 90,6% e *Dif3* menor ou igual a 97,1%, a partir dessas características já sendo considerado “Bom”.

O valor do *PISG* menor ou igual a 92,4% da vazão requerida por sua vez, levou a identificação do subconjunto com a maior quantidade de simulações para a classe apontadas na Tabela 23, apresentando 30,04% da amostra classificada como “Bom”, acima da média de 25,0% da representatividade das respostas para a classe.

Outros padrões de alocação que também poderiam ser identificadas a partir das respostas da árvore de decisão para a classe “Bom”, seriam a manutenção das vazões para *Dif2*

e *Dif3*, porém com vazão do *PISG* acima de 92,4% nos casos em que a precipitação anual estivesse entre 1083,3 mm e 1321,2 mm. Estes cenários, portanto, dependeriam das condições meteorológicas para o maior fornecimento de vazão ao *PISG*.

Para as demais classes observadas na árvore, a vazão do *PISG* foi aquela que mais apareceu nas ramificações após a definição dos limites para as vazões *Dif2* e *Dif3*. A vazão do *PISG* foi considerada como critério de definição para as respostas com maior probabilidade de ocorrência.

Para a classe “Aceitável”, a vazão *Dif2* se manteve entre 34,4% e 59,4%, enquanto a *Dif3* se manteve acima de 67,6% com o *PISG* entre 71,6% e 91% da vazão requerida. Já para a classe “Ruim”, o subconjunto considerou *Dif2* menor ou igual a 59,4%, com *PISG* entre 47,1% e 65,1% da vazão requerida, sendo a principal diferença entre as classes “Aceitável” e “Ruim” o limite definido em *Dif2* e as faixas de atendimento para o *PISG*.

A partir do subconjunto com *Dif2* menor ou igual a 34,4%, as simulações já passaram a ser classificadas “Péssimo”, sendo *PISG* abaixo de 41,4% a ramificação com maior quantidade de simulações para a classe, com 93,34% da representatividade das respostas para a classe.

### 4.3 CENÁRIO OPERACIONAL 3

Neste cenário a simulação foi considerada a entrada da vazão do *PISF* no *SEA*, sem o atendimento das demandas após a confluência com o Rio *Piancó*, contudo, o atendimento da vazão enviada para o Rio Grande do Norte foi simulado com a mesma prioridade de atendimento que as demandas urbanas do *SEASG*.

No geral, o cenário operacional apresentou resultados muito semelhantes com o *CO2*, sendo juntamente com o *CO2*, o cenário o que obteve os melhores resultados. O fornecimento de água adicional ocasionado pelo *PISF* continuou sendo o maior destaque do cenário, mostrando que o envio de vazão integral e prioritária ao *RN* não possuiu efeitos significativos balanço hídrico das simulações.

A análise do indicador *C* obtido para cada uma das demandas do *CO3*, pode ser realizada através da Tabela 24, contendo dados estatísticos descritivos para todas as demandas, juntamente com a Figura A13, em que estão apresentados os histogramas de distribuição do indicador *C* para as demandas rurais e urbanas do *CO3*.

**Tabela 24** – Estatística descritiva das confiabilidades para as demandas do CO3

	<i>AbastEA</i>	<i>AbastSG</i>	<i>Dif1</i>	<i>Dif2</i>	<i>Dif3</i>	<i>PISG</i>
$\bar{X}$	100,00	100,00	100,00	71,63	71,08	72,15
<b>Med</b>	100,00	100,00	100,00	72,50	72,50	73,33
<b>Std</b>	0,00	0,00	0,00	18,67	18,93	18,39
<b>Mín</b>	100,00	100,00	100,00	15,83	14,17	15,83
<b>Máx</b>	100,00	100,00	100,00	100,00	100,00	100,00
<b>q1</b>	100,00	100,00	100,00	58,33	57,50	58,33
<b>q3</b>	100,00	100,00	100,00	86,67	86,67	87,50

Fonte: Autor (2021)

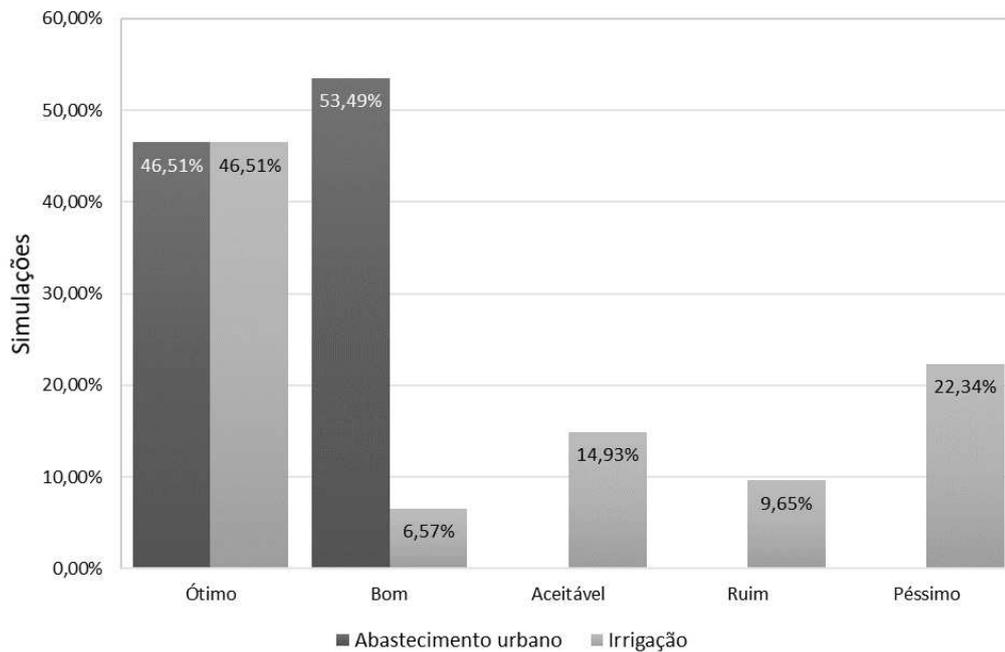
O valor médio de *C* para as demandas urbanas, junto com os valores de *Dif1* obtiveram o valor máximo para todos as simulações, fato que reforça o indício de semelhança entre os CO2 e CO3, novamente explicado pelo aumento da disponibilidade e sem aumento das demandas do SEASG.

As diferenças de valor do *C* são muito pequenas se comparados os valores de todas as demandas tanto rurais como urbanas, contudo a correspondência dos resultados não indica a necessária obtenção dos mesmos resultados de abastecimento observando simulação por simulação, avaliadas par a par. Para o CO3, a interpretação segue o mesmo raciocínio de que a proximidade das demandas do PISF, considerando a estrutura topológica de como o sistema foi representado influenciou na obtenção dos resultados e no pleno atendimento somente para a área difusa 1, de todas as demandas rurais.

O valor do desvio padrão para as demandas que não apresentaram atendimento pleno, foi ainda mais estabilizado, com porcentagens de 18% aproximadamente. O desvio padrão para as demandas rurais após o RSG, sugere que este seria o cenário com melhor estabilidade dos atendimentos em função das variações hidroclimáticas dos eventos sintéticos.

#### 4.3.1 Árvore de decisão para o abastecimento urbano do CO3

A Figura 19 apresenta o percentual médio de *C* para as simulações para os usos na irrigação e no abastecimento urbano para o CO3 respectivamente, classificadas de acordo com cada uma das classes de risco estabelecidas.

**Figura 19** – Percentuais de simulações conforme a classificação de risco para o CO3.

Fonte: Autor (2022).

As diferenças de valor do  $C$  entre os CO2 e CO3 são muito pequenas se comparados os valores de todas as demandas tanto rurais como urbanas, contudo a correspondência dos resultados não indica a necessária obtenção dos mesmos resultados para as mesmas condições das simulações, avaliadas individualmente. Para o CO3, a interpretação segue o mesmo raciocínio de que o fornecimento às demandas pelo PISF, considerando a estrutura topológica de como o sistema foi representado, influenciou na obtenção dos resultados e no pleno atendimento para o abastecimento urbano.

As classificações para o abastecimento urbano comparadas com a irrigação também apresentaram diferenças para o valor das classificações “Bom”, “Aceitável”, “Ruim” e “Péssimo”. Assim como no CO2, o CO3 também apresentou falhas na confiabilidade para as demandas de irrigação, com o comportamento da classificação semelhante aos demais cenários. Foi possível observar que a classe “Péssimo” foi a que apresentou um maior número de simulações para ambos os cenários na irrigação, mesmo com a disponibilidade de água.

O CO3 também obteve 99,8% de acurácia no modelo para o abastecimento, o que indica um bom desempenho na representação das características das simulações para cada classe, favorecido pela menor quantidade de classes identificadas.

Os resultados para os modelos gerados para as confiabilidades de abastecimento urbano do CO3 obtiveram precisão e *recall*, conforme os valores da Tabela 25.

**Tabela 25** – Desempenho da árvore de decisão para o abastecimento urbano do CO3.

Classes	Precisão	Recall
Ótimo	1,00	1,00
Bom	1,00	1,00
Aceitável	1,00	1,00
Ruim	0,99	1,00
Péssimo	1,00	1,00

Fonte: Autor (2022).

O resultado obtido para a precisão indicou o melhor desempenho em todas as classificações, apesar da classe “Ruim” não haver alcançado o valor máximo. O resultado indicou que o modelo possui alta capacidade de classificação, com os dados de teste plenamente identificados conforme as classes observadas nas simulações.

As classes obtiveram o valor máximo para o *recall*, com bons resultados assim como para a precisão, sendo todas as simulações identificadas corretamente pelo modelo.

Os níveis de importância das variáveis independentes da árvore de decisão para o abastecimento urbano estão apresentados na Tabela 26.

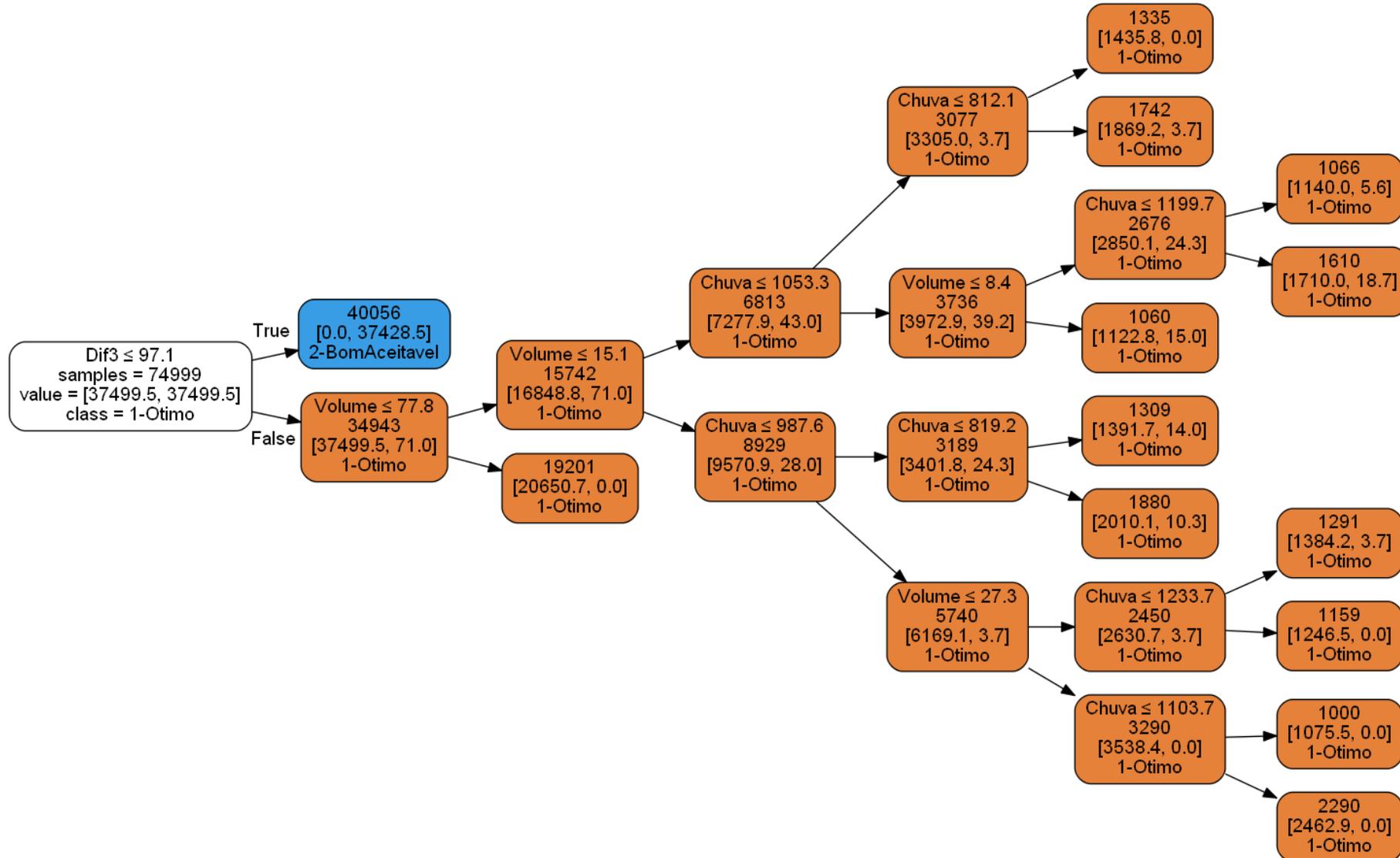
**Tabela 26** – Níveis de importância das variáveis para abastecimento urbano do CO3.

Variável independente	Nível de importância
<i>Dif1</i>	0,0%
<i>Dif2</i>	0,0%
<i>Dif3</i>	99,8%
<i>PISG</i>	0,0%
<i>AbastEA</i>	0,0%
<i>AbastSG</i>	0,0%
<i>P</i>	0,1%
<i>VfimEA</i>	0,1%

Fonte: Autor (2022).

A árvore de decisão para classificação do risco no abastecimento urbano para o CO3 está apresentada na Figura 20.

**Figura 20** – Árvore de decisão para as confiabilidades médias no abastecimento urbano no CO3.



Fonte: Autor (2022).

O teste de normalidade de Shapiro-Wilk das acurácias calculadas na validação cruzada, para a árvore de decisão da confiabilidade para o abastecimento urbano, resultou no p-valor de 1,0, mantendo a hipótese nula de que os dados apresentam distribuição normal, assim sendo possível confirmar a robustez dos modelos em detrimento da variação das variáveis de treino.

Os níveis de importância verificados para o abastecimento urbano do CO3 demonstraram uma forte influência da  $Dif3$  na definição das classificações “Ótimo” e “Bom”, juntamente com uma pequena porcentagem de 0,1% para a precipitação e 0,1% para o volume.

A definição das classes dependeu predominantemente da vazão da  $Dif3$  assim como no cenário anterior, indicando mais uma vez que quando o sistema começa a apresentar falhas no atendimento, a  $Dif3$  é a primeira demanda a ser restringida.

As ramificações com o maior número de simulações de cada situação de risco para o abastecimento urbano do CO3 resultaram nos diagnósticos de operação mais prováveis de acordo com as  $Cer$ , que foram organizadas e apresentadas na Tabela 27.

**Tabela 27** – Diagnóstico do sistema de acordo com as  $Cer$  do abastecimento urbano no CO3.

Situação	Descrição	Operação
Ótimo	Confiabilidade média anual igual a 100%	$VolEA > 26,5\%$ da capacidade de armazenamento $Dif3 > 0,165 \text{ m}^3/\text{s}$
Bom	Confiabilidade média anual entre 95% e 99,99%	$0 \text{ m}^3/\text{s} < Dif3 \leq 0,165 \text{ m}^3/\text{s}$

Fonte: Autor (2022).

O diagnóstico dos padrões de alocação de acordo com o risco para o abastecimento urbano do CO3 apresentou uma menor quantidade de ramificações para manipulação das vazões de demanda, com apenas duas variáveis com intervalos de operação para a classe “Ótimo” e uma variável de demanda para a classe “Bom”.

O subconjunto da classe “Ótimo” foi formado pelo subconjunto com vazão de  $Dif3$  acima de 97,1% da vazão requerida já considerado como pertencente a classe, sendo o subconjunto da classe com maior quantidade de simulações que possuíram como principal característica a influência do  $VfimEA$ , com valores maiores que  $77,8 \text{ hm}^3$ . A ramificação que foi considerada para o diagnóstico da operação da Tabela 27, apresentou 54,95% de representatividade do total de simulações da classe, acima da média de 8,33%.

Outras ramificações da classe “Ótimo” com o segundo e o terceiro maior grupo da amostra, apresentaram condições que consideraram  $27,3 < VfimEA \leq 77,8 \text{ hm}^3$  e  $P$  acima de 1103,7 mm no segundo grupo e  $VfimEA > 15,4 \text{ hm}^3$  e  $P$  próximo a média, entre 819,2 mm e

987,6 mm. Tais resultados indicaram que assim como nos cenários anteriores, as variáveis relacionadas as condições hidrológicas do sistema tem maior interferência na “Ótimo”, indicando quais as condições mínimas de volume e de precipitação para que o sistema atenda as demandas plenamente.

O CO3 também possui o fornecimento da vazão do PISF, sugerindo que a manutenção do volume do REA é uma condição importante para o atendimento das demandas, principalmente para a *Dif3* que têm sido a demanda que primeiro apresenta falhas no atendimento quando sob condições hidrológicas desfavoráveis.

Também foi possível identificar o subconjunto da classe “Bom” para o abastecimento, sendo definição do subconjunto a partir da vazão de *Dif3* menor ou igual a 97,1% da vazão requerida, sendo o subconjunto já considerado na classe “Bom”, com representatividade de 53,47% da amostra. Assim é possível deduzir que para todos as simulações em que o valor de *C* esteve entre 95% e 99,99%, a vazão de *Dif3* é a vazão prejudicada e é mantido o abastecimento para as demais demandas.

#### 4.3.2 Árvore de decisão para a irrigação do CO3

A acurácia para a irrigação do CO3 obteve 99,7%, representando o um bom desempenho na representação das características das simulações para cada classe.

Os resultados para os modelos gerados para as confiabilidades de abastecimento urbano obtiveram precisão e *recall*, conforme os valores da Tabela 28.

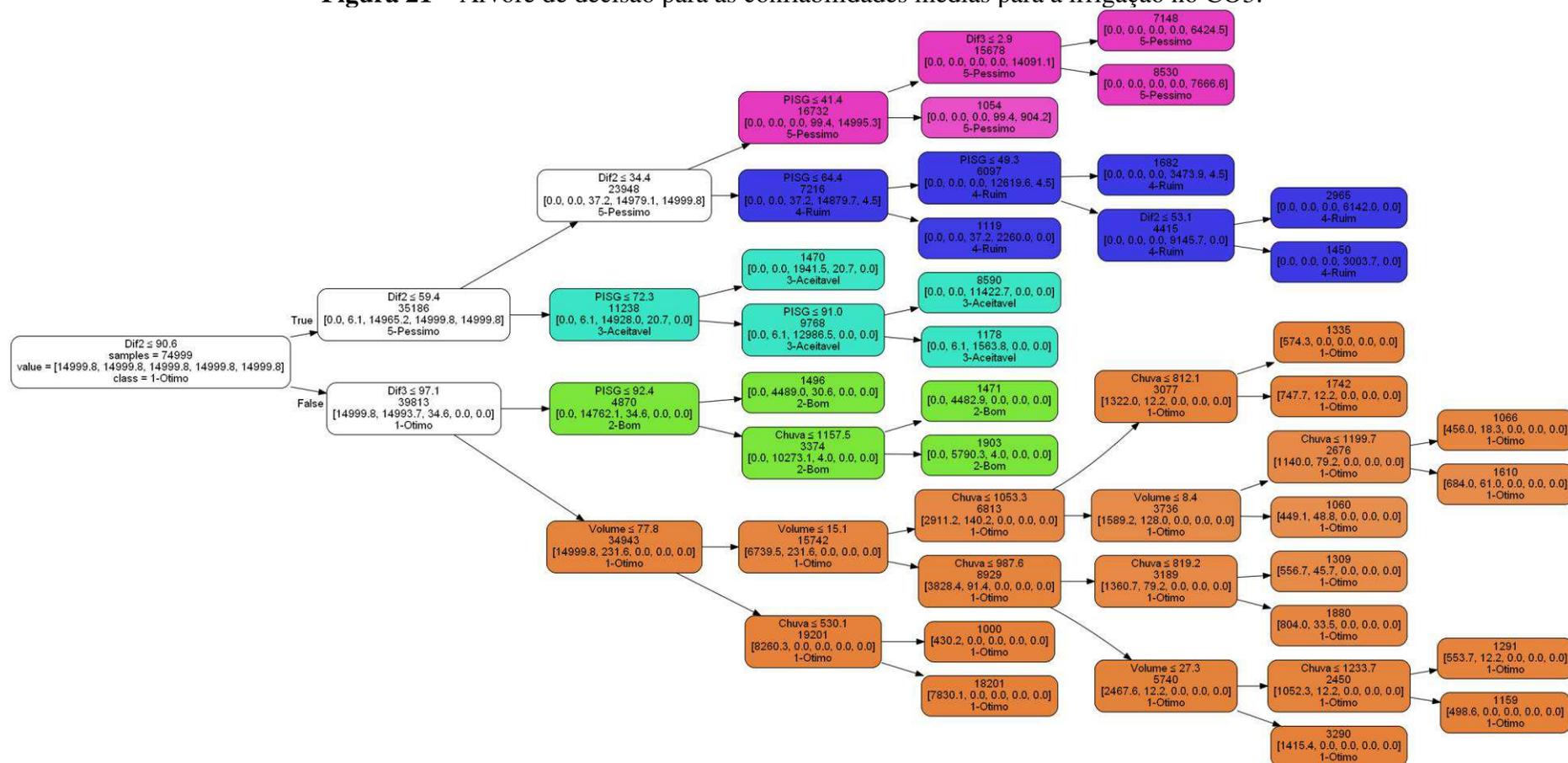
**Tabela 28** – Desempenho da árvore de decisão para a irrigação do CO3.

Classes	Precisão	<i>Recall</i>
Ótimo	1,00	1,00
Bom	1,00	0,99
Aceitável	1,00	0,99
Ruim	0,99	1,00
Péssimo	1,00	1,00

Fonte: Autor (2022).

A árvore de decisão para classificação do risco na irrigação para o CO2 está apresentada na Figura 21.

Figura 21 – Árvore de decisão para as confiabilidades médias para a irrigação no CO3.



Fonte: Autor (2022).

A árvore de decisão para a irrigação do CO3 apresentou semelhanças com a árvore do CO2, com valores de precisão e recall considerados altos, com todas as classes com valor 1,00 ou 0,99, significando que as classificações observadas foram adequadamente previstas para as simulações utilizadas no teste.

O teste de normalidade de Shapiro-Wilk das acurácias calculadas na validação cruzada, para a árvore de decisão da confiabilidade para a irrigação do CO3, resultou no p-valor de 1,00, mantendo a hipótese nula com distribuição normal dos resultados de acurácia

Os níveis de importância das variáveis independentes da árvore de decisão para a irrigação estão apresentados na Tabela 29.

**Tabela 29** – Níveis de importância das variáveis para a irrigação do CO3.

Variável independente	Nível de importância
<i>Dif1</i>	0,0%
<i>Dif2</i>	82,9%
<i>Dif3</i>	16,5%
<i>PISG</i>	0,4%
<i>AbastEA</i>	0,0%
<i>AbastSG</i>	0,0%
<i>P</i>	0,1%
<i>VfimEA</i>	0,2%

Fonte: Autor (2022).

As ramificações com o maior número de simulações de cada situação de risco para a irrigação resultaram nos diagnósticos de operação mais prováveis de acordo com as *Cer* do cenário, apresentadas na Tabela 30.

**Tabela 30** – Diagnóstico do sistema de acordo com as *Cer* para a irrigação no CO3.

Situação	Descrição	Operação
Ótimo	Confiabilidade média anual igual a 100%	$P > 530,1 \text{ mm}$ $VfimEA > 26,5\%$ da capacidade de armazenamento $0,145 \text{ m}^3/\text{s} < Dif2 \leq 0,16 \text{ m}^3/\text{s}$ $0,165 \text{ m}^3/\text{s} < Dif3 \leq 0,17 \text{ m}^3/\text{s}$
Bom	Confiabilidade média anual entre 95% e 99,99%	$Dif2 > 0,145 \text{ m}^3/\text{s}$ $0 \text{ m}^3/\text{s} < Dif3 \leq 0,165 \text{ m}^3/\text{s}$ $PISG \leq 1,294 \text{ m}^3/\text{s}$ $P > 1157,5 \text{ mm}$
Aceitável	Confiabilidade média anual entre 70% e 94,99%	$0,095 \text{ m}^3/\text{s} < Dif2 \leq 0,145 \text{ m}^3/\text{s}$ $1,012 \text{ m}^3/\text{s} < PISG \leq 1,274 \text{ m}^3/\text{s}$
Ruim	Confiabilidade média anual entre 50% e 69,99%	$0,055 \text{ m}^3/\text{s} < Dif2 \leq 0,085 \text{ m}^3/\text{s}$ $0,69 \text{ m}^3/\text{s} < PISG \leq 0,902 \text{ m}^3/\text{s}$
Péssimo	Confiabilidade média anual abaixo de 50%	$0 \text{ m}^3/\text{s} < Dif2 \leq 0,055 \text{ m}^3/\text{s}$ $Dif3 > 0,005 \text{ m}^3/\text{s}$ $0 \text{ m}^3/\text{s} < PISG \leq 0,58 \text{ m}^3/\text{s}$

Fonte: Autor (2022).

As variáveis que mais estiveram presentes na estruturação foram *Dif2*, *Dif3* e *PISG* como as variáveis mais importantes para a classificação do risco, com destaque para o *PISG* que apresentou mais participação para a irrigação do CO3 do que o CO2. Com exceção da classe “Ótimo”, o *PISG* apareceu em todas as classes no primeiro nível dos subconjuntos da amostra, sendo considerada a mais importante para a definição das ramificações com maior representatividade de simulações dentro das classes.

A irrigação do CO3 apresentou comportamento semelhante ao do CO2 sendo a classe “Ótimo” definida após o atendimento da vazão requerida de 90,6% para *Dif2* e de 97,1% para a *Dif3*. O *VfimEA* e *P* foram observados apenas dentro da classe “Ótimo”, sendo responsáveis pela divisão das ramificações de intervalos para obtenção dos intervalos para o subconjunto da amostra com maior representatividade.

Os padrões de alocação para a classe “Ótimo” obtiveram a maior quantidade de simulações para a classe, com representatividade de 52,09% da amostra classificadas “Ótimo”, acima da média de 9,09%.

A operação mais favorável considerou o *VfimEA* a partir de 77,8 hm<sup>3</sup>, com a precipitação por sua vez considerada maior que 530,1 mm, o que levou a situação “Ótimo”, sugerindo que assim como no CO2 o sistema atinge o pleno atendimento com condições hidrológicas menos favoráveis.

Para a classe “Bom”, a árvore identificou um número pequeno de subconjuntos, sendo os padrões de alocação com maior representatividade, com *Dif2* maior que 90,6% da vazão requerida, *Dif3* com vazão menor ou igual a 97,1%, *PISG* maior que 92,4% e *P* maior que 1157,5 mm. Além dos intervalos apontados para a classe, as demais ramificações também apresentaram possibilidade de ocorrência, o que pode ser indicado pelo percentual da amostra do maior subconjunto da classe com 39,08% próximo da média de 33,33% e desvio padrão 3,83%.

As demais ramificações da árvore consideraram a possibilidade de o sistema operar sem a obediência dos intervalos definidos para o *PISG* e *P*, sendo apenas as variáveis de vazão *Dif2* e *Dif3* responsáveis por agrupar o subconjunto de simulações classificadas “Bom”.

A classe “Aceitável” apresentou um mesmo número de ramificações que a classe “Bom”, divergindo apenas na observação da precipitação, ocorrendo na possibilidade de a vazão da *Dif2* ser menor ou igual a 90,6% da vazão requerida, com intervalo de atendimento a vazão do *PISG* entre 72,3% e 91%.

As classes “Ruim” e “Péssimo”, consideraram o intervalo de vazão de 34,4% da vazão da *Dif2* como determinante para a classificação, sendo as simulações com a vazão média

enviada de *Dif2* acima de 34,4% classificadas “Ruim” e as simulações com a vazão média enviada abaixo desse intervalo classificadas “Péssimo”.

O pior cenário existente classificado como “Péssimo”, com a maior quantidade de simulações no subconjunto, considerou o atendimento das vazões de *Dif2*, *Dif3* e *PISG* abaixo de 34,4%, 2,9% e 41,4% da vazão requerida respectivamente, indicando que em condições com problemas mais graves de falha no atendimento, as demandas da ramificação da classe “Péssimo” são as que possuíram o atendimento mais fragilizado.

#### 4.4 CENÁRIO OPERACIONAL 4

A partir deste cenário foram implementados mais elementos de demanda nas simulações do SEASG, onde além da adição da vazão do PISF, são adicionadas demandas que passam a não ser mais de responsabilidade pelo sistema Coremas-Mãe D’água como preconizado na Resolução conjunta ANA/IGARN/AESA N° 65/2019 – *AbastCO*. Junto com esses elementos também foi considerado o atendimento prioritário da vazão enviada ao RN.

O CO4 em geral apresentou um cenário com maior equilíbrio entre as demandas e a oferta de água variadas para cada evento sintético implementado nas milhares de simulações. Ainda assim, o fornecimento de água pelo PISF continuou impactando significativamente conforme observado nos cenários anteriormente apresentados.

A análise do indicador *C* obtido para cada uma das demandas do CO4, está apresentada na Tabela 31, contendo dados estatísticos descritivos para todas as demandas, juntamente com a Figura A16, em que estão apresentados os histogramas de distribuição do indicador *C* para as demandas rurais e urbanas do CO4.

**Tabela 31** – Estatística descritiva das confiabilidades para as demandas do CO4

	<i>AbastEA</i>	<i>AbastSG</i>	<i>AbastCO</i>	<i>Dif1</i>	<i>Dif2</i>	<i>Dif3</i>	<i>Dif4</i>	<i>PISG</i>
$\bar{X}$	100,00	100,00	61,99	57,73	45,96	45,30	40,67	47,10
<b>Med</b>	100,00	100,00	61,67	57,50	45,00	44,17	39,17	45,83
<b>Std</b>	0,00	0,00	12,13	13,14	15,95	16,08	16,79	15,84
<b>Mín</b>	100,00	100,00	28,33	22,50	8,33	8,33	5,83	8,33
<b>Máx</b>	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
<b>q1</b>	100,00	100,00	53,33	48,33	34,17	33,33	27,50	35,00
<b>q3</b>	100,00	100,00	70,00	66,67	56,67	55,83	51,67	57,50

Fonte: Autor (2021)

Com relação as demandas urbanas do sistema para o CO4, os valores de *C* apresentaram uma diferença na demanda urbana após a confluência com o Rio Piancó, estando bem abaixo das demandas urbanas do REA e do RSG, sendo também a demanda mais distante do PISF e localizada a maioria dos elementos de demanda do SEASG.

No geral, o cenário apresentou os valores bastante diferentes dos cenários anteriores não somente para as demandas urbanas, como também para as demandas rurais, com destaque para a *Difl* que nos CO2 e CO3 apresentou atendimento pleno para todas as simulações e que neste cenário obteve valores estatísticos menores que os do CO1, em que não há atuação do PISF, com uma distribuição bastante característica se comparada com as outras demandas rurais.

O CO4 possui um maior número de demandas urbanas rurais e que também possuem uma vazão considerável, o que fez com que o atendimento passasse a ser dificultado em relação a distribuição para as demandas rurais que apresentaram o pior desempenho na confiabilidade, apesar do atendimento pleno das demandas correspondentes aos reservatórios.

O comparativo entre os cenários CO2 e CO3 com relação ao CO4 mostrou que o fornecimento de água do PISF, consegue suprir satisfatoriamente as demandas urbanas, mas não é o suficiente para as demandas rurais, mostrando que a segurança de atendimento urbano pode ser rapidamente afetada com o acréscimo de mais elementos ao sistema.

#### **4.4.1 Árvore de decisão para o abastecimento urbano do CO4**

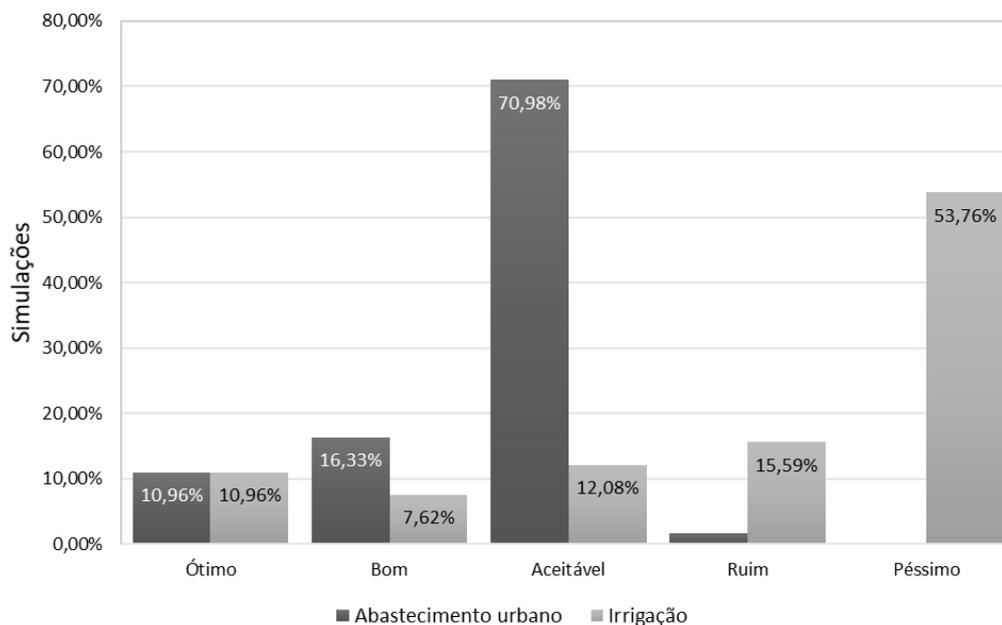
A Figura 22 apresenta o percentual médio de *C* para as simulações para os usos na irrigação e no abastecimento urbano para o CO4 respectivamente, classificadas de acordo com cada uma das classes de risco estabelecidas.

O CO4 em geral apresentou um cenário com maior equilíbrio entre as classes de risco para as demandas das simulações sintéticas implementadas. Mesmo com um desempenho para as faixas de *C* menor do que o observado nos CO2 e CO3, o fornecimento de água pelo PISF continuou impactando significativamente conforme observado nos cenários anteriormente apresentados.

A classe com maior quantidade de simulações classificadas para o abastecimento urbano foi a classe “Aceitável” com 70,98% de simulações. O cenário se manteve com valores de *C* entre 50% e 70%, valor que foi obtido devido as confiabilidades da demanda *AbastCO*, que reduziu a confiabilidade média do uso, já que as demandas *AbastEA* e *AbastSG* obtiveram

a confiabilidade em 100% em todas as simulações para o cenário. A demanda de abastecimento urbano mais distante e com maior vazão requerida portanto, impactou significativamente as classificações das simulações.

**Figura 22** – Percentuais de simulações conforme a classificação de risco para o CO4.



Fonte: Autor (2022).

Mesmo com poucas simulações sintéticas classificadas “Ótimo” se comparado com outros cenários, não houve simulações classificadas “Péssimo”, semelhante aos CO2 e CO3 que foram estruturadas com um número de demandas menor, indicando que ainda assim, o cenário possui condições de atendimento aceitáveis.

Entre as classes “Ótimo” e “Bom” a diferença dos percentuais de simulações foi a menor dentre os cenários anteriores, sendo a classificação realizada para a irrigação com uma melhor distribuição, o que configura uma pior situação para esse uso, sendo o percentual da classe “Péssimo” o maior para a irrigação, assim como o comportamento observado nos cenários anteriores.

O CO4 também obteve 99,03% de acurácia no modelo para o abastecimento, o que indica um bom desempenho na representação das características das simulações para cada classe, favorecido pela menor quantidade de classes identificadas.

Os resultados para os modelos gerados para as confiabilidades de abastecimento urbano do CO4 obtiveram precisão e *recall*, conforme os valores da Tabela 32.

**Tabela 32** – Desempenho da árvore de decisão para o abastecimento urbano do CO4.

Classes	Precisão	Recall
Ótimo	1,00	1,00
Bom	1,00	0,98
Aceitável	1,00	0,99
Ruim	0,74	1,00

Fonte: Autor (2022).

A precisão obtida par o modelo do CO4 se mostrou com desempenho máximo para todas as classes identificadas com exceção da classe “Ruim”, que apresentou o menor valor para a precisão dentre todos os cenários já apresentados, justificado pela quantidade de simulações da classe, sendo então insuficiente para a melhor precisão da classe.

Já com relação ao *recall*, os valores máximos foram obtidos para as classes “Ótimo” e “Ruim”, com destaque para a classe “Ruim” que não obteve o mesmo desempenho para a precisão, indicando que as simulações das demais classes não apresentaram subconjuntos com condições de intervalos distintos para as simulações observadas para a classe. Os menores valores de *recall* foram os das classes “Aceitável” e “Bom” respectivamente possuindo, portanto, simulações observadas nestas classes previstas para outra classificação.

Os níveis de importância das variáveis independentes obtidos a partir da árvore de decisão para o abastecimento urbano estão apresentados na Tabela 33.

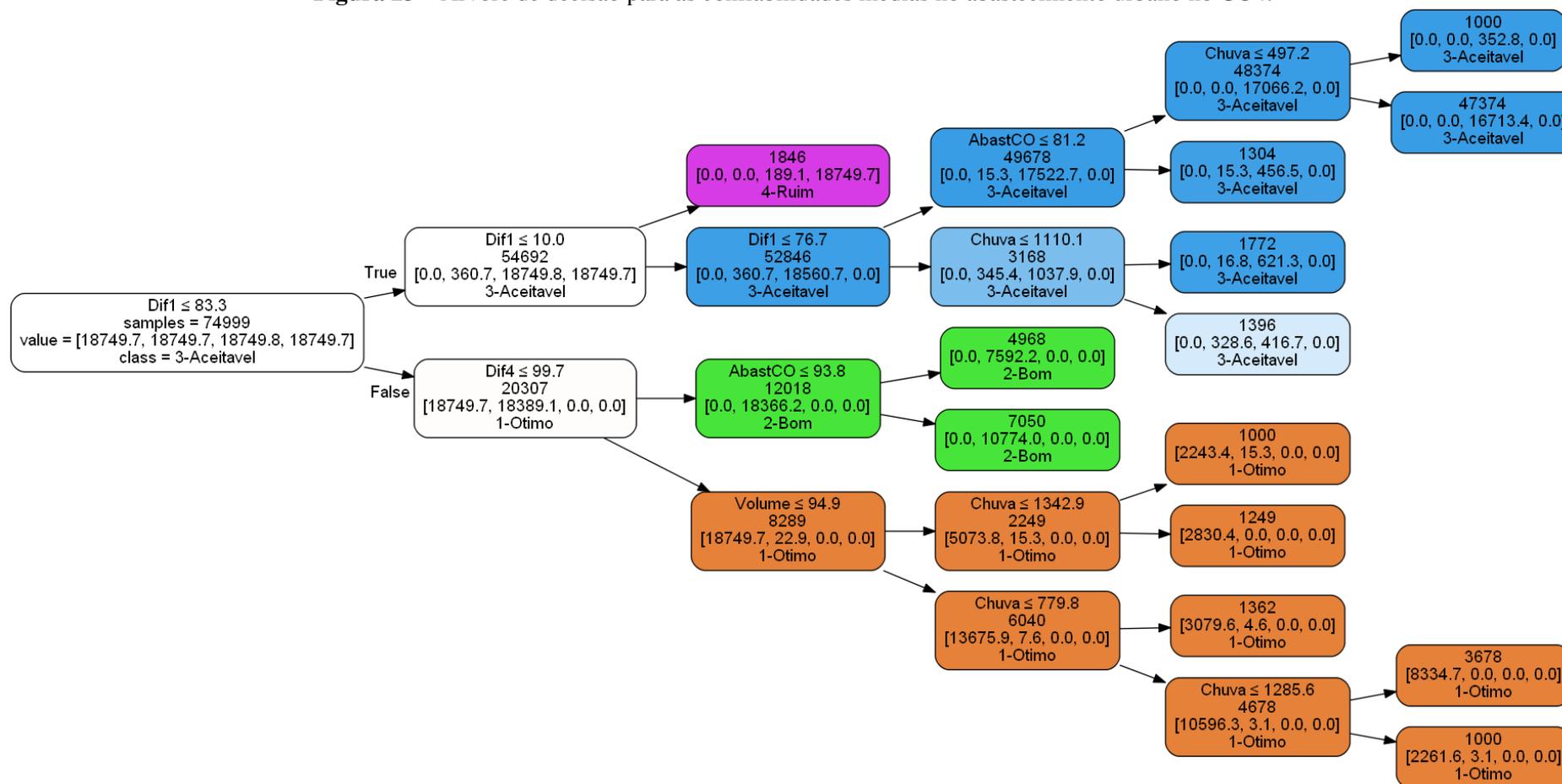
**Tabela 33** – Níveis de importância das variáveis para abastecimento urbano do CO4.

Variável independente	Nível de importância
<i>Dif1</i>	74,7%
<i>Dif2</i>	0,0%
<i>Dif3</i>	0,0%
<i>Dif4</i>	25,0%
<i>PISG</i>	0,0%
<i>AbastEA</i>	0,0%
<i>AbastSG</i>	0,0%
<i>AbastCO</i>	0,1%
<i>P</i>	0,2%
<i>VfimEA</i>	0,0%

Fonte: Autor (2022).

A árvore de decisão para classificação do risco no abastecimento urbano para o CO4 está apresentada na Figura 23.

Figura 23 – Árvore de decisão para as confiabilidades médias no abastecimento urbano no CO4.



Fonte: Autor (2022).

O teste de normalidade de Shapiro-Wilk das acurácias calculadas na validação cruzada, para a árvore de decisão da confiabilidade para o abastecimento urbano, resultou no p-valor de 0,99, mantendo a hipótese nula de que os dados apresentam distribuição normal, assim sendo possível confirmar a robustez dos modelos em detrimento da variação das variáveis de treino.

A variável com maior nível de importância para a árvore do abastecimento urbano do CO4 foi a *Dif1* com 74,7%, sendo o cenário diferente dos demais já observados, onde a variável de demanda não apareceu com percentuais de importância, a exemplo dos CO2 e CO3.

A demanda *Dif4* também apresentou influência sobre as ramificações da árvore com 25,0% de importância, sendo a segunda variável mais importante para o cenário e a que mais impacta na divisão dos subconjuntos após a variável raiz se comparados os níveis de importância das variáveis das árvores do abastecimento urbano já apresentadas.

Mesmo a variável de demanda *AbastCO* sendo importante para a determinação do valor médio de *C* do abastecimento, esta apresentou apenas 0,1% de importância para a divisão dos subconjuntos da árvore de decisão. Tal fato indica que a modelagem da árvore considerou a identificação das simulações com melhor resultado acima da definição dos limites das variáveis, sendo necessário que as demandas de abastecimento urbano sejam atendidas conforme a vazão requerida, quando não especificados seus intervalos nas ramificações da árvore.

As ramificações com o maior número de simulações de cada situação de risco para o abastecimento urbano resultaram nos diagnósticos de operação mais prováveis de acordo com as *Cer* no cenário, que foram organizadas e apresentadas na Tabela 34.

**Tabela 34** – Diagnóstico do sistema de acordo com as *Cer* do abastecimento urbano no CO4.

Situação	Descrição	Operação
Ótimo	Confiabilidade média anual igual a 100%	$P > 779,8$ mm $VfimEA > 32,32\%$ da capacidade de armazenamento $Dif1 > 0,125$ m <sup>3</sup> /s $Dif4 > 1,595$ m <sup>3</sup> /s
Bom	Confiabilidade média anual entre 95% e 99,99%	$AbastCO > 0,291$ m <sup>3</sup> /s $Dif1 > 0,125$ m <sup>3</sup> /s $0$ m <sup>3</sup> /s < $Dif4 \leq 1,595$ m <sup>3</sup> /s
Aceitável	Confiabilidade média anual entre 70% e 94,99%	$P > 497,2$ mm $0$ m <sup>3</sup> /s < $AbastCO \leq 0,252$ m <sup>3</sup> /s $0,015$ m <sup>3</sup> /s < $Dif1 \leq 0,125$ m <sup>3</sup> /s
Ruim	Confiabilidade média anual entre 50% e 69,99%	$0$ m <sup>3</sup> /s < $Dif1 \leq 0,015$ m <sup>3</sup> /s

Fonte: Autor (2022).

A estruturação da árvore apresentou as variáveis de demanda *Dif1* e *Dif4* como principais nós de divisão dos principais subconjuntos, que dividiram a árvore entre as quatro classes apresentadas. O número de variáveis escolhidas para criação das ramificações no CO4 é relativamente menor do que nos cenários anteriores, visto que o CO4 possui mais elementos de demanda.

Para a classe “Ótimo”, a variável com maior ganho de informação e que primeiro dividiu a amostra para o subconjunto da classe foi *Dif1*, sendo as simulações com valores de *Dif1* menores ou iguais a 83,3% da vazão requerida consideradas fora da classe, seguido da vazão *Dif4* que dividiu o subconjunto da amostra, sendo por sua vez considerada a classe como “Ótimo” quando os valores de *Dif4* foram maiores a 99,7% da vazão requerida.

A ramificação utilizada para a identificação do padrão de alocação para a classe “Ótimo” foi a que resultou no subconjunto com a maior quantidade de simulações, sendo consideradas as variáveis *VfimEA* e *P* como determinantes para a divisão dos subconjuntos. Para o melhor cenário identificado para a classe, o *VfimEA* deve ser maior que 94,9 hm<sup>3</sup> no REA e a *P* deve ser de pelo menos 779,8 mm, sendo o nó folha com maior representatividade com *P* menor que 1258,6 mm, que estabelece a faixa de precipitações favoráveis aos altos valores de *C*.

Os padrões de alocação para a classe “Ótimo” obtiveram a maior quantidade de simulações para a classe, sendo representada por 56,44% da amostra da classe “Ótimo”, acima da média de 20,0%. O nó folha que definiu o intervalo de *P* entre 779,8 mm e 1258,6 mm apresentou 44,37%, dentre os quatro outros nós folha alternativos.

A classe “Bom” foi a que possuiu o menor número de nós folha, com a divisão dos subconjuntos feita considerando apenas a *Dif1* e *Dif4*, onde *Dif1* segue com o mesmo intervalo da classe “Ótimo” acima de 83,3%, porém com *Dif4* sendo atendida com vazões menores ou iguais a 99,7%. A definição da ramificação com maior representatividade para a classe foi feita a partir da observação da *AbastCO*, sendo enviado no mínimo 93,8% da vazão requerida.

A classe apresentou apenas duas ramificações como possibilidades de situação para o CO4, sendo a ramificação com os padrões de alocação mais prováveis representada por 58,66% da amostra. A segunda possibilidade de operação para o cenário considerou o atendimento da demanda *AbastCO* menor ou igual a 93,8%, desde que sejam obedecidos os limites estabelecidos para as demandas *Dif1* e *Dif4*.

Para a classe “Aceitável” foi identificada a mesma quantidade de nós folha que a classe “Ótimo”, porém considerando a variável *AbastCO* para definição do subconjunto mais

representativo. O intervalo de operação para a *Dif1* foi entre 10,0% e 83,3% da vazão requerida, contudo a subdivisão da ramificação mais representativa considerou o intervalo entre 10,0% e 76,4%. A regra de operação definida para a demanda *AbastCO* considerou a vazão menor ou igual a 81,2%, atrelado ao valor da precipitação sendo pelo menos maior que 497,2 mm.

A ramificação escolhida para identificação dos padrões de alocação possuiu 89,65% da amostra da classe, sendo o conjunto de intervalos com maior representatividade, acima da média que foi de 20,0%.

A classe “Ruim” foi a classe com menor subdivisão e envolveu apenas a demanda *Dif1* na definição do subconjunto da classe. Também foi a classe com a menor quantidade de simulações classificadas, sendo representada por apenas 2,46% das simulações. Foi possível observar que a determinação dos intervalos para a classe esteve diretamente relacionada com o tamanho da amostra, sendo a classe “Ruim” definida para todas as simulações com vazão de *Dif1* abaixo de 10,0%.

Tal resultado significa que quando a situação do sistema se encontra em uma situação considerada ruim no que diz respeito a confiabilidade, a vazão média de *Dif1* não consegue alcançar valores maiores do que 10,0%, sendo essa característica considerada a mais comum na simulação, podendo ser subentendido que ao fornecer menos que essa quantidade para *Dif1*, não haveria disponibilidade suficiente para as demandas de abastecimento.

#### 4.4.2 Árvore de decisão para a irrigação do CO4

Os resultados para os modelos gerados para as confiabilidades de abastecimento urbano obtiveram precisão e *recall*, conforme os valores da Tabela 35.

**Tabela 35** – Desempenho da árvore de decisão para a irrigação do CO4.

Classes	Precisão	<i>Recall</i>
Ótimo	1,00	1,00
Bom	1,00	0,99
Aceitável	0,99	0,94
Ruim	0,88	1,00
Péssimo	1,00	0,98

Fonte: Autor (2022).

A árvore de decisão para classificação do risco na irrigação para o CO4 está apresentada na Figura 24.



A acurácia para a irrigação do CO4 obteve 96,81%, representando o um bom desempenho na representação das características das simulações para cada classe.

Os resultados para a previsão indicaram o melhor desempenho na classificação da classe “Ótimo”, “Bom” e “Péssimo”, seguidas das classes “Aceitável”, sendo a classe “Ruim” aquela que possuiu a menor precisão, o que sugere a menor homogeneidade das amostras causada pela pequena quantidade de simulações da amostra para a classe.

Para o *recall* a classe “Ótimo” e “Aceitável” obtiveram o melhor resultado, mostrando que estas são mais importantes para calibração da previsão, apesar de não haver diferença significativa se comparado ao das demais classes, mesmo sendo classes com poucas simulações.

O teste de normalidade de Shapiro-Wilk das acurácias calculadas na validação cruzada, para a árvore de decisão da confiabilidade para a irrigação do CO4, resultou no p-valor de 1,0, mantendo a hipótese nula com distribuição normal dos resultados de acurácia.

Os níveis de importância das variáveis independentes da árvore de decisão para a irrigação estão apresentados na Tabela 36.

**Tabela 36** – Níveis de importância das variáveis para a irrigação do CO4.

Variável independente	Nível de importância
<i>Dif1</i>	1,5%
<i>Dif2</i>	37,8%
<i>Dif3</i>	42,0%
<i>Dif4</i>	17,6%
<i>PISG</i>	1,0%
<i>AbastEA</i>	0,0%
<i>AbastSG</i>	0,0%
<i>AbastCO</i>	0,0%
<i>P</i>	0,0%
<i>VfimEA</i>	0,0%

Fonte: Autor (2022).

As ramificações com o maior número de simulações de cada situação de risco para a irrigação resultaram em nos diagnósticos de operação mais prováveis de acordo com as *Cer* para o cenário, apresentadas na Tabela 37.

A árvore de decisão para a irrigação do CO4 apresentou uma estruturação semelhante ao que foi modelado para o abastecimento urbano, porém com a realização de mais divisões dos subconjuntos para a classe “Ruim”. As principais variáveis utilizadas para a divisão primária das classes na árvore foram as vazões médias das áreas difusas, *Dif2*, *Dif3* e *Dif4*, com a presença de intervalos para o *PISG* na classe “Aceitável” e mais uma vez sendo observados os valores de *VfimEA* e *P* na classe “Ótimo”.

**Tabela 37** – Diagnóstico do sistema de acordo com as *Cer* da irrigação no CO4.

Situação	Descrição	Operação
Ótimo	Confiabilidade média anual igual a 100%	$P > 779,8$ mm $VfimEA > 32,32\%$ da capacidade de armazenamento $0,155 \text{ m}^3/\text{s} < Dif3 \leq 0,17 \text{ m}^3/\text{s}$ $1,595 \text{ m}^3/\text{s} < Dif4 \leq 1,6 \text{ m}^3/\text{s}$
Bom	Confiabilidade média anual entre 95% e 99,99%	$Dif3 > 0,155 \text{ m}^3/\text{s}$ $1,405 \text{ m}^3/\text{s} < Dif4 \leq 1,595 \text{ m}^3/\text{s}$
Aceitável	Confiabilidade média anual entre 70% e 94,99%	$Dif2 > 0,115 \text{ m}^3/\text{s}$ $0 \text{ m}^3/\text{s} < Dif3 \leq 0,155 \text{ m}^3/\text{s}$ $1,056 \text{ m}^3/\text{s} < PISG \leq 1,275 \text{ m}^3/\text{s}$
Ruim	Confiabilidade média anual entre 50% e 69,99%	$0 \text{ m}^3/\text{s} < Dif1 \leq 0,115 \text{ m}^3/\text{s}$ $0,075 \text{ m}^3/\text{s} < Dif2 \leq 0,115 \text{ m}^3/\text{s}$ $0 \text{ m}^3/\text{s} < Dif3 \leq 0,155 \text{ m}^3/\text{s}$ $Dif4 > 0,734 \text{ m}^3/\text{s}$ $0,755 \text{ m}^3/\text{s} < PISG \leq 1,044 \text{ m}^3/\text{s}$
Péssimo	Confiabilidade média anual abaixo de 50%	$0 \text{ m}^3/\text{s} < Dif2 \leq 0,075 \text{ m}^3/\text{s}$ $0 \text{ m}^3/\text{s} < Dif3 \leq 0,155 \text{ m}^3/\text{s}$ $0 \text{ m}^3/\text{s} < Dif4 \leq 0,666 \text{ m}^3/\text{s}$

Fonte: Autor (2022).

Para a classe “Ótimo”, a variável com maior ganho de informação e que primeiro dividiu a amostra para o subconjunto da classe foi *Dif3*, sendo as simulações com valores maiores que 91,2% da vazão requerida pertencentes a classe, seguido da vazão *Dif1* que dividiu o subconjunto da amostra, sendo por sua vez considerado “Ótimo” quando os valores de *Dif4* foram maiores que 99,7% da vazão requerida.

A partir dos intervalos definidos para *Dif3* e *Dif4* as simulações já passaram a ser consideradas na classe “Ótimo”, porém o maior subconjunto da amostra foi obtido a partir da observação de *VfimEA* que deve ser pelo menos de 94,9 hm<sup>3</sup> e *P* pelo menos igual a 779,8 mm, configurando a situação com maior representatividade de simulações para a classe.

Os padrões de alocação para a classe “Ótimo” obtiveram a maior quantidade de simulações para a classe, sendo representada por 56,44% da amostra classificada como “Ótimo”, acima da média de 20,0%.

A classe “Bom” foi a classe com menor quantidade de ramificações para identificação dos padrões de alocação. A variável com maior influência foi a vazão da *Dif4* dividindo as simulações com vazão acima de 87,8% e menor ou igual a 99,7%, após a divisão do subconjunto com vazão de *Dif3* acima de 91,2% da vazão requerida, sendo o nó folha com o subconjunto amostral possuindo 80,39% das simulações classificadas “Bom”, acima da média de 50% da representatividade das respostas para a classe.

A classe “Aceitável” apresentou uma quantidade de variáveis maiores do que as classes com intervalos maiores de  $C$ . As variáveis observadas para a definição do subconjunto da amostra da classe foram  $Dif2$ ,  $Dif3$  e  $PISG$ , sendo esta última utilizada nos níveis finais para definição do nó folha mais representativo. A principal divisão para definição do subconjunto amostral ocorreu quando  $Dif2$  entre 46,9% e 71,9%, juntamente com  $Dif3$  abaixo de 91,2% da vazão requerida.

A vazão do  $PISG$  entre 75,4% e 91,1%, possibilitou a definição do nó folha com a maior quantidade de simulações enquadradas na classe. Os padrões de alocação identificados para a classe “Aceitável” apresentaram 75,81% das simulações do subconjunto amostral.

A partir da divisão das simulações através da observação da vazão da  $Dif2$ , as simulações já puderam ser classificadas “Ruim”, sendo o subconjunto da classe definido para vazões abaixo de 71,9% do requerido. A classe “Ruim”, foi a que possuiu uma maior quantidade de nós folha assim como a maior quantidade de intervalos observados para as variáveis do modelo.

As variáveis de maior destaque foram  $Dif4$  e  $PISG$ , ambas com as maiores vazões demandadas, sendo recursivamente observadas nos últimos níveis da árvore, o que sugere que são as demandas mais sensíveis ao estado de colapso do sistema, na eminência do cenário apresentar 50% das demandas de irrigação com falha de atendimento.

Mesmo com a maior quantidade de ramificações, o nó folha considerado no diagnóstico do comportamento das variáveis obteve uma percentagem significativa, sendo o nó folha mais representativo com 45,62% das simulações da amostra, acima da média de 14,29%.

A classe “Péssimo”, foi definida como sendo o subconjunto de simulações com os valores de  $Dif3$  abaixo de 91,2% e  $Dif2$  abaixo de 46,9% da vazão requerida, sugerindo que esta seja a característica mais presente para as situações em que o sistema apresente valores médios de  $C$  abaixo de 50% na irrigação.

A partir do subconjunto com  $Dif2$  menor ou igual a 46,9%, as simulações já passaram a ser classificadas “Péssimo”, sendo  $Dif4$  abaixo de 41,6% a ramificação com maior quantidade de simulações para a classe, com 97,34% da representatividade das respostas para a classe.

#### 4.5 CENÁRIO OPERACIONAL 5

Neste cenário operacional, foram implementados os mesmos elementos de demanda do CO4, relacionadas as demandas urbanas e rurais que em princípio foram consideradas de

responsabilidade do sistema Coremas-Mãe D'água e com fornecimento da vazão do PISF, porém diferente do cenário anterior, o atendimento da vazão enviada para o Estado do Rio Grande do Norte, foi simulado com a mesma prioridade das demandas urbanas do SEASG.

Os resultados obtidos para os indicadores no CO5 foram praticamente idênticos aos dados do CO4, tanto em relação as características estatísticas quanto em relação ao comportamento das distribuições.

A análise do indicador *C* obtido para cada uma das demandas do CO5, está apresentada na Tabela 38, contendo dados estatísticos descritivos para todas as demandas, juntamente com a Figura A19, em que estão apresentados os histogramas de distribuição do indicador *C* para as demandas rurais e urbanas do CO5.

**Tabela 38** - Estatística descritiva das confiabilidades para as demandas do CO5

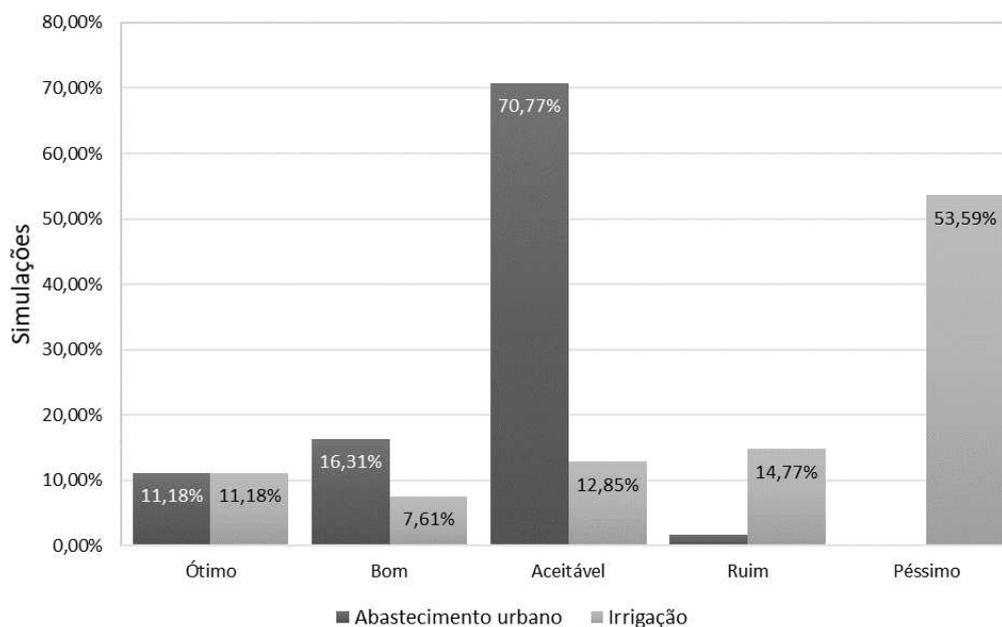
	<i>AbastEA</i>	<i>AbastSG</i>	<i>AbastCO</i>	<i>Dif1</i>	<i>Dif2</i>	<i>Dif3</i>	<i>Dif4</i>	<i>PISG</i>
$\bar{X}$	100,00	100,00	62,07	57,91	46,15	45,48	40,82	47,23
<b>Med</b>	100,00	100,00	61,67	57,50	45,00	44,17	39,17	46,67
<b>Std</b>	0,00	0,00	12,18	13,16	16,01	16,14	16,88	15,91
<b>Mín</b>	100,00	100,00	28,33	22,50	8,33	8,33	5,83	8,33
<b>Máx</b>	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
<b>q1</b>	100,00	100,00	53,33	48,33	34,17	33,33	27,50	35,00
<b>q3</b>	100,00	100,00	70,83	66,67	56,67	56,67	52,50	58,33

Fonte: Autor (2021)

A confiabilidade obtida para o cenário indica a forte influência da disponibilidade hídrica no sistema, além do estabelecimento das prioridades de atendimento, pois apesar da maior prioridade da demanda urbana após a confluência com o Rio Piancó, esta não foi plenamente atendida em todas as simulações realizadas, contudo mantendo-se com melhor desempenho se comparada com as demandas rurais.

#### 4.5.1 Árvore de decisão para o abastecimento urbano do CO5

A Figura 25 apresenta o percentual médio de *C* para as simulações para os usos na irrigação e no abastecimento urbano para o CO5 respectivamente, classificadas de acordo com cada uma das classes de risco estabelecidas.

**Figura 25** – Percentuais de simulações conforme a classificação de risco para o CO5.

Fonte: Autor (2022).

O CO5 apresentou um cenário com quantidades de simulações muito similares com as classes de risco para as demandas do CO4. Assim como no CO4, o desempenho para as faixas de  $C$  menor do que o observado foi inferior ao que foi avaliado nos CO2 e CO3, porém o fornecimento de água pelo PISF continuou impactando significativamente a distribuição de água.

A classe com maior quantidade de simulações classificadas para o abastecimento urbano também foi a classe “Aceitável” com 70,77% de simulações. O cenário se manteve com valores de  $C$  entre 50% e 70%. É importante salientar que a semelhança dos valores percentuais de simulações classificados para o cenário, apesar de possuir tais percentuais, não há indicativo de que as árvores e os comportamentos das variáveis apresentem estruturas similares.

Para o abastecimento urbano o CO5 também não apresentou a maior quantidade de simulações sintéticas classificadas “Ótimo”, se considerados os percentuais de outras classes. A classe “Péssimo”, assim como no CO4 não apresentou simulações, já para o uso na irrigação a quantidade de simulações pode ser equiparada com a classe de outros cenários operacionais. O CO5 obteve 99,11% de acurácia no modelo para o abastecimento, o que indica um bom desempenho na representação das características das simulações para cada classe.

Os resultados para os modelos gerados para as confiabilidades de abastecimento urbano do CO5 obtiveram precisão e *recall*, conforme os valores da Tabela 39.

**Tabela 39** – Desempenho da árvore de decisão para o abastecimento urbano do CO5.

Classes	Precisão	Recall
Ótimo	1,00	1,00
Bom	1,00	0,98
Aceitável	1,00	0,99
Ruim	0,74	1,00

Fonte: Autor (2022).

A precisão obtida par o modelo do CO5 se mostrou com desempenho máximo para todas as classes identificadas com exceção da classe “Ruim”, juntamente com o do CO4 um valor considerado baixo do que foi encontrado para as demais classes, sendo causado pela quantidade de simulação da amostra da classe. O valor do *recall* se manteve no máximo para as classes “Ótimo” e “Ruim”, sendo as classes “Bom” e “Aceitável”, resultado compatível com as quantidades de simulações da amostra e também pelo maior grau de heterogeneidade dos dados sintéticos implementados.

Os níveis de precisão e recall encontrados para o modelo do CO5, apesar de possuir um baixo desempenho para o estado “Ruim”, não interferem diretamente na divisão dos subconjuntos das classes, devido a acurácia do modelo. Como nenhum modelo passou por processos de balanceamento em que fossem preservadas as características do sistema em diferentes situações, os valores de previsão poderão futuramente alimentarem o modelo com o intuito de tornar mais especificados para o cliente.

Os níveis de importância das variáveis independentes obtidos a partir da árvore de decisão para o abastecimento urbano estão apresentados na Tabela 40.

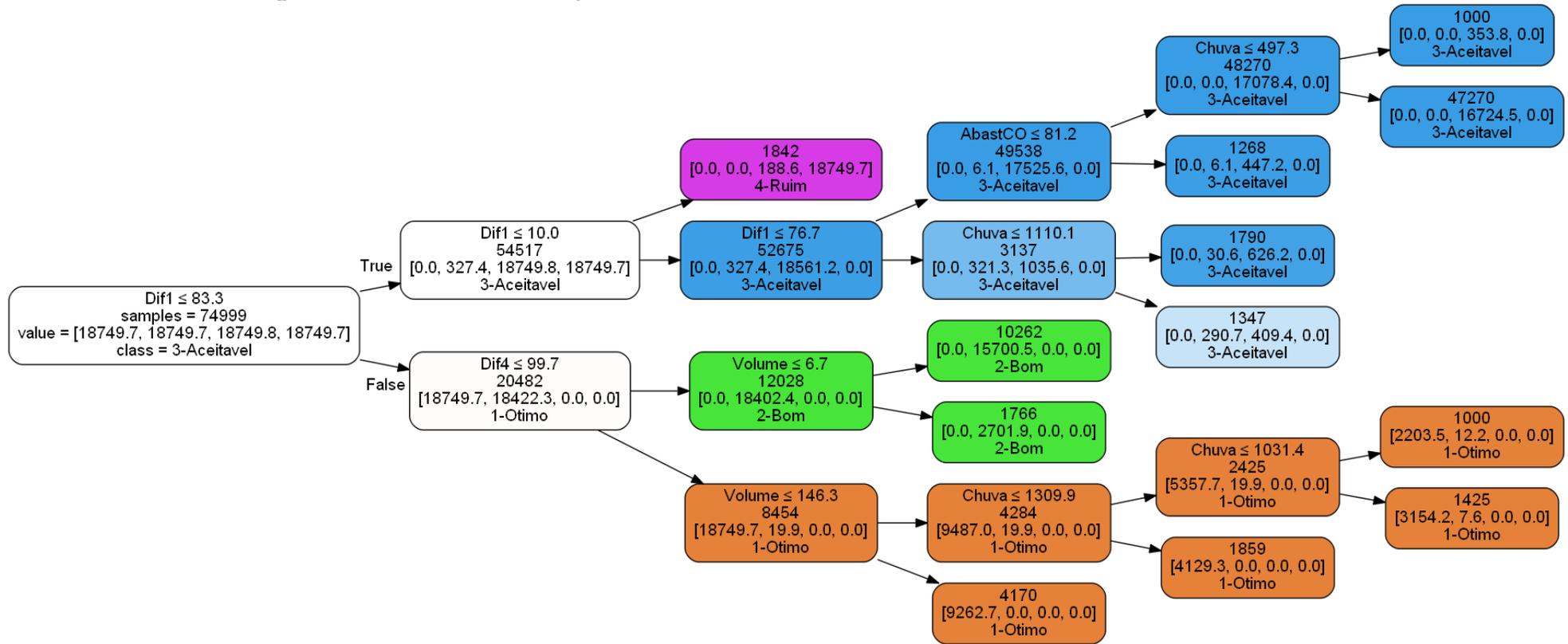
**Tabela 40** – Níveis de importância das variáveis para abastecimento urbano do CO5.

Variável independente	Nível de importância
<i>Dif1</i>	74,7%
<i>Dif2</i>	0,0%
<i>Dif3</i>	0,0%
<i>Dif4</i>	25,1%
<i>PISG</i>	0,0%
<i>AbastEA</i>	0,0%
<i>AbastSG</i>	0,0%
<i>AbastCO</i>	0,0 %
<i>P</i>	0,2%
<i>VfimEA</i>	0,0%

Fonte: Autor (2022).

A árvore de decisão para classificação do risco no abastecimento urbano para o CO5 está apresentada na Figura 26.

Figura 26 – Árvore de decisão para as confiabilidades médias no abastecimento urbano no CO5.



Fonte: Autor (2022).

O teste de normalidade de Shapiro-Wilk das acurácias calculadas na validação cruzada, para a árvore de decisão da confiabilidade para o abastecimento urbano, resultou no p-valor de 1,00, mantendo a hipótese nula de que os dados apresentam distribuição normal, assim sendo possível confirmar a robustez dos modelos em detrimento da variação das variáveis de treino.

Na árvore de decisão desenvolvida para o abastecimento urbano apresentou apenas quatro variáveis como sendo as mais importantes para a estrutura da árvore e divisão dos subconjuntos das amostras agrupadas para cada classe, sendo estas a *Dif1*, *Dif4*, *VfimEA* e *P*. A variável com maior nível de importância para a árvore do abastecimento urbano do CO5 foi a *Dif1* com 74,7%, percentual igual ao encontrado para CO4, considerada como sendo a variável de vazão com características melhor definidas e com melhores condições de percepção do tipo de situação de risco.

Assim como nos demais cenários operacionais, as variáveis hidrológicas de volume de reservatório e precipitação foram utilizadas mais uma vez para a identificação das ramificações com os padrões de alocação para a classe “Ótimo”. As variáveis *VfimEA* e *P*. Além da importância indicada para a variável *Dif1*, outra variável que subdividiu a classificação foi *Dif4*. A demanda *Dif4* também apresentou influência sobre as ramificações da árvore, porém em apenas uma única classe, sendo está a classe “Aceitável”.

A árvore de decisão para o risco no abastecimento do CO5 apresentou uma quantidade maior de subdivisões. A *VfimEA* e *P*, estas variáveis estiveram presentes pela primeira vez nas divisões do subconjunto amostral de outras classes, compondo as ramificações da classe “Aceitável”.

Outro fator importante foi a diferença observada entre este e o cenário anterior, apresentando características de subdivisão foram as variáveis *Dif4* e a precipitação. Mesmo que as demais variáveis mantenham os valores de vazão estabelecidos conforme o desempenho apontado pelo modelo, foi possível observar que ainda sim pode haver diferenças significativas na estrutura da árvore.

A principal diferença observada para os modelos desenvolvidos tanto para o CO4 como o CO5, está apenas na quantidade de intervalos observados para a operação, sendo neste cenário considerado o *VfimEA* como sendo de maior importância que a precipitação.

As ramificações com o maior número de simulações de cada situação de risco para o abastecimento urbano resultaram nos diagnósticos de operação mais prováveis de acordo com as *Cer* para o cenário, que foram organizadas e apresentadas na Tabela 41.

**Tabela 41** – Diagnóstico do sistema de acordo com as *Cer* do abastecimento urbano no CO5.

Situação	Descrição	Operação
Ótimo	Confiabilidade média anual igual a 100%	$V_{fimEA} > 49,83\%$ da capacidade de armazenamento $0,125 \text{ m}^3/\text{s} < Dif1 \leq 0,15 \text{ m}^3/\text{s}$ $1,595 \text{ m}^3/\text{s} < Dif4 \leq 1,6 \text{ m}^3/\text{s}$
Bom	Confiabilidade média anual entre 95% e 99,99%	$0\% < V_{fimEA} \leq 2,28\%$ $0,125 \text{ m}^3/\text{s} < Dif1 \leq 0,15 \text{ m}^3/\text{s}$ $0 \text{ m}^3/\text{s} < Dif4 \leq 1,595 \text{ m}^3/\text{s}$
Aceitável	Confiabilidade média anual entre 70% e 94,99%	$P > 497,3 \text{ mm}$ $0 \text{ m}^3/\text{s} < AbastCO \leq 0,252 \text{ m}^3/\text{s}$ $0,015 \text{ m}^3/\text{s} < Dif1 \leq 0,125 \text{ m}^3/\text{s}$
Ruim	Confiabilidade média anual entre 50% e 69,99%	$0 \text{ m}^3/\text{s} < Dif1 \leq 0,015 \text{ m}^3/\text{s}$

Fonte: Autor (2022).

Para a classe “Ótimo”, a variável com maior ganho de informação e que primeiro dividiu a amostra para o subconjunto da classe foi *Dif1*, sendo as simulações com valores de *Dif1* menores ou iguais a 83,3% da vazão requerida consideradas fora da classe, seguido da vazão *Dif4* que dividiu o subconjunto da amostra, sendo por sua vez considerada a classe como “Ótimo” quando os valores de *Dif4* foram maiores a 99,7% da vazão requerida.

A ramificação utilizada para a identificação dos padrões de alocação para a classe “Ótimo” do CO5 resultou no subconjunto com a maior quantidade de simulações, sendo consideradas as variáveis *VfimEA* e *P* como determinantes para a divisão dos subconjuntos. Tal resultado manteve um comportamento semelhante já observadas no CO4, contudo o *VfimEA* para o sistema obteve valores de intervalo distintos, onde o *VfimEA* deve ser maior que 146,3 hm<sup>3</sup> no REA.

Apenas os demais nós folha para a classe consideraram a observação dos valores de *P*, sendo para a situação de risco “Ótimo” a avaliação dos intervalos de 146,3 hm<sup>3</sup> sendo a ramificação mais representativa aquela que deve ser de pelo menos máxima precipitação de 1031,4 mm e mínima de 1309,9 mm, sendo o nó folha com maior representatividade quanto aos intervalos de *P*, sendo o nó folha com maior quantidade de simulações e com precipitação.

Os padrões de alocação para a classe “Ótimo” que obtiveram a maior quantidade de simulações para a classe, foram representadas por 50,67% da amostra da classe “Ótimo”, acima da média de 20,0%. O nó folha que definiu o intervalo de *P* entre 1031,4 mm e 1309,9 mm apresentou 16,86%, dentre os quatro outros nós folha alternativos.

A classe “Bom” foi a que possuiu o menor número de nós folha, com a divisão dos subconjuntos feita considerando apenas a *Dif1* e *Dif4*, assim como no cenário anterior. A vazão

*Dif1* para a classe se manteve acima de 83,3%, porém com *Dif4* sendo atendida com vazões menores ou iguais a 99,7%.

A classe também apresentou o *VfimEA* como variável importante para a avaliação das possibilidades de conforto ambiental e saneamento rural estão correlacionados. Apenas duas ramificações como possibilidades de situação para o CO4, sendo escolhida a ramificação representada por 85,31% da amostra. A segunda possibilidade de operação para o cenário considerou o atendimento da demanda *VfimEA* menor ou igual a 6,7 hm<sup>3</sup>, desde que sejam obedecidos os limites estabelecidos para as demandas *Dif1* e *Dif4*.

Para a classe “Aceitável” foi identificada uma maior quantidade de nós folha do que as demais classes já mencionadas. O intervalo de operação para a *Dif1* foi entre 10,0% e 83,3% da vazão requerida, contudo a subdivisão da ramificação mais representativa considerou o intervalo entre 10,0% e 76,4%. A regra de operação definida, considerou *AbastCO* com vazão menor ou igual a 81,2%, além da avaliação das precipitações mínimas e máxima respectivamente, sendo necessário para o atendimento “Aceitável” que a precipitação seja acima de 497,3 mm.

A ramificação escolhida para identificação dos padrões de alocação possuiu 89,73% da amostra da classe, sendo o conjunto de intervalos com maior representatividade, acima da média que foi de 20,0%.

Assim como para CO4, a classe “Ruim” foi a classe com menor subdivisão e envolveu apenas a demanda *Dif1* na definição do subconjunto da classe. Também foi a classe com a menor quantidade de simulações classificadas, sendo representada por apenas 2,51% das simulações, sendo a vazão de *Dif1* abaixo de 10,0%.

#### 4.5.2 Árvore de decisão para a irrigação do CO5

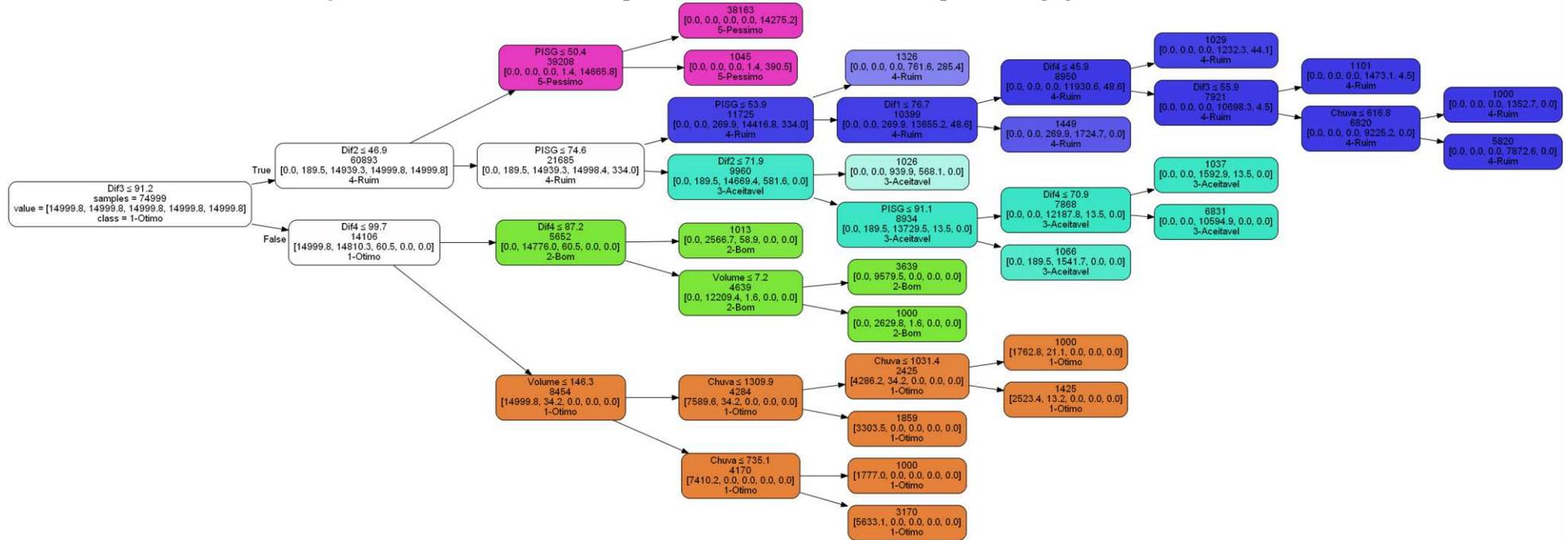
Os resultados para os modelos gerados para as confiabilidades de abastecimento urbano obtiveram precisão e *recall*, conforme os valores da Tabela 42. A árvore de decisão para classificação do risco na irrigação para o CO5 está apresentada na Figura 27.

**Tabela 42** – Desempenho da árvore de decisão para a irrigação do CO5.

Classes	Precisão	<i>Recall</i>
Ótimo	1,00	1,00
Bom	1,00	0,99
Aceitável	0,95	0,98
Ruim	0,91	0,96
Péssimo	1,00	0,98

Fonte: Autor (2022).

Figura 27 – Árvore de decisão para as confiabilidades médias para a irrigação no CO5.



Fonte: Autor (2022).

A acurácia para a irrigação do CO5 obteve 96,78%, representando o um bom desempenho na representação das características das simulações para cada classe.

Os resultados para a previsão indicaram o melhor desempenho na classificação da classe “Ótimo”, “Bom” e “Péssimo”, sendo as classes “Aceitável” e “Ruim” aquela que possuíram a menor precisão, o que sugere a menor homogeneidade das amostras causada pela pequena quantidade de simulações da amostra para as classes.

Para o *recall*, apenas a classe “Ótimo” obteve desempenho máximo, sendo os valores do *recall* das demais classes entre 0,96 e 0,99, não indicando uma perda significativa para o indicador. Os resultados mostraram que no geral o modelo para os valores médios de *C* da irrigação, apesar da menor acurácia obtida em comparação com o CO4, o modelo manteve o potencial de classificação.

O teste de normalidade de Shapiro-Wilk das acurácias calculadas na validação cruzada, para a árvore de decisão da confiabilidade para a irrigação do CO5, resultou no p-valor de 0,98, mantendo a hipótese nula com distribuição normal dos resultados de acurácia.

Os níveis de importância das variáveis independentes da árvore de decisão para a irrigação estão apresentados na Tabela 43.

**Tabela 43** – Níveis de importância das variáveis para a irrigação do CO5.

Variável independente	Nível de importância
<i>Dif1</i>	0,5%
<i>Dif2</i>	24,1%
<i>Dif3</i>	41,9%
<i>Dif4</i>	17,6%
<i>PISG</i>	1,0%
<i>AbastEA</i>	0,0%
<i>AbastSG</i>	0,0%
<i>AbastCO</i>	0,0%
<i>P</i>	0,0%
<i>VfimEA</i>	0,0%

Fonte: Autor (2022).

As variáveis com maior nível de importância que foram utilizadas para a divisão dos subconjuntos das amostras das classes para a irrigação foram as *Dif1*, *Dif2*, *Dif3*, *Dif4* e *PISG*, não havendo participação das demandas de abastecimento urbano, o volume final do REA e a precipitação.

Dentre estas, a variável com maior percentual de importância foi a *Dif3* com 41,9% de importância, sendo apresentada na árvore imediatamente no primeiro nível, seguida das divisões feitas a partir das variáveis *Dif2* e *Dif4*.

Mesmo sem apresentarem níveis de importância acima de 0,0%,  $V_{fimEA}$  e  $P$  também foram observadas nas ramificações da classe “Ótimo”, sendo utilizadas na definição dos nós folha de maior representatividade, assim como fora percebido nos demais cenários.

As ramificações com o maior número de simulações de cada situação de risco para a irrigação resultaram nos diagnósticos de operação mais prováveis de acordo com as  $Cer$  para o cenário, apresentadas na Tabela 44.

**Tabela 44** – Diagnóstico do sistema de acordo com as  $Cer$  da irrigação no CO5.

Situação	Descrição	Operação
Ótimo	Confiabilidade média anual igual a 100%	$P > 735,1$ mm $V_{fimEA} > 49,83\%$ da capacidade de armazenamento $0,155 \text{ m}^3/\text{s} < Dif3 \leq 0,17 \text{ m}^3/\text{s}$ $1,595 \text{ m}^3/\text{s} < Dif4 \leq 1,6 \text{ m}^3/\text{s}$
Bom	Confiabilidade média anual entre 95% e 99,99%	$V_{fimEA} > 2,45\%$ da capacidade de armazenamento $Dif3 > 0,155 \text{ m}^3/\text{s}$ $1,395 \text{ m}^3/\text{s} < Dif4 \leq 1,595 \text{ m}^3/\text{s}$
Aceitável	Confiabilidade média anual entre 70% e 94,99%	$Dif2 > 0,115 \text{ m}^3/\text{s}$ $0 \text{ m}^3/\text{s} < Dif3 \leq 0,155 \text{ m}^3/\text{s}$ $Dif4 > 1,134 \text{ m}^3/\text{s}$ $1,044 \text{ m}^3/\text{s} < PISG \leq 1,275 \text{ m}^3/\text{s}$
Ruim	Confiabilidade média anual entre 50% e 69,99%	$P > 616,8$ mm $0 \text{ m}^3/\text{s} < Dif1 \leq 0,115 \text{ m}^3/\text{s}$ $Dif2 > 0,075 \text{ m}^3/\text{s}$ $0,095 \text{ m}^3/\text{s} < Dif3 \leq 0,155 \text{ m}^3/\text{s}$ $Dif4 > 0,734 \text{ m}^3/\text{s}$ $0 \text{ m}^3/\text{s} < Dif2 \leq 0,075 \text{ m}^3/\text{s}$
Péssimo	Confiabilidade média anual abaixo de 50%	$0 \text{ m}^3/\text{s} < Dif3 \leq 0,155 \text{ m}^3/\text{s}$ $Dif4 > 0 \text{ m}^3/\text{s}$ $0 \text{ m}^3/\text{s} < PISG \leq 0,706 \text{ m}^3/\text{s}$

Fonte: Autor (2022).

Para a classe “Ótimo”, a variável com maior ganho de informação e que primeiro dividiu a amostra para o subconjunto da classe foi  $Dif3$ , sendo as simulações com valores maiores que 91,2% da vazão requerida pertencentes a classe, seguido da vazão  $Dif4$  que dividiu o subconjunto da amostra, sendo por sua vez considerado “Ótimo” com vazões maiores que 99,7% da vazão requerida. O subconjunto amostral da classe com o nó folha de maior representatividade para a classe foi obtido a partir da observação de  $V_{fimEA}$  que deve ser pelo menos de 146,3 hm<sup>3</sup> e  $P$  pelo menos igual a 735,1 mm, configurando a situação com maior representatividade de simulações para a classe.

Os padrões de alocação para a classe “Ótimo” obtiveram a maior quantidade de simulações para a classe, sendo representada por 37,49% da amostra classificada como “Ótimo”, acima da média de 20,0%.

A classe “Bom” foi a classe com menor quantidade de ramificações para identificação dos padrões de alocação, com um nó folha a mais que no CO4. A variável com maior influência foi a vazão da *Dif4*, com a observação do intervalo do *VfimEA* nas ramificações, que não foi observado em outros cenários para a classe “Bom”. A ramificação com maior representatividade dividiu as simulações com vazão abaixo de 2,5% da capacidade de armazenamento, o que indica que o nível do reservatório apresenta níveis baixos para a maioria dos anos, sendo o sistema com confiabilidades acima de 95%, apesar dos baixos níveis do reservatório.

A ramificação que definiu as regras e operação para a classe obteve uma representatividade de 64,38% das simulações classificadas “Bom”, acima da média de 33,33% de representatividade para as ramificações da classe.

A classe “Aceitável” para o CO5 especificamente apresentou um nível a mais para definição do subconjunto da amostra da classe, sendo necessária a divisão das simulações a partir da observação dos intervalos das variáveis *Dif2*, *Dif3* e *PISG*. As simulações com vazão de *Dif3* acima de 91,2%, *Dif2* acima de 46,9% e *PISG* acima de 74,6% da vazão requerida formaram o subconjunto da amostra para a classe “Aceitável”, sendo observadas as vazões *Dif2*, *Dif4* e *PISG* para obtenção das ramificações de maior representatividade, com *Dif2* acima de 71,9%, *PISG* abaixo de 91,1% e *Dif4* acima de 70,9% da vazão requerida.

A ramificação que definiu as regras e operação para a classe obteve uma representatividade de 68,58% das simulações classificadas “Aceitável”, acima da média de 25,0% de representatividade para as ramificações da classe.

A partir da divisão das simulações para a classe “Aceitável”, considerando a observação da vazão do *PISG*, foi formado o subconjunto da amostra de simulações classificadas “Ruim”, sendo o subconjunto da classe definido para vazões abaixo de 74,6% do requerido. A classe “Ruim”, foi a que possuiu uma maior quantidade de nós folha assim como a maior quantidade de intervalos observados para as variáveis do modelo.

Para a classe, todas as variáveis de vazão para uso na irrigação foram consideradas para a definição das regras de operação, juntamente com a precipitação, o que sugere que o impacto das falhas de atendimento para o sistema, causariam déficits de atendimento para as demandas de irrigação de forma homogênea, com danos distribuídos. O nó folha de maior

representatividade considerou as simulações com *Dif1* abaixo de 76,7%, *PISG* acima de 53,9%, *Dif3* acima de 55,9%, *Dif4* acima de 45,9% e *P* acima de 616,8 mm.

A ramificação que definiu as regras e operação para a classe obteve uma representatividade de 68,58% das simulações classificadas “Aceitável”, acima da média de 25,0% de representatividade para as ramificações da classe.

A classe “Péssimo”, foi definida como sendo o subconjunto de simulações com os valores de *Dif3* abaixo de 91,2% e *Dif2* abaixo de 46,9% da vazão requerida, sugerindo que esta seja a característica mais presente para as situações em que o sistema apresente valores médios de *C* abaixo de 50% na irrigação.

A partir do subconjunto com *Dif2* menor ou igual a 46,9%, as simulações já passaram a ser classificadas “Péssimo”, sendo *PISG* abaixo de 50,4% a ramificação com maior quantidade de simulações para a classe, com 97,33% da representatividade das respostas para a classe.

## 5 DISCUSSÃO

### 5.1 ANÁLISE DOS CENÁRIOS

Através da análise conjunta dos cenários operacionais foi possível definir a pior e a melhor situação operacional, assim como para cada cenário também foi possível determinar quais as características mais evidentes nas piores e melhores situações de risco.

A obtenção dos diferentes percentuais de *C* e de diferentes características para os intervalos de valores das variáveis, utilizadas para análise do comportamento do sistema nas faixas de *C* estudadas, apresentaram aspectos característicos de acordo com o equilíbrio das condições de oferta e demanda para as condições de operação e de atendimento ocorrido em cada simulação.

Os cenários elaborados para a análise do SEASG nas diversas situações de planejamento, considerando as variações das regras de operação do sistema, resultaram na análise apurada das variações das condições de atendimento das demandas, sendo possível observar os efeitos das mudanças no desempenho do sistema a partir da análise de risco.

As confiabilidades do sistema para todos os cenários observados apresentaram um comportamento muito semelhante quando analisadas as mudanças propostas nas simulações que representaram o CO2 e CO3, como também para o CO4 e CO5. A medida que a

disponibilidade hídrica aumentou, houve maiores concentrações de simulações com porcentagem 100%, sugerindo que nos eventos sintéticos acima das condições mínimas posteriormente observadas nos modelos de árvore de decisão, a tendência foi de atendimento pleno e imediato conforme o aumento das reservas de água nos reservatórios.

Os CO2, CO3, CO4 e CO5 obtiveram percentuais de confiabilidade acima do limite mínimo para classificação “Ótimo”, diferindo apenas nas médias entre os cenários CO2 e CO3 comparados aos CO4 e CO5.

Os CO4 e CO5 obtiveram médias de confiabilidade menores, pois apesar do atendimento pleno das demandas urbanas ligadas aos reservatórios, a demanda urbana após a confluência com o Rio Piancó não obteve bons resultados para a confiabilidade interferindo no desempenho do cenário, já que as classificações de risco dependeram diretamente desse indicador.

As classes definidas e os agrupamentos das simulações, de acordo com a média de *C*, mostraram que conforme as condições de disponibilidade hídrica do sistema mudaram de suficientemente satisfatória para totalmente insatisfatória, as demandas do sistema foram apresentando padrões de vazão a serem mantidos conforme o sistema realizou a alocação da água.

Assim na medida que a situação hídrica do sistema foi se agravando, as demandas começaram a apresentar falhas no atendimento, de maneira que algumas demandas foram sendo impactadas mais diretamente do que outras, não dependendo da quantidade de vazão requerida para a maioria dos casos.

De todo modo, os intervalos identificados para as ramificações das árvores mostraram quais as demandas que mais comumente apresentaram falha e que mais possuem suas confiabilidades relacionadas com a confiabilidade geral para um determinado cenário ou seja, se o sistema tende a perder desempenho na confiabilidade, a falha de atendimento de algumas demandas específicas será uma característica presente, o que não necessariamente indica que outras demandas também não estejam com falha.

Para todos os casos operacionais estudados, com exceção das situações observadas no CO1, tanto as árvores de decisão para o abastecimento urbano como as árvores para a irrigação mostraram que as variáveis *AbastEA*, *AbastSG* e *AbastCO*, que são as demandas de abastecimento urbano tiveram pouco significância para identificação das principais características das situações de risco, mesmo em situações onde a confiabilidade média do abastecimento esteve dentro das classes “Ruim” e “Péssimo”.

Tal fato pode estar relacionado as prioridades das demandas de abastecimento que possuíram níveis mais altos, sendo sempre consideradas em primeiro lugar, o que fez com que o comportamento do atendimento das vazões se tornasse homogêneo nas simulações, sendo para o modelo mais fácil perceber mudanças no comportamento das vazões das demais variáveis.

Outro fato importante foi o comportamento das árvores quanto a determinação das ramificações, sendo possível estabelecer dentre a grande quantidade simulações diversos padrões de comportamento das vazões e das condições hidrológicas, mesmo em um único subconjunto amostral de uma classe.

Foi possível observar que as variáveis  $V_{fimEA}$  e  $P$  foram as mais presentes dentro dos subconjuntos da classe “Ótimo” em todos os cenários operacionais, o que levou a conclusão de que ao observar a estrutura das árvores de decisão e as variáveis que compõem as ramificações da classe “Ótimo”, estão identificados os limites mínimos das condições hidrológicas necessárias para que o sistema possua o maior nível de confiabilidade, haja vista que em vários cenários também foi recomendado o atendimento total ou quase total de algumas demandas, algo que já é desejado quando se pretende obter o desempenho ótimo da confiabilidade. Contudo, tais condições de atendimento pleno seriam obtidas de maneira mais segura com disponibilidade hídrica favorável, representada pelos valores de  $V_{fimEA}$  e  $P$ .

Com relação aos cenários operacionais, foi possível afirmar que os resultados obtidos para o risco apresentaram relação direta com o comportamento das séries sintéticas, sendo necessária considerá-la na análise dos cenários, pois caso as séries sintéticas possuíssem valores diferentes dada a aleatoriedade do método de geração dos dados, os seus efeitos impactariam no comparativo entre os cenários.

Os padrões das demandas observados para o CO1 mostraram que o cenário apresentou uma maior distribuição das simulações entre as classes se comparada com os demais cenários. Do ponto de vista da análise, esse resultado pode ser considerado positivo, pois possibilitou uma melhor geração das árvores e identificação das classificações, sendo os subconjuntos de cada classe com um maior contingente de simulações.

Os resultados obtidos para os CO2 e CO3 convergem para essa conclusão, já que as árvores de decisão para o abastecimento urbano não apresentaram simulações classificadas para todas as classes de situação de risco estabelecidas. Essa característica presente nestes cenários está diretamente associada aos aportes de vazão enviados pelo PISF. Apesar de os cenários CO4 e CO5 também receberem fornecimento do PISF, estes não obtiveram o mesmo desempenho,

com maior quantidade de simulações com falhas de atendimento, devido ao aumento da demanda, representado pela inserção das demandas *AbastCO* e *Dif4* nas simulações.

Com relação as árvores de decisão para o abastecimento urbano, o pior cenário operacional de atendimento foi o CO1, com a distribuição e classificação das simulações com relação a *C* piores do que para a irrigação, como nos CO2 e CO3. Os CO4 e CO5 estão em segundo lugar como piores cenários, sendo o CO4 um pouco pior que o CO5. Por fim, os cenários CO2 e CO3 obtiveram os melhores valores e distribuições das simulações com relação a *C*, sendo também estes os mesmos cenários que obtiveram os melhores resultados para a irrigação.

Diferente dos demais cenários, no CO1, além do menor valor de *C*, as frequências dos estados de risco obtiveram os valores mais distribuídos. Embora o atendimento sendo mais bem distribuído no CO1, devido a maior prioridade estabelecida para as demandas urbanas através da Lei nº 9433/1997, o sistema necessita inevitavelmente direcionar o atendimento a depender das condições de disponibilidade hídrica, o que é observado nos demais cenários. Sendo assim, mesmo com os melhores desempenhos, preferencialmente é desejável uma melhor distribuição.

As árvores de decisão para o abastecimento urbano, assim como as árvores para as irrigações, destacaram a influência das decisões na configuração e desempenho dos cenários operacionais. A diferença dos valores de *C* entre o CO1 e os cenários que possuíram o fornecimento de água do PISF foi mais evidente que quando comparado com os cenários de acréscimo das demandas após a confluência com o Rio Piancó, indicando que a presença da vazão PISF obteve maior impacto no risco que o aumento da demanda no SEASG.

Para a decisão associada ao atendimento de forma prioritária para a vazão regularizada enviada ao RN ou atendimento da vazão considerando a prioridade equivalente as demandas urbanas, a diferença observada entre os pares CO2 e CO3, junto com CO4 e CO5, mostrou que o impacto da regra de operação foi desprezível sem pouca ou nenhuma diferença para a modelagem das árvores para os CO2 e CO3, assim como para os CO4 e CO5. Vale salientar que a disponibilidade hídrica existente entre os cenários CO2 e CO3 é maior que entre CO4 e CO5, haja vista o aumento do número de elementos de demanda nestes últimos cenários, fazendo com que o impacto seja minimamente sentido nos CO4 e CO5.

Já com relação a modelagem e análise das árvores de decisão para a irrigação, a importância da árvore se deu também pela obtenção de cenários operacionais com resultados muito semelhantes para os indicadores e desempenho dos modelos de árvore de decisão, o que dificultou a avaliação das piores e melhores situações ou para situações com considerável diferença com relação as regras de operação.

A avaliação levou em consideração a ordem crescente, sendo assim o pior cenário de atendimento as demandas rurais foi o CO4, que obteve *C* muito próximo do CO5 que ficou em quarto lugar, baseando-se na estatística descritiva e na distribuição das simulações conforme as classes. Em terceiro lugar ficou o CO1, apesar do cenário possuir a menor oferta de água dentre os cenários. Em segundo e em primeiro lugar ficaram os CO2 e CO3, respectivamente, com uma pequena diferença entre os valores de *C* médios, ambos sendo cenários com máxima oferta de água e mesma quantidade de demandas rurais originalmente.

O CO4 foi aquele com a classe “Péssimo” com 53,76% das simulações classificadas, sendo comparado com a classe “Péssimo” do CO1, com 33,61%. Comparando o *C* entre os cenários com e sem atendimento prioritário ao RN, a diferença entre o CO2 e o CO3 foi maior do que entre o CO4 e o CO5, o que indicou que o aumento das demandas do sistema fez com que a situação dos cenários se tornasse mais semelhante com relação ao risco, diminuindo o impacto das regras de operação.

As frequências das classes de risco mostraram que quanto mais desbalanceado o cenário se apresenta, maiores as chances de este ser considerado como pior ou melhor, enquanto um cenário balanceado dependerá muito mais da média das confiabilidades.

Vale salientar que a disponibilidade hídrica do cenário não é o fator mais importante de definição do risco do cenário, a exemplo do CO1 que não recebeu vazão enviada do PISF e mesmo assim, conseguiu se manter em melhor situação com relação aos CO4 e CO5, que foram simulados considerando a vazão de projeto, podendo esta oscilar ao longo do tempo.

Através da árvore de decisão dos cenários para as irrigações, também foi possível observar o impacto da interferência das decisões nos dados de risco dos cenários por meio da diferença média entre cada nível de análise para cada decisão. Tais aspectos sugerem que dentre as regras de operação, a de maior efeito sobre a irrigação foi o aumento da demanda, resultado que corrobora com a análise dos cenários, juntamente com as deduções referentes a oferta do sistema.

A classe “Bom” para todos os cenários foi aquela que possuiu maior amplitude, com mínimo de 9,37% e máximo de 54,65%. Os cenários CO4 e CO5 obtiveram resultados semelhantes com 12,08 e 12,85%, respectivamente, sendo os cenários CO2 e CO3 os de menor valor.

Por fim, os resultados mostraram a importância da regulação e alerta para o atendimento das demandas de uso para irrigação, principalmente *Dif2* e *Dif3*, no que diz respeito ao nível de importância dessas variáveis observado em todos os cenários operacionais, tanto para a melhoria das condições de atendimento do abastecimento quanto para a irrigação,

sendo estas as variáveis que mais estão correlacionadas com o comportamento da confiabilidade, servindo também de indicador para o atendimento adequado do sistema.

O risco das demandas do sistema para todos os cenários demonstrou que o melhor fornecimento de água foi realizado para as demandas urbanas, mesmo nos cenários com menor disponibilidade hídrica em relação a vazão do PISF e a quantidade de elementos presentes na topologia da simulação. Enquanto para as demandas de irrigação, os cenários mostraram uma diminuição na qualidade do atendimento, que já era esperado devido a diferença das prioridades contudo, os cenários não realizaram o atendimento uniforme para a irrigação, sendo o melhor desempenho geralmente observado em elementos mais próximos dos reservatórios, com destaque para a *Difl*, posicionada entre os REA e RSG, sendo a única diretamente atendida pelo REA e que possui maior capacidade comparada ao RSG.

## 5.2 ÁRVORES DE DECISÃO PARA MODELAGEM DO RISCO

No geral, a vazão média anual das séries sintéticas foi a variável que mais impactou o atendimento do sistema, tendo a precipitação e o volume do reservatório a finalidade de refinar o agrupamento de simulações conforme as classificações.

As árvores de decisão para cada cenário também apresentaram bom desempenho para a previsão do risco através da observação das variáveis, com acurácias acima de 99,0% que pode ser comparada com outras técnicas de aprendizado de máquina, como as redes neurais, floresta randômica, *XGBoost*, entre outros (LU; MA, 2020; DOTA; CUGNASCA; BARBOSA, 2015; ANJOS *et al.*, 2017).

Além da funcionalidade e da robustez verificada, o modelo também se mostrou bastante prático, devido sua fácil interpretação dos resultados por parte dos gestores e planejadores dos sistemas de recursos hídricos, por ser visualmente compreendido, diferente de outros modelos, indo além da comumente utilizada análise gráfica dos resultados dos indicadores de risco.

Juntamente com a análise das variáveis e seus efeitos para a variação do risco, outro ganho trazido pelo modelo foi a possibilidade de análise das regras de operação por meio da análise dos cenários, que viabilizou de forma prática a correlação entre as vazões afluentes às demandas e aspectos hidrológicos dos sistemas.

Através dos modelos de árvore de decisão, foi possível considerar que a regra de operação de maior impacto foi o fornecimento da vazão do PISF que fez com que o atendimento

das demandas urbanas fosse atendido plenamente na maioria dos cenários, estando mais bem supridas em todas as simulações realizadas. Vale salientar que, devido ao não supervisionamento do atendimento, durante as simulações executadas de forma automática, mesmo entre demandas de mesma prioridade a alocação de água não foi uniforme em alguns casos, o que sugere a necessidade de que no contexto real, a distribuição do atendimento seja melhor realizada visando a uniformidade, sendo orientada principalmente pelas árvores de decisão.

A principal vantagem percebida para as variáveis criadas na caracterização do comportamento dos dados sintéticos, foi a versatilidade da aplicação do modelo de árvore de decisão assim como fora identificado também em outros trabalhos como o realizado por Lu e Ma (2020), assim como Murphy *et al.* (2016), tanto para as demandas urbanas quanto de irrigação, podendo ser aplicado para cenários com expansão da área de estudo sem a necessidade de desconsiderar os dados sintéticos e resultados já calculados, por meio de outras técnicas como fora realizado por Zhao e Cai (2020) que utilizaram cadeias de Markov, sendo comparados os dados sintéticos antigos e novos.

Com relação a outros trabalhos envolvendo as árvores de decisão e os recursos hídricos, os conhecimentos e a modelagem realizada para a geração dos diagnósticos do risco na operação dos sistemas podem contribuir significativamente com a análise de resultados em outras aplicações, como é o caso de Zhao e Cai (2020) no desenvolvimento de regras de operação para os reservatórios, sendo essas regras muito mais assertivas a partir da consideração da confiabilidade no atendimento às demandas.

Não somente nos sistemas de reservatórios, mas de forma indireta, a aplicação pode ser de grande valia para a análise do risco em outras linhas de pesquisa, como o esgotamento sanitário que envolve também questões de saúde pública. Um exemplo que pode servir de comparação são os dos trabalhos de Syachrani, Jeong e Chung (2013), em que o gerenciamento do esgotamento pode passar pela identificação de pontos críticos nas redes de esgotamento que possuem padrões de comportamentos das variáveis para as situações de degradação das tubulações. Contudo, com a aplicação da modelagem para análise de risco para esse caso, facilitaria a prévia localização dos trechos de intervenção na rede, dando mais tempo para que os gestores possam atuar no melhor condicionamento dos sistemas, tendo a minimização dos riscos como base.

### 5.3 Avaliação da proposta metodológica

Foi possível concluir que o RAMSS demonstrou que mesmo nos cenários operacionais com piores condições de atendimento, foi possível identificar em quais condições de precipitação, vazão e volumes de reservatório, o atendimento do sistema pode ser continuado ou adequado, de modo a compensar a falta de água.

Por fim, a análise de risco forneceu subsídios para a verificação do risco existente no sistema de acordo com as condições hidroclimatológicas. Nos casos que houve seca foi possível identificar os cenários com melhor desempenho, avaliando os possíveis eventos de forma estocástica, aumentando a precisão das análises de risco, viabilizando a verificação do atendimento as demandas do sistema, abrangendo tanto as situações com melhor como pior disponibilidade hídrica.

A simulação estocástica do sistema, implementada dentro do procedimento metodológico do RAMSS, se mostrou robusta com relação a observação das inúmeras possibilidades de análise das variações da disponibilidade hídrica em detrimento da precipitação e conseqüentemente da vazão afluente ao sistema.

A metodologia proposta neste trabalho, possui alta capacidade de aplicação em outros sistemas de recursos hídricos, sem a necessidade de adaptações na estrutura do procedimento metodológico, podendo agregar na análise de resultados para outros estudos correlacionados, que dispõem da aplicação dos indicadores trabalhados por Hashimoto, Stedinger e Loucks (1982), com destaque para a avaliação da confiabilidade de atendimento das demandas.

É possível concluir através do método RAMSS que não somente as falhas no atendimento estão relacionadas ao nível de atendimento e o risco analisado, mas além disso, foi possível relacionar as condições hidroclimáticas com o risco e até mesmo prever qual o risco do sistema baseado nessas tais condições. Um exemplo dessa relação está na obtenção dos percentuais de confiabilidade para uma demanda específica, que podem ser obtidos exclusivamente do quantitativo de falhas em um período observado, enquanto com as árvores de decisão, a confiabilidade pôde ser indicada pelos valores de precipitação, volume de reservatório e até mesmo pelas vazões de outras demandas no sistema.

## 6 CONCLUSÃO

Foi possível realizar o desenvolvimento e aplicação das séries sintéticas de precipitação e vazão para obtenção de resultados que ajudassem a definir as principais características e comportamentos do sistema com relação ao risco e contribuir para a melhor alocação de água do sistema.

A elaboração dos cenários, a simulação do balanço hídrico e a árvores de decisão, mostraram a influência significativa das vazões afluentes às demandas do sistema no risco nos usos específicos do sistema, a exemplo das demandas de irrigação influenciando na confiabilidade do abastecimento urbano.

Também foi possível verificar a ação da precipitação e o volume do reservatório sobre o risco, indicando que a análise e previsão dessas variáveis pode antecipar situações indesejáveis de alocação, facilitando o planejamento dos recursos hídricos e minimizando os impactos em cenários de crise. Possibilitou, portanto, um melhor entendimento sobre como o sistema de recursos hídricos responde as variações hidrológicas.

Os principais pontos críticos do sistema foram as demandas de irrigação, mais próximas do RSG que possuíam menor quantidade de vazão requerida a exemplo das áreas difusas, enquanto que o volume do REA demonstrou ser mais importante para o risco no abastecimento urbano. Além disso, as árvores de decisão conseguiram apontar as faixas de vazão e os limites mínimos de precipitação de volume de reservatório mais prováveis, relacionados ao risco, indicando os elementos mais vulneráveis para as faixas de risco estudadas.

A modelagem das árvores de decisão mostrou ser uma técnica prática, confiável e robusta para a análise dos indicadores e previsão dos riscos do sistema, em detrimento dos aportes e consumos de água nos sistemas de reservatórios auxiliando, portanto, na definição de estratégias de operação para o planejamento dos recursos hídricos.

Se espera que a metodologia desenvolvida neste trabalho possa ser aplicada e com isso promover soluções e estratégias para o planejamento dos sistemas de recursos hídricos, melhorando assim a qualidade de vida das populações do semiárido e do nordeste brasileiro.

## 7 RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Dentre as recomendações para realização de trabalhos futuros, podem ser elencados os seguintes pontos:

- Devido ao RAMSS ser aplicado apenas utilizando os modelos com a definição de classes de risco para a confiabilidade, poderiam ser investigadas formas de compatibilizar a aplicação de outros indicadores de Hashimoto, Stedinger e Loucks (1982), de forma integrada, considerando também a abordagem de outros parâmetros de medem o desempenho dos sistemas de recursos hídricos;
- As árvores de decisão modeladas no trabalho utilizaram apenas a precipitação, a vazão e o volume do reservatório como variáveis predictoras das classes de risco. É possível testar a modelagem das árvores de decisão levando em consideração o ganho de informação para estabelecimento dos padrões de classificação, através de outras variáveis como a evaporação nos reservatórios, a evapotranspiração e a inclusão de outros usos como a geração de energia elétrica, tornando a identificação das condições para o risco, muito mais especificadas, utilizando até mesmo variáveis com maior poder de controle para o planejamento;
- Devido ao grande número de séries pluviométricas sintéticas e simulações executadas para cada cenário operacional, a divisão das faixas e estabelecimento das classes de risco para a confiabilidade tornaram a modelagem das árvores de decisão limitadas a observância de apenas cinco situações distintas. Embora a capacidade de interpretação das árvores de decisão possa ser reduzida, que é a uma das grandes vantagens do modelo, é possível dividir a análise da variável dependente em mais classes, tornando a identificação das condições de ocorrência das situações de risco ainda mais precisa;
- Além da utilização das árvores de decisão para a classificação dos cenários sintéticos, também pode ser feita a investigação da aplicação de outras metodologias como a Floresta Aleatória e *XGBoost*, com foco na análise de risco, verificando o desempenho dos resultados e o manuseio por parte dos profissionais dos órgãos que lidam com a gestão e o planejamento dos sistemas;
- O RAMSS considerou apenas a modelagem sintética da precipitação, sendo esta responsável por ser a componente estocástica, pela qual as demais variáveis de correlação com a análise de risco, como a vazão e o volume do reservatório, foram

geradas. Assim, é possível considerar que em trabalhos futuros a implementação da componente estocástica no método possa vir a ser representada não somente pela precipitação, mas também pela evaporação, vazão, consumo de água, entre outras, com dados gerados sinteticamente;

- A aplicação do método foi realizada em um sistema de recursos hídricos com apenas dois reservatórios e poucos elementos de demanda, numa bacia hidrográfica gerida pela União. Como forma de melhorar a análise, é possível expandir a aplicação para sistemas maiores e mais complexos, sendo avaliados considerando todos os elementos que influenciam ou são influenciados pela bacia, seus usuários, os usos e conflitos associados;
- Algo que foi inicialmente pensando para o trabalho foi a aplicação da otimização, impossibilitada por motivos inerentes a pesquisa. Contudo, para trabalhos futuros fica a sugestão de aplicação de modelos de otimização.
- Por fim, também se recomenda a utilização dos dados dos indicadores sendo aplicados como variáveis previsoras para um modelo de árvore de decisão que possa auxiliar na determinação do índice de sustentabilidade, tendo em vista os diferentes comportamentos das vazões afluentes às demandas.

## REFERÊNCIAS

ABDELWAHED KHALIFA, Hamiden. Study on multi-objective nonlinear programming in optimization of the rough interval constraints. **International Journal of Industrial Engineering & Production Research**, v. 29, n. 4, p. 407-413, 2018.

Agência Executiva de Gestão de Águas – AESA. **Meteorologia, Chuvas.**, 2020. Disponível em: <<http://www.aesa.pb.gov.br/aesa-website/meteorologia-chuvas/>>. Acesso em: 9 jan. 2021.

\_\_\_\_\_. **Volume Mensal. João Pessoa: AESA**, 2021a. Disponível em: <<http://www.aesa.pb.gov.br/aesa-website/monitoramento/volume-mensal/?tipo=anterior>>. Acesso em: 9 jan. 2021.

AGUIAR, Renata Gonçalves. **Balanco de Energia em Ecossistema Amazônico por Modelo de Regressão Robusta com Bootstrap e Validação Cruzada**. 2013. Universidade Federal de Mato Grosso, Instituto de Física, Programa de Pós-Graduação em Física Ambiental, Cuiabá, 2013.

AHMADI, Mohsen; HADDAD, Omid Bozorg; LOÁICIGA, Hugo A. Adaptive reservoir operation rules under climatic change. **Water Resources Management**, v. 29, n. 4, p. 1247-1266, 2015.

Agência Nacional de Águas – ANA. **Plano de recursos hídricos da bacia hidrográfica do Rio Piancó- Piranhas-Açu**. Brasília: ANA. 2016.

\_\_\_\_\_. **Nota Técnica no 11/2018/COMAR/SRE**. Marco Regulatório estabelecendo condições de uso dos recursos hídricos no sistema hídrico Eng. Ávidos, São Gonçalo e rio Piranhas, no Estado da Paraíba. 2018a.

\_\_\_\_\_. **Resolução conjunta ANA/AESA-PB nº 76, de 09 de outubro de 2018**. Dispõe sobre condições de uso dos recursos hídricos no sistema hídrico Engenheiro Ávidos/São Gonçalo e rio Piranhas, localizado no Estado da Paraíba. Brasília. 2018b.

ANJOS, Camila Souza dos *et al.* Análise do nível de legenda de classificação de áreas urbanas empregando imagens multiespectrais e hiperespectrais com os métodos árvore de decisão C4.5 e Floresta randômica. **Boletim de Ciências Geodésicas**, v. 23, p. 371-388, 2017.

BORYCZKO, Krzysztof; TCHÓRZEWSKA-CIESLAK, Barbara. Application of Fuzzy Fault Tree in Risk Analysis of Collective Water Supply Systems/ZASTOSOWANIE ROZMYTYCH DRZEW NIEZDATNOSCI W ANALIZIE RYZYKA SYSTEMOW ZBIOROWEGO ZAOPATRZENIA W WODE. **Journal of KONBiN**, v. 24, n. 1, p. 13, 2012.

BREIMAN, Leo *et al.* **Classification and regression trees**. Routledge, 2017.

CASTRO, Cristiane Martins de Sousa Nava. **Emprego de indicadores e índices de sustentabilidade na avaliação de medidas emergenciais de gestão da água: caso da bacia**

**do Rio Descoberto**. 2019. Dissertação (Mestrado em gestão e regulação de recursos hídricos) - Programa de Pós-Graduação PROFÁGUA – UFRGS, Porto Alegre, 2019.

CASTRO, Cristiane Martins de Sousa Nava; MARQUES, Guilherme Fernandes. Emprego de indicadores e índices de sustentabilidade: caso da bacia do rio Descoberto, DF. **Simpósio Brasileiro de Recursos Hídricos (23.: Foz do Iguaçu, 2019)**. Anais [recurso eletrônico]. Porto Alegre: ABRH, 2019, 2019.

CELESTE, Alcigeimes Batista; CURI, Wilson Fadlo; CURI, Rosires Catão. Análise de métodos para geração sintética de vazões em rios intermitentes. **Proceedings of the XXVII Simpósio Brasileiro de Recursos Hídricos**, 2007.

CHANG, Fi-John; WANG, Yu-Chung; TSAI, Wen-Ping. Modelling intelligent water resources allocation for multi-users. **Water resources management**, v. 30, n. 4, p. 1395-1413, 2016.

CHENG, Wen-Ming *et al.* Risk analysis of reservoir operations considering short-term flood control and long-term water supply: A case study for the Da-Han Creek Basin in Taiwan. **Water**, v. 9, n. 6, p. 424, 2017.

CHENG, Xi *et al.* Optimal water resources management and system benefit for the Marcellus shale-gas reservoir in Pennsylvania and West Virginia. **Journal of Hydrology**, v. 540, p. 412-422, 2016.

CHIODI, Rafael Eduardo; SARCINELLE, Oscar; UEZU, Alexandre. Gestão dos recursos hídricos na área do Sistema Produtor de Água Cantareira: um olhar para o contexto rural. **Revista Ambiente & Água**, v. 8, p. 151-165, 2013.

Confederação Nacional dos Municípios – CNM. **Análise sobre a seca do Nordeste**. 2013. Disponível em: <[https://www.cnm.org.br/cms/biblioteca\\_antiga/ET Vol 6 - 09. Análise sobre a seca do nordeste.pdf](https://www.cnm.org.br/cms/biblioteca_antiga/ET_Vol_6_-_09_Analise_sobre_a_seca_do_nordeste.pdf)>.

COSTA NETO, Francisco de Assis da *et al.* **Análise de alternativas para o uso das águas do projeto de integração do rio São Francisco-eixo norte, no Estado da Paraíba através de um modelo de otimização multiobjetivo**. 2020. Dissertação (Mestrado em Engenharia Civil e Ambiental) - Pós-graduação em Engenharia Civil e Ambiental – UFCG, Campina Grande, 2020.

DE ALVARENGA JÚNIOR, Wagner José. **Métodos de otimização hiperparamétrica: um estudo comparativo utilizando árvores de decisão e florestas aleatórias na classificação binária**. 2018. Dissertação (Mestrado em Engenharia Elétrica) – Programa de Pós-Graduação em Engenharia Elétrica – UFMG, Belo Horizonte, 2018.

DE ANDRADE, Paulo Romero Guimarães Serrano; CURI, Wilson Fadlo; CURI, Rosires Catão. Aplicação de critério de sustentabilidade na avaliação de desempenho de um sistema hídrico de usos múltiplos. **XVIII Simpósio Brasileiro De Recursos Hídricos**, p. 1–20, 2009.

DE OLIVEIRA, Luiz F. C. *et al.* Comparação de metodologias de preenchimento de falhas de séries históricas de precipitação pluvial anual. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 14, p. 1186-1192, 2010.

DE SANTANA SANTOS, Itallo Henrique; MACIEL, Alexandre Magno Andrade. Desenvolvimento de um Sistema de Apoio a Decisão para priorização de Pedidos de Desembolso no Estado de Pernambuco. **Revista de Engenharia e Pesquisa Aplicada**, v. 3, n. 3, 2018.

DETZEL, Daniel H. Marco; BESSA, Marcelo Rodrigues; MINE, Miriam Rita Moro. Amostragem de Séries Sintéticas Hidrológicas. **Revista Brasileira de Recursos Hídricos**, v. 18, n. 4, p. 173-184, 2013.

DO NASCIMENTO SILVA, Semirames *et al.* Efeitos da escassez hídrica na economia do perímetro irrigado de São Gonçalo, Paraíba. **Revista Verde de Agroecologia e Desenvolvimento Sustentável**, v. 12, n. 1, p. 132-137, 2017.

DONG, Ningpeng *et al.* Climate-induced hydrological impact mitigated by a high-density reservoir network in the Poyang Lake Basin. **Journal of Hydrology**, v. 579, p. 124148, 2019.

DOTA, Mara Andrea; CUGNASCA, Carlos Eduardo; BARBOSA, Domingos Sávio. Comparative analysis of decision tree algorithms on quality of water contaminated with soil. **Ciência Rural**, v. 45, p. 267-273, 2015.

DUARTE, Filipe Coelho de Lima; GIRÃO, Luiz Felipe de Araújo Pontes; PAULO, Edilson. Avaliando modelos lineares de value relevance: Eles captam o que deveriam captar?. **Revista de Administração Contemporânea**, v. 21, p. 110-134, 2017.

EVERAERT, Gert; BENNETSEN, Elina; GOETHALS, Peter LM. An applicability index for reliable and applicable decision trees in water quality modelling. **Ecological Informatics**, v. 32, p. 1-6, 2016.

FAN, Xudong *et al.* Machine learning based water pipe failure prediction: the effects of engineering, geology, climate and socio-economic factors. **Reliability Engineering & System Safety**, v. 219, p. 108185, 2022

FERNANDES, Fabiano Rodrigues. **Emprego de diferentes algoritmos de árvores de decisão na classificação da atividade celular in vitro para tratamentos de superfícies de titânio**. 2017. Dissertação (Mestrado em Engenharia) – Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e dos Materiais – UFRGS, Porto Alegre, 2017.

GIL, Antonio Carlos. **Como Elaborar Projetos de Pesquisa**. 2007

GIULIANI, Matteo; GALELLI, Stefano; SONCINI-SESSA, Rodolfo. A dimensionality reduction approach for many-objective Markov Decision Processes: Application to a water reservoir operation problem. **Environmental Modelling & Software**, v. 57, p. 101-114, 2014.

GOHARIAN, Erfan *et al.* Vulnerability assessment to support integrated water resources management of metropolitan water supply systems. **Journal of Water Resources Planning and Management**, v. 143, n. 3, p. 04016080, 2017.

HANNAN, Abdul; ANMALA, Jagadeesh. Classification and Prediction of Fecal Coliform in Stream Waters Using Decision Trees (DTs) for Upper Green River Watershed, Kentucky, USA. **Water**, v. 13, n. 19, p. 2790, 2021.

HASHIMOTO, Tsuyoshi; STEDINGER, Jery R.; LOUCKS, Daniel P. Reliability, resiliency, and vulnerability criteria for water resource system performance evaluation. **Water Resources Research**, v. 18, n. 1, p. 14-20, 1982.

HIGGINS, Andrew; ARCHER, Ainsley; HAJKOWICZ, Stefan. A stochastic non-linear programming model for a multi-period water resource allocation with multiple objectives. **Water Resources Management**, v. 22, n. 10, p. 1445-1460, 2008.

IBGE. **Instituto Brasileiro de Geografia e Estatística**, 2020. Disponível em: <<https://memoria.ibge.gov.br/sinteses-historicas/historicos-dos-censos/censodemograficos.html%3E>>. Acesso em: 15 dez. 2020.

JAISWAL, Sanjay K. Estimation of Surplus Water for Groundwater Recharging from Ravishankar Sagar Reservoir Basin. **International Journal of Applied Engineering Research**, v. 13, n. 17, p. 13103-13107, 2018.

KLEMEŠ, Vít. Reliability estimates for a storage reservoir with seasonal input. **Journal of Hydrology**, v. 7, n. 2, p. 198-216, 1969.

KUO, Chih-Liang; HSU, Nien-Sheng. An optimization model for crucial key pipes and mechanical reliability: a case study on a water distribution system in Taiwan. **Water resources management**, v. 25, n. 2, p. 763-775, 2011.

LEE, Saro; LEE, Chang-Wook. Application of decision-tree model to groundwater productivity-potential mapping. **Sustainability**, v. 7, n. 10, p. 13416-13432, 2015.

LEONETI, Alexandre Bevilacqua; PRADO, Eliana Leão do; OLIVEIRA, Sonia Valle Walter Borges de. Saneamento básico no Brasil: considerações sobre investimentos e sustentabilidade para o século XXI. **Revista de Administração Pública**, v. 45, p. 331-348, 2011.

LI, Yi-Ping; HUANG, Guan-Hua. Risk analysis and management for water resources systems. **Stochastic Environmental Research and Risk Assessment**, v. 27, n. 3, p. 593-597, 2013.

LISBOA, Severina Sarah; HELLER, Léo; SILVEIRA, Rogério Braga. Desafios do planejamento municipal de saneamento básico em municípios de pequeno porte: a percepção dos gestores. **Engenharia Sanitária e Ambiental**, v. 18, p. 341-348, 2013.

LOUCKS, Daniel Pete. Quantification des tendances de la durabilité des systèmes. **Hydrological Sciences Journal**, v. 42, n. 4, p. 513-530, 1997.

LU, Hongfang; MA, Xin. Hybrid decision tree-based machine learning models for short-term water quality prediction. **Chemosphere**, v. 249, p. 126169, 2020.

MORGADO, Fernando. Biodiversidade e sustentabilidade dos recursos biológicos no contexto das alterações climáticas: desafios no contexto africano. **Revista Captar: Ciência e Ambiente para Todos**, v. 8, n. 1, p. 1-12, 2019.

MURPHY, Heather M. *et al.* Using decision trees to predict drinking water advisories in small water systems. **Journal-American Water Works Association**, v. 108, n. 2, p. E109-E118, 2016.

NEPOMUCENO, Thiago Cabral *et al.* **Modelo para simulação de nitrogênio e fósforo em sistemas de recursos hídricos, aplicação em região do alto curso do rio Piranhas-PB**. 2018. Dissertação (Mestrado em Engenharia Civil e Ambiental) - Pós-graduação em Engenharia Civil e Ambiental – UFCG, Campina Grande, 2018.

NOURANI, Vahid *et al.* An integrated simulation-optimization framework to optimize the reservoir operation adapted to climate change scenarios. **Journal of Hydrology**, v. 587, p. 125018, 2020.

OKADA, Hugo Kenji Rodrigues; DAS NEVES, André Ricardo Nascimento; SHITSUKA, Ricardo. Análise de Algoritmos de Indução de Árvores de Decisão. **Research, Society and Development**, v. 8, n. 11, p. e298111473-e298111473, 2019.

Organização das Nações Unidas – ONU. **World Water Development Report**. 2019.

PARAÍBA. **Plano Diretor de Recursos Hídricos do Estado da Paraíba: Sub-bacia do Médio Piranhas**. 1996.

PEIXINHO, Frederico Cláudio. **Gestão sustentável dos recursos hídricos**. 2010.

PEREIRA, A. S. *et al.* Método qualitativo, quantitativo ou quali-quantitativo. **Metodologia da Pesquisa Científica**. Disponível em: [https://repositorio.ufsm.br/bitstream/handle/1/15824/Lic\\_Computacao\\_Metodologia-Pesquisa-Cientifica.pdf](https://repositorio.ufsm.br/bitstream/handle/1/15824/Lic_Computacao_Metodologia-Pesquisa-Cientifica.pdf), 2018.

PICCOLI, Andrezza de Souza *et al.* Environmental Education as a social mobilization strategy to face water scarcity. **Ciencia & saude coletiva**, v. 21, p. 797-808, 2016.

PRUSKI, Fernando F. *et al.* Precipitação média anual e vazão específica média de longa duração, na Bacia do São Francisco. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 8, p. 247-253, 2004.

PULIDO-VELAZQUEZ, Manuel; CABRERA MARCET, Enrique; GARRIDO COLMENERO, Alberto. Economía del agua y gestión de recursos hídricos. **Ingeniería del agua**, v. 18, n. 1, p. 99-110, 2014.

QUINLAN, J. Ross. Discovering rules by induction from large collections of examples. **Expert systems in the micro electronics age**, 1979.

QUINLAN, J. Ross. **C4.5: Programs for machine learning**. Morgan Kaufman, 1993.

REDDY, P. J. **Stochastic Hydrology**. Delhi, Índia: Laxmi Publications, 1987.

REIS, Cristiane Queiroz; SANTOS, Valterlin da Silva; FARIAS, Camyllo Allyson Simões de. Avaliação da sustentabilidade hídrica dos reservatórios Engenheiro Ávidos e São Gonçalo no Estado da Paraíba. **Revista Brasileira de Gestão Ambiental**, v. 13, n. 2, 2019.

RIVERO, Norah Patricia Panozo. **Modelo de escolha estratégica para a identificação de alternativas para o combate à escassez de água em La Paz–Bolívia**. Dissertação (Mestrado) - Programa de Pós- Graduação em Engenharia de Produção UFPE, Recife, 2018.

RONG, Qiangqiang *et al.* Identification of the optimal agricultural structure and population size in a reservoir watershed based on the water ecological carrying capacity under uncertainty. **Journal of Cleaner Production**, v. 234, p. 340-352, 2019.

SAAB, Saad Mawlood *et al.* Review on generating optimal operation for dam and reservoir water system: simulation models and optimization algorithms. **Applied Water Science**, v. 12, n. 4, p. 1-28, 2022.

SAMPAIO, A. P. C. *et al.* Pegada hídrica da água de coco verde nas principais regiões produtoras do Nordeste. **LALCA - Revista Latino-Americana em Avaliação do Ciclo de Vida**, v. 2, p. 128–141, 2018.

SANTOS, Valterlin da Silva. Modelo de otimização quali-quantitativo multiobjetivo para o planejamento dos recursos hídricos superficiais, com aplicação à bacia do rio Paraíba. 2011. Tese (Doutorado em Recursos Naturais) – Programa de Pós-Graduação em Recursos Naturais – UFCG, Campina Grande, 2011.

SILVA, Airton Marques da. da. **Metodologia da Pesquisa**. 2. ed. Fortaleza-CE: EDUECE, 2015.

SILVA NETO, Enéas Dantas da. **Modelo de cobrança pelo uso dos recursos hídricos em Bacias Hidrográficas Controladas por reservatórios: proposta metodológica e simulação na bacia hidrográfica dos Rios Piancó-Piranhas-Açu**. 2018. Tese (Doutorado em Recursos Naturais) – Programa de Pós-Graduação em Recursos Naturais – UFCG, Campina Grande, 2018.

SILVEIRA, Sandra Maria Batista; SILVA, Maria das Graças. Conflitos socioambientais por água no Nordeste brasileiro: expropriações contemporâneas e lutas sociais no campo. **Revista Katálisis**, v. 22, p. 342-352, 2019.

SUGAHARA, Cibele Roberta *et al.* Avaliação da sustentabilidade das bacias PCJ a partir de indicadores de disponibilidade e demandas hídricas. In: II Sustentare – Seminário de Sustentabilidade da PUC-Campinas, Campinas-SP. **Anais...** Campinas-SP: 2020.

SUN, Yimeng *et al.* Risk analysis for reservoir real-time optimal operation using the scenario tree-based stochastic optimization method. **Water**, v. 10, n. 5, p. 606, 2018.

SUQUINA, Paulina Da Silva Orlando. **Estudo e construção de árvores de decisão: aplicação ao ensino**. 2019. Dissertação (Mestrado em Matemática e Computação) – Matemática – Universidade do Minho, Braga – Portugal, 2019.

- SYACHRANI, Syadaruddin; JEONG, Hyung Seok “David”; CHUNG, Colin S. Decision tree-based deterioration model for buried wastewater pipelines. **Journal of Performance of Constructed Facilities**, v. 27, n. 5, p. 633-645, 2013.
- TEDESCO, Andrea; ANTUNES, Alzir Felipe Buffara; OLIANI, Luiz Octávio. Detecção de formação erosiva (voçoroca) por meio de classificação hierárquica e por árvore de decisão. **Boletim de Ciências Geodésicas**, v. 20, p. 1005-1026, 2014.
- VAROUCHAKIS, Emmanouil A.; PALOGOS, Ioannis.; KARATZAS, George P. Application of Bayesian and cost benefit risk analysis in water resources management. **Journal of Hydrology**, v. 534, p. 390-396, 2016.
- VELOSO, Mariana Faria *et al.* Utilização do algoritmo de árvore de decisão para mapeamento do uso e cobertura da terra da área de contribuição de uma barragem localizada na Bacia Hidrográfica do São Francisco. **Caderno de Ciências Agrárias**, v. 13, p. 1-9, 2021.
- VIEIRA, A. S. **Modelo de Simulação Quali-quantitativo Multiobjetivo para o Planejamento Integrado dos Sistemas de Recursos Hídricos**. 2011. Tese (Doutorado em Recursos Naturais) – Programa de Pós-Graduação em Recursos Naturais – UFCG, Campina Grande, 2011.
- WILGES, Beatriz *et al.* Avaliação da aprendizagem por meio de lógica de fuzzy validado por uma Árvore de Decisão ID3. **Renote**, v. 8, n. 3, 2010.
- WU, Jiefeng *et al.* Impacts of reservoir operations on multi-scale correlations between hydrological drought and meteorological drought. **Journal of Hydrology**, v. 563, p. 726-736, 2018.
- WURBS, Ralph A. Modeling River/reservoir system management, water allocation, and supply reliability. **Journal of Hydrology**, v. 300, n. 1-4, p. 100-113, 2005.
- YAMOUT, Ghina M.; HATFIELD, Kirk; ROMEIJN, H. Edwin. Comparison of new conditional value-at-risk-based management models for optimal allocation of uncertain water supplies. **Water Resources Research**, v. 43, n. 7, 2007.
- YANG, Shuyu *et al.* Real-time reservoir operation using recurrent neural networks and inflow forecast from a distributed hydrological model. **Journal of Hydrology**, v. 579, p. 124229, 2019.
- YASSIN, Fuad *et al.* Representation and improved parameterization of reservoir operation in hydrological and land-surface models. **Hydrology and Earth System Sciences**, v. 23, n. 9, p. 3735-3764, 2019.
- ZHANG, Di *et al.* Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. **Journal of Hydrology**, v. 565, p. 720-736, 2018.
- ZHANG, Xiaoli *et al.* 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.

ZHAO, Qiankun; CAI, Ximing. Deriving representative reservoir operation rules using a hidden Markov-decision tree model. **Advances in Water Resources**, v. 146, p. 103753, 2020.

ZHU, Mengyuan *et al.* A review of the application of machine learning in water quality evaluation. **Eco-Environment & Health**, 2022.

## ANEXOS

**Tabela A1 - Série histórica de precipitações em mm para o reservatório Engenheiro Ávidos**

Ano/mês	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
1963	124,7	454,4	384,2	84,1	32,6	26,0	0,2	0,0	1,0	21,8	36,3	98,0
1964	142,1	190,6	310,2	255,2	109,2	37,9	13,2	1,7	41,7	8,5	20,7	10,6
1965	113,7	30,9	240,2	339,4	29,1	96,4	13,1	22,3	1,1	39,5	0,2	0,6
1966	94,8	112,3	62,8	148,3	57,1	65,2	19,8	0,2	20,6	1,3	1,7	50,7
1967	104,5	230,1	306,8	447,8	256,9	4,9	4,9	3,0	0,7	0,3	0,6	29,1
1968	142,1	48,9	492,5	77,1	79,3	21,3	11,2	0,7	0,0	1,2	38,6	37,5
1969	110,8	189,9	178,7	255,7	78,4	64,3	30,9	7,8	2,5	13,4	0,4	31,3
1970	131,2	83,9	171,9	109,9	1,0	3,5	2,9	0,1	2,7	28,3	8,8	28,3
1971	185,1	234,6	147,0	301,6	68,2	76,9	27,9	1,2	4,7	46,8	2,2	3,2
1972	192,2	165,7	239,9	128,2	41,2	27,5	21,6	23,5	3,1	3,0	0,0	157,5
1973	57,6	134,9	211,9	454,7	78,9	69,6	35,3	36,1	29,4	1,3	1,5	44,9
1974	192,8	370,2	332,3	403,0	57,4	21,6	8,2	0,7	3,2	24,4	46,7	41,2
1975	159,2	199,6	342,6	216,1	80,9	62,5	76,5	0,1	6,2	16,6	1,7	35,6
1976	67,9	346,3	277,4	106,2	17,3	18,9	0,1	1,1	85,5	40,2	66,2	6,6
1977	133,2	196,4	259,1	293,9	154,5	95,3	47,9	1,9	0,0	11,3	0,0	96,3
1978	93,7	433,9	176,8	132,4	138,5	33,8	92,8	2,8	6,0	1,2	31,7	10,9
1979	292,1	322,1	248,8	188,4	74,0	11,2	4,2	8,5	38,9	48,8	16,6	10,5
1980	101,3	409,9	234,6	33,6	9,1	32,6	23,2	0,0	1,8	32,0	89,2	11,6
1981	111,4	74,9	460,2	49,3	0,3	14,3	1,0	4,2	0,0	0,1	27,7	131,3
1982	98,9	103,6	173,0	163,5	32,8	4,8	1,9	1,7	3,7	48,5	30,1	13,4
1983	92,8	200,8	91,7	109,8	5,3	14,0	10,4	2,3	0,4	0,0	1,9	9,7
1984	141,5	139,5	242,3	263,6	78,8	3,9	8,8	4,6	0,0	9,6	28,6	19,7
1985	351,7	315,7	386,6	529,5	141,1	60,8	32,1	14,2	1,5	0,5	10,7	183,9
1986	91,2	297,5	315,1	246,2	59,7	58,4	33,1	17,7	18,6	33,2	76,0	5,2
1987	85,6	129,4	304,3	122,0	15,0	50,7	1,6	0,6	0,8	0,0	0,0	0,8
1988	86,1	143,6	255,3	314,6	42,9	4,4	10,8	0,0	2,3	3,0	1,3	132,9
1989	169,4	46,4	210,6	311,1	197,2	22,4	17,7	11,2	34,8	28,1	2,6	171,2
1990	16,9	144,5	62,6	256,8	41,9	10,8	3,8	39,5	29,5	5,3	27,9	8,6
1991	169,2	129,9	224,7	88,0	97,0	3,5	1,0	1,9	0,3	0,0	12,7	4,6
1992	127,5	164,7	164,0	92,4	19,2	56,7	19,5	6,7	11,8	1,2	6,8	3,4
1993	28,2	93,7	75,2	79,7	62,8	9,0	36,8	3,7	13,6	21,2	6,8	7,8
1994	382,0	190,3	116,6	134,8	159,3	100,7	23,3	2,0	0,0	0,8	1,0	89,4
1995	148,3	139,5	165,8	243,1	204,0	16,4	18,7	0,0	0,0	0,0	35,3	9,5
1996	179,3	162,2	238,8	275,3	170,7	23,8	1,7	3,7	0,0	3,9	16,1	15,5
1997	234,8	92,7	185,8	127,0	99,8	0,8	13,7	5,6	0,4	3,6	70,9	30,9
1998	149,4	97,2	123,4	117,4	2,7	6,7	10,3	0,0	0,0	0,0	0,0	19,0
1999	179,0	117,0	264,1	97,0	110,2	7,5	3,1	3,5	5,9	8,1	33,2	176,1
2000	137,5	256,6	212,6	230,4	32,2	18,8	32,7	12,0	34,6	2,9	0,0	78,7
2001	49,9	119,8	303,2	55,6	9,6	38,2	2,4	3,9	0,0	29,0	7,8	151,0
2002	387,9	91,1	199,8	58,5	102,4	35,6	6,6	0,0	3,2	1,0	4,3	21,9
2003	160,7	76,2	195,7	140,5	47,6	15,8	5,2	1,3	9,5	15,5	16,9	51,0
2004	361,0	323,7	114,4	52,0	86,9	39,8	23,5	8,1	1,1	0,0	10,7	34,4
2005	93,0	71,5	385,3	76,0	48,3	26,7	4,3	5,0	0,0	0,0	0,0	107,4
2006	54,1	240,8	381,7	281,8	92,5	11,6	17,6	0,4	0,3	20,1	7,4	30,5
2007	47,8	301,4	69,5	94,1	60,1	13,4	0,9	0,5	0,6	3,3	17,1	39,5
2008	156,4	273,1	614,6	202,6	111,9	5,5	30,3	4,7	0,8	0,0	0,0	45,4
2009	132,3	149,6	144,8	347,1	181,8	10,4	21,0	20,0	0,0	4,3	8,9	53,4
2010	189,5	27,2	68,7	129,6	29,2	42,9	1,0	0,0	0,0	99,7	0,0	133,1
2011	327,6	266,1	61,7	139,5	105,8	16,0	42,5	11,0	0,0	129,2	44,9	17,4
2012	88,1	158,3	192,1	55,4	55,6	26,6	5,1	0,0	0,2	1,3	0,0	2,7
2013	63,1	80,4	220,7	132,2	57,9	70,8	10,8	2,1	1,2	3,1	8,1	98,2
2014	39,4	131,6	227,8	166,7	55,5	16,2	29,4	10,2	4,1	27,0	138,0	90,1
2015	45,8	75,9	256,6	43,3	51,9	53,4	59,5	1,7	0,0	1,5	2,3	16,5
2016	170,5	39,1	192,4	104,6	55,0	2,3	0,6	0,0	1,0	4,4	0,0	13,2
2017	179,5	101,3	108,8	151,8	39,0	31,2	19,8	0,0	0,0	0,4	7,1	2,6
2018	131,5	258,7	202,3	286,0	61,9	2,4	1,5	0,0	0,0	2,2	50,4	120,4
2019	83,1	118,3	233,1	197,1	76,6	29,2	11,3	5,5	4,9	1,4	20,8	26,0

Fonte: Autor (2022).

**Tabela A2 - Série histórica de precipitações em mm para o reservatório São Gonçalo**

Ano/mês	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
1963	187,5	340,7	402,5	142,8	27,9	19,6	0,0	0,0	2,4	24,7	38,9	116,9
1964	91,5	268,6	277,8	223,9	98,1	46,1	30,1	10,3	46,6	17,5	2,0	2,8
1965	115,9	25,7	188,5	387,3	18,7	98,8	19,4	0,6	0,6	70,8	0,0	15,0
1966	43,0	141,0	87,9	78,1	78,1	79,4	26,1	1,7	8,8	4,2	9,6	41,3
1967	96,1	254,9	308,1	342,2	298,3	4,0	10,8	0,0	0,0	0,0	0,9	12,8
1968	158,8	71,5	422,9	81,1	81,0	8,4	21,0	0,9	0,0	0,0	5,8	29,8
1969	130,0	189,4	159,1	301,5	73,2	39,0	45,6	6,4	0,2	10,9	0,0	8,7
1970	149,1	100,7	243,9	78,4	0,0	22,7	0,6	3,7	2,5	12,2	10,0	4,8
1971	234,8	230,1	245,9	262,8	73,9	177,5	24,1	0,5	9,0	42,2	1,4	11,6
1972	201,4	150,7	221,0	104,5	79,4	48,3	2,3	39,4	0,0	0,1	0,0	133,2
1973	23,1	158,1	154,2	498,9	85,6	67,4	46,6	5,5	19,0	34,5	0,3	74,7
1974	252,1	217,6	265,4	445,8	113,2	11,0	16,5	0,0	25,3	13,1	35,5	7,6
1975	177,6	128,1	258,5	207,4	59,0	37,0	111,1	1,0	8,5	4,9	0,0	23,2
1976	26,7	330,2	304,5	89,6	4,4	12,0	1,0	0,0	15,1	58,2	20,5	24,3
1977	175,4	263,3	327,8	204,8	218,4	43,3	33,7	0,3	0,0	2,9	0,5	132,9
1978	162,2	423,9	195,0	48,4	113,9	19,0	42,2	1,8	0,1	2,7	30,2	5,0
1979	218,5	300,7	188,7	113,4	90,1	20,7	41,2	5,4	17,7	18,0	37,7	5,4
1980	110,2	327,1	217,9	17,8	11,7	35,5	0,3	0,0	4,9	15,5	33,4	14,1
1981	127,2	86,5	523,2	50,9	0,6	6,4	0,8	5,2	0,0	0,0	55,1	55,0
1982	109,4	112,4	204,3	197,9	48,6	4,6	3,4	11,5	0,0	12,5	11,1	1,0
1983	63,2	206,2	110,5	50,8	3,0	0,9	0,3	7,7	0,0	1,6	0,0	22,4
1984	95,7	40,9	245,8	245,0	141,1	5,9	6,3	4,5	0,5	6,6	43,1	98,0
1985	241,3	284,8	307,7	518,0	153,4	88,8	40,4	8,0	10,6	0,3	4,6	130,5
1986	85,7	240,8	317,9	216,4	55,3	63,2	39,5	13,7	5,6	21,9	65,0	16,7
1987	74,9	124,9	328,5	78,1	15,2	45,6	1,2	0,2	0,3	0,0	0,6	0,3
1988	79,3	108,9	239,5	367,1	75,3	7,2	6,7	0,0	0,5	0,8	0,5	103,1
1989	122,3	45,8	170,2	303,9	167,8	24,0	14,9	5,4	24,6	11,5	4,1	118,4
1990	8,1	110,9	137,5	298,1	35,5	4,4	6,7	12,7	27,0	2,6	38,4	2,5
1991	138,1	108,3	230,4	118,3	179,3	3,5	0,5	2,6	0,2	0,9	10,1	2,6
1992	76,5	180,9	200,7	109,3	32,7	77,3	28,6	12,1	10,7	1,9	8,0	2,3
1993	39,7	31,9	57,5	61,6	49,5	15,8	37,6	4,3	12,5	30,3	5,7	10,1
1994	373,1	174,1	85,0	110,0	87,6	169,8	42,5	6,9	1,7	0,0	0,0	115,1
1995	105,3	136,5	197,6	216,8	241,8	20,3	19,3	0,2	0,0	3,1	28,8	0,0
1996	243,4	239,6	197,4	389,3	135,9	19,4	5,6	12,2	0,4	10,0	21,1	34,1
1997	264,3	79,2	262,4	159,6	117,4	2,8	13,6	12,6	0,0	3,2	22,2	66,4
1998	209,2	89,0	84,4	98,5	8,4	9,5	7,0	8,5	0,0	1,8	0,0	15,9
1999	154,6	76,3	264,0	78,9	221,1	8,1	1,0	0,0	7,9	29,7	49,1	149,7
2000	173,0	299,3	214,4	172,1	49,9	31,5	35,0	32,4	12,8	7,4	0,0	67,3
2001	12,2	134,7	300,8	47,9	3,4	30,4	7,6	9,1	0,0	14,7	1,1	110,3
2002	329,4	87,9	229,0	136,6	158,4	35,0	5,3	0,0	0,3	1,0	6,3	14,3
2003	160,8	169,2	232,7	216,6	49,7	17,1	18,0	3,4	0,4	0,0	5,6	17,0
2004	294,4	261,2	213,6	84,0	60,0	92,8	15,8	1,1	0,0	0,0	0,0	44,6
2005	41,5	25,4	473,0	123,2	77,1	16,9	0,3	3,4	1,3	0,0	0,0	42,0
2006	74,3	196,9	289,9	319,6	123,1	8,2	11,6	7,8	0,0	5,7	8,3	88,4
2007	62,9	321,0	104,2	136,4	80,8	40,7	5,1	0,0	0,0	0,0	43,0	14,7
2008	175,3	161,1	624,1	242,4	228,3	18,5	25,5	4,3	0,0	2,6	0,0	51,6
2009	121,4	159,6	188,3	329,4	255,1	23,8	21,6	14,8	0,0	0,0	0,0	57,9
2010	72,1	50,3	116,3	154,3	15,7	64,4	0,0	0,0	0,0	78,9	0,0	121,9
2011	339,0	360,4	81,9	145,5	88,0	41,5	22,0	3,3	0,0	180,6	20,3	15,6
2012	94,2	165,3	78,6	111,1	27,9	22,5	4,7	0,0	0,0	0,0	0,1	0,5
2013	52,1	76,5	203,7	102,5	87,7	43,0	11,4	4,0	0,8	21,1	8,0	92,9
2014	63,6	124,2	218,1	149,2	62,0	16,5	21,9	0,1	8,0	43,8	49,9	21,6
2015	61,3	103,6	304,3	54,9	21,4	42,3	47,7	2,3	0,0	2,6	0,0	13,8
2016	171,1	32,5	328,7	120,4	46,9	4,7	0,2	0,0	0,0	5,6	0,0	6,1
2017	112,2	132,2	97,2	116,2	44,4	24,5	45,2	0,0	0,0	0,0	4,9	4,2
2018	107,9	286,0	97,4	288,4	36,8	11,8	1,1	0,0	0,0	0,0	17,1	104,4
2019	92,2	164,4	231,9	154,8	90,4	18,5	21,7	3,6	3,4	0,1	7,3	37,9

Fonte: Autor (2022).

**Tabela A3** - Equações de correlação para obtenção das precipitações de São Gonçalo

	Parâmetros	Equação	R <sup>2</sup>	Nº iterações
Jan	a = 0,0815 b = 4,6550	$Y = a * (Ln(X))^b$	0,786	14
Fev	a = 0,1698*10 <sup>-5</sup> b = -236,5767 c = 0,0157 d = 0,7742	$Y = \frac{1}{a * (X + b)^2 + c} + d * X$	0,806	44
Mar	a = 296.0734 b = 734.8427 c = 0.4911	$Y = a * \left(\frac{X}{b}\right)^c * e^{\left(\frac{X}{b}\right)}$	0,7899	42
Abr	a = -0,2478*10 <sup>4</sup> b = 0,1058*10 <sup>4</sup> c = -0,5155*10 <sup>3</sup> d = 0,1198*10 <sup>3</sup>	$Y = \frac{a}{X} + b * e^{\left(\frac{c}{X}\right)} + d$	0,8711	8
Mai	a = 1,3227 b = -0,2193*10 <sup>2</sup> c = 0,5339*10 <sup>-5</sup>	$Y = a * X + b * X^2 + c * X^3$	0,7659	7
Jun	a = 0,1545*10 <sup>-5</sup> b = -0,3559*10 <sup>4</sup> c = 1,0319	$Y = \frac{1}{(a + b * LnX)} + c * X$	0,7495	27
Jul	a = 2,3837 b = 0,9945 c = 0,4563*10 <sup>-4</sup>	$Y = \frac{(a * X)}{(b + c * X^2)}$	0,6122	7
Ago	a = 4,5879 b = 0,04917	$Y = A * Cosh(b * X)$	0,9279	5
Set	a = 0,08731 b = 0,7459 c = 0,02888	$Y = (a * X)/(b + c * X)$	0,6015	8
Out	a = 0,56263 b = 0,5346*10 <sup>-2</sup>	$Y = a * X + b * X^2$	0,8134	6
Nov	a = 1,92763 b = 1,07692 c = -0,5668*10 <sup>-2</sup> d = -0,012	$Y = \frac{a * X}{b + c * X^2} + d * X$	0,6175	22
Dez	a = 1,8612 b = 0,8216 c = 1,3154	$Y = a * X^{b+c}$	0,7228	41

Fonte: Autor (2022).

**Tabela A4** - Propriedades estatísticas da SH do reservatório Engenheiro Ávidos

Mês	$\bar{X}$	S	med	mad	min	max	skew	Curt	q1	q3	C. Var.	C. Ass	% zeros
Jan	143,52	86,36	131,46	59,71	16,85	387,90	1,28	1,28	91,19	170,50	1,66	1,28	0,00
Fev	177,17	105,39	144,46	83,52	27,15	454,45	0,82	0,82	97,23	240,84	1,68	0,82	0,00
Mar	229,19	111,82	220,68	84,07	61,69	614,59	0,89	0,89	165,76	277,42	2,05	0,89	0,00
Abr	184,94	115,01	140,50	95,67	33,59	529,47	0,92	0,92	96,97	256,80	1,61	0,92	0,00
Mai	74,84	55,59	60,10	41,40	0,25	256,91	1,09	1,09	38,98	99,78	1,35	1,09	0,00
Jun	30,63	26,21	22,40	23,02	0,75	100,71	1,00	1,00	10,84	42,88	1,17	1,00	0,00
Jul	17,70	18,81	11,34	14,24	0,07	92,78	1,83	1,83	3,75	23,53	0,94	1,83	0,00
Ago	5,63	8,47	2,07	3,06	0,00	39,50	2,31	2,31	0,42	6,66	0,67	2,31	19,30
Set	7,62	15,14	1,12	1,66	0,00	85,52	3,00	3,00	0,00	5,94	0,50	3,00	26,32
Out	14,97	24,06	3,57	5,29	0,00	129,24	2,75	2,75	1,18	21,78	0,62	2,75	14,04
Nov	19,25	26,76	8,09	12,00	0,00	137,96	2,17	2,17	1,31	28,62	0,72	2,17	17,54
Dez	50,36	52,51	30,49	31,08	0,59	183,93	1,11	1,11	10,63	89,45	0,96	1,11	0,00

Fonte: Autor (2022).

**Tabela A5 - Dados de entrada para modelagem SMAP**

Reservatórios	Dados Iniciais				
	Ad (Km <sup>2</sup> )	Tu (%)	Eb (m <sup>3</sup> /s)	Ai (mm)	Cc (%)
Eng. Ávidos	942,11				
São Gonçalo	306,1	0	1,051	4	50

Fonte: AESA (2016)

**Tabela A6 - Dados do posto fluviométrico de referência para modelagem SMAP**

Estação	Período	Parâmetros			
Jardim de Piranhas	Jan/94 a dez/2007	Sat (mm)	Pes	CREC	k
		1250	4,5	0,0	1

Fonte: AESA (2016)

**Tabela A7 - Propriedades estatísticas da SH do reservatório São Gonçalo**

Mês	$\bar{X}$	S	med	mad	min	max	skew	Curt	q1	q3	C. Var	C. Ass	% zeros
Jan	136,42	84,22	115,88	78,04	8,12	373,10	0,82	0,16	74,91	175,45	1,62	0,82	0,00
Fev	170,39	97,50	158,10	106,14	25,39	423,90	0,51	-0,63	100,72	240,76	1,75	0,51	0,00
Mar	232,67	111,21	220,97	91,68	57,48	624,08	1,03	1,76	170,17	289,89	2,09	1,03	0,00
Abr	184,26	120,09	145,46	99,94	17,76	517,99	0,96	0,15	98,54	244,96	1,53	0,96	0,00
Mai	85,99	70,81	75,33	59,02	0,00	298,31	1,09	0,58	35,52	113,91	1,21	1,09	0,18
Jun	34,60	36,23	22,51	22,75	0,94	177,55	2,15	5,26	10,98	43,01	0,95	2,15	0,00
Jul	18,78	19,73	14,90	18,69	0,00	111,06	1,91	6,06	3,43	28,64	0,95	1,91	0,35
Ago	5,19	7,35	3,38	4,95	0,00	39,38	2,68	8,74	0,14	7,71	0,71	2,68	2,11
Set	5,09	9,03	0,38	0,56	0,00	46,63	2,40	6,58	0,00	7,95	0,56	2,40	4,04
Out	14,65	28,47	3,24	4,81	0,00	180,58	3,86	18,12	0,14	15,51	0,51	3,86	2,11
Nov	13,44	17,44	5,68	8,42	0,00	64,95	1,23	0,31	0,00	21,06	0,77	1,23	2,63
Dez	44,48	45,32	22,39	29,14	0,00	149,66	0,84	0,00	8,74	74,70	0,98	0,84	0,18

Fonte: Autor (2022).

**Tabela A8 - Coeficientes do modelo RLM para a série temporal**

Parâmetros	Coeficientes	Erro padrão	Teste T	p-valor
$\alpha_1$	153,112	11,068	13,834	$<-2,00*10^{-16}$
$\gamma_1$	-0,026	0,018	-1,454	0,14644
$\beta_2$	42,135	14,049	6,723	$2,83*10^{-03}$
$\beta_3$	94,450	14,049	3,447	$4,46*10^{-11}$
$\beta_4$	48,429	14,049	-4,817	$6,10*10^{-04}$
$\beta_5$	-67,669	14,049	-8,181	$1,89*10^{-06}$
$\beta_6$	-114,929	14,049	-9,129	$1,96*10^{-15}$
$\beta_7$	-128,258	14,049	-9,940	$<-2*10^{-16}$
$\beta_8$	-139,652	14,049	-9,738	$<-2*10^{-16}$
$\beta_9$	-136,812	14,049	-9,493	$<-2*10^{-16}$
$\beta_{10}$	-133,378	14,049	-9,124	$<-2*10^{-16}$
$\beta_{11}$	-128,189	14,098	-6,813	$<-2*10^{-16}$
$\beta_{12}$	-95,725	14,050	2,999	$2,51*10^{-11}$

Fonte: Autor (2022)

**Tabela A9 - Valores das previsões, limites inferiores e superiores da SH Eng. Ávidos**

Meses	Previsão	Inf. 80%	Sup. 80%	Inf. 80%	Sup. 95%
Jan	137,25	48,6	225,9	1,54	272,97
Fev	179,35	90,72	268	43,64	315,08
Mar	226,42	137,77	315,07	90,71	362,14
Abr	185,61	96,96	274,25	49,89	321,31
Mai	69,47	0,00	158,13	0,00	205,19
Jun	22,19	0,00	110,85	0,00	157,91
Jul	8,85	0,00	97,49	0,00	144,55
Ago	0,00	0,00	86,08	0,00	133,14
Set	0,25	0,00	88,89	0,00	135,95
Out	3,65	0,00	92,29	0,00	139,35
Nov	8,81	0,00	97,46	0,00	144,52
Dez	41,25	0,00	129,89	0,00	176,95

Fonte: Autor (2022).

**Tabela A10 - Características dos dados de séries sintéticas (mm) obtidos pelo modelo TF**

Mês	$\bar{X}$	<i>S</i>	<i>med</i>	<i>mad</i>	<i>min</i>	<i>máx.</i>	<i>skew</i>	<i>Curt</i>	<i>CVar</i>	<i>%Zeros</i>
Jan	144,1	86,2	126,5	75,9	2,6	1070,2	81,3	188,3	1,7	0,0
Fev	177,1	104,5	163,1	99,8	0,0	890,2	101,1	237,6	1,7	13,4
Mar	229,2	112,0	213,0	105,4	0,0	1089,5	147,9	292,6	2,0	0,2
Abr	185,3	114,9	167,6	107,9	0,0	1154,7	101,6	250,2	1,6	14,3
Mai	75,5	55,3	65,7	50,9	0,0	568,7	34,7	105,1	1,4	41,9
Jun	31,0	25,4	26,3	24,3	0,0	230,6	11,6	45,1	1,2	87,2
Jul	18,0	18,5	13,2	14,2	0,0	245,9	4,9	25,3	1,0	109,3
Ago	6,0	8,3	2,9	4,3	0,0	128,8	0,0	8,4	0,7	259,7
Set	8,5	14,5	2,6	3,9	0,0	249,7	0,0	10,9	0,6	364,3
Out	15,5	23,4	6,6	9,8	0,0	334,5	0,0	20,9	0,7	261,6
Nov	20,1	25,8	11,3	16,8	0,0	418,2	1,1	28,6	0,8	219,5
Dez	52,0	49,8	40,6	47,7	0,0	435,1	11,7	78,0	1,0	155,9

Fonte: Autor (2022).

**Tabela A11 - Características dos dados de séries sintéticas (mm) obtidos pelo modelo MDF**

Mês	$\bar{X}$	<i>S</i>	<i>med</i>	<i>mad</i>	<i>min</i>	<i>máx.</i>	<i>skew</i>	<i>Curt</i>	<i>CVar</i>	<i>%Zeros</i>
Jan	141,4	86,4	126,2	62,2	12,0	512,9	86,0	170,5	1,6	0,0
Fev	167,2	100,7	130,2	81,9	27,0	460,5	90,5	232,0	1,7	0,0
Mar	243,1	112,1	232,3	99,6	32,0	626,6	176,0	305,0	2,2	0,0
Abr	181,7	119,1	133,4	111,3	33,3	772,3	85,8	274,1	1,5	0,0
Mai	75,8	59,5	61,5	52,6	0,3	262,1	32,9	100,8	1,3	0,0
Jun	30,0	26,2	22,4	23,3	0,7	101,6	9,1	40,0	1,1	0,0
Jul	16,0	17,3	10,8	12,5	0,1	92,8	3,1	23,1	0,9	0,0
Ago	4,3	5,7	2,8	3,4	0,0	39,5	0,6	5,4	0,8	15,5
Set	6,6	15,6	0,8	1,2	0,0	86,2	0,0	4,7	0,4	29,6
Out	14,8	23,3	3,5	5,2	0,0	131,3	1,0	25,2	0,6	15,4
Nov	19,6	25,5	7,9	11,7	0,0	138,3	0,6	30,5	0,8	18,6
Dez	54,7	51,8	31,0	31,9	0,6	268,3	15,4	97,8	1,1	0,0

Fonte: Autor (2022).

**Tabela A12 – Fragmentos da série histórica utilizados no modelo MDF**

Anos	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
1963	0,099	0,360	0,304	0,067	0,026	0,021	0,000	0,000	0,001	0,017	0,029	0,078
1964	0,124	0,167	0,272	0,224	0,096	0,033	0,012	0,002	0,036	0,007	0,018	0,009
1965	0,123	0,033	0,259	0,366	0,031	0,104	0,014	0,024	0,001	0,043	0,000	0,001
1966	0,149	0,177	0,099	0,234	0,090	0,103	0,031	0,000	0,032	0,002	0,003	0,080
1967	0,075	0,166	0,221	0,322	0,185	0,004	0,004	0,002	0,000	0,000	0,000	0,021
1968	0,150	0,051	0,518	0,081	0,083	0,022	0,012	0,001	0,000	0,001	0,041	0,039
1969	0,115	0,197	0,185	0,265	0,081	0,067	0,032	0,008	0,003	0,014	0,000	0,032
1970	0,229	0,147	0,300	0,192	0,002	0,006	0,005	0,000	0,005	0,049	0,015	0,049
1971	0,168	0,213	0,134	0,274	0,062	0,070	0,025	0,001	0,004	0,043	0,002	0,003
1972	0,192	0,165	0,239	0,128	0,041	0,027	0,021	0,023	0,003	0,003	0,000	0,157
1973	0,050	0,117	0,183	0,393	0,068	0,060	0,031	0,031	0,025	0,001	0,001	0,039
1974	0,128	0,246	0,221	0,268	0,038	0,014	0,005	0,000	0,002	0,016	0,031	0,027
1975	0,133	0,167	0,286	0,180	0,068	0,052	0,064	0,000	0,005	0,014	0,001	0,030
1976	0,066	0,335	0,268	0,103	0,017	0,018	0,000	0,001	0,083	0,039	0,064	0,006
1977	0,103	0,152	0,201	0,228	0,120	0,074	0,037	0,001	0,000	0,009	0,000	0,075
1978	0,081	0,376	0,153	0,115	0,120	0,029	0,080	0,002	0,005	0,001	0,027	0,009
1979	0,231	0,255	0,197	0,149	0,059	0,009	0,003	0,007	0,031	0,039	0,013	0,008
1980	0,103	0,419	0,240	0,034	0,009	0,033	0,024	0,000	0,002	0,033	0,091	0,012
1981	0,127	0,086	0,526	0,056	0,000	0,016	0,001	0,005	0,000	0,000	0,032	0,150
1982	0,146	0,153	0,256	0,242	0,049	0,007	0,003	0,002	0,005	0,072	0,045	0,020
1983	0,172	0,372	0,170	0,204	0,010	0,026	0,019	0,004	0,001	0,000	0,003	0,018
1984	0,150	0,148	0,258	0,280	0,084	0,004	0,009	0,005	0,000	0,010	0,030	0,021
1985	0,173	0,156	0,191	0,261	0,070	0,030	0,016	0,007	0,001	0,000	0,005	0,091
1986	0,073	0,238	0,252	0,197	0,048	0,047	0,026	0,014	0,015	0,027	0,061	0,004
1987	0,120	0,182	0,428	0,172	0,021	0,071	0,002	0,001	0,001	0,000	0,000	0,001
1988	0,086	0,144	0,256	0,315	0,043	0,004	0,011	0,000	0,002	0,003	0,001	0,133
1989	0,139	0,038	0,172	0,254	0,161	0,018	0,014	0,009	0,028	0,023	0,002	0,140
1990	0,026	0,223	0,097	0,396	0,065	0,017	0,006	0,061	0,046	0,008	0,043	0,013
1991	0,231	0,177	0,307	0,120	0,132	0,005	0,001	0,003	0,000	0,000	0,017	0,006
1992	0,189	0,244	0,243	0,137	0,029	0,084	0,029	0,010	0,018	0,002	0,010	0,005
1993	0,064	0,214	0,172	0,182	0,143	0,021	0,084	0,008	0,031	0,048	0,016	0,018
1994	0,318	0,159	0,097	0,112	0,133	0,084	0,019	0,002	0,000	0,001	0,001	0,075
1995	0,151	0,142	0,169	0,248	0,208	0,017	0,019	0,000	0,000	0,000	0,036	0,010
1996	0,164	0,149	0,219	0,252	0,156	0,022	0,002	0,003	0,000	0,004	0,015	0,014
1997	0,271	0,107	0,215	0,147	0,115	0,001	0,016	0,006	0,000	0,004	0,082	0,036
1998	0,284	0,185	0,235	0,223	0,005	0,013	0,019	0,000	0,000	0,000	0,000	0,036
1999	0,178	0,116	0,263	0,097	0,110	0,007	0,003	0,003	0,006	0,008	0,033	0,175
2000	0,131	0,245	0,203	0,220	0,031	0,018	0,031	0,011	0,033	0,003	0,000	0,075
2001	0,065	0,156	0,394	0,072	0,012	0,050	0,003	0,005	0,000	0,038	0,010	0,196
2002	0,425	0,100	0,219	0,064	0,112	0,039	0,007	0,000	0,003	0,001	0,005	0,024
2003	0,218	0,103	0,266	0,191	0,065	0,022	0,007	0,002	0,013	0,021	0,023	0,069
2004	0,342	0,307	0,108	0,049	0,082	0,038	0,022	0,008	0,001	0,000	0,010	0,033
2005	0,114	0,087	0,471	0,093	0,059	0,033	0,005	0,006	0,000	0,000	0,000	0,131
2006	0,048	0,211	0,335	0,247	0,081	0,010	0,015	0,000	0,000	0,018	0,007	0,027
2007	0,074	0,465	0,107	0,145	0,093	0,021	0,001	0,001	0,001	0,005	0,026	0,061
2008	0,108	0,189	0,425	0,140	0,077	0,004	0,021	0,003	0,001	0,000	0,000	0,031
2009	0,123	0,139	0,135	0,323	0,169	0,010	0,020	0,019	0,000	0,004	0,008	0,050
2010	0,263	0,038	0,095	0,180	0,040	0,059	0,001	0,000	0,000	0,138	0,000	0,185
2011	0,282	0,229	0,053	0,120	0,091	0,014	0,037	0,009	0,000	0,111	0,039	0,015
2012	0,151	0,270	0,328	0,095	0,095	0,045	0,009	0,000	0,000	0,002	0,000	0,005
2013	0,084	0,107	0,295	0,177	0,077	0,095	0,014	0,003	0,002	0,004	0,011	0,131
2014	0,042	0,141	0,243	0,178	0,059	0,017	0,031	0,011	0,004	0,029	0,147	0,096
2015	0,075	0,125	0,422	0,071	0,085	0,088	0,098	0,003	0,000	0,002	0,004	0,027
2016	0,292	0,067	0,330	0,179	0,094	0,004	0,001	0,000	0,002	0,007	0,000	0,023
2017	0,280	0,158	0,170	0,237	0,061	0,049	0,031	0,000	0,000	0,001	0,011	0,004
2018	0,118	0,232	0,181	0,256	0,055	0,002	0,001	0,000	0,000	0,002	0,045	0,108
2019	0,103	0,147	0,289	0,244	0,095	0,036	0,014	0,007	0,006	0,002	0,026	0,032

Fonte: Autor (2022).

**Tabela A13 - Propriedades estatísticas das séries sintéticas do São Gonçalo**

Mês	$\bar{X}$	<i>S</i>	<i>med</i>	<i>mad</i>	<i>min</i>	<i>máx</i>	<i>skew</i>	<i>Curt</i>	<i>q1</i>	<i>q3</i>	<i>C. Var.</i>	<i>C. Ass</i>	% zeros
Jan	136,4	84,2	115,9	78,0	8,1	373,1	0,8	0,2	74,9	175,4	1,6	0,8	0,0
Fev	170,4	97,5	158,1	106,1	25,4	423,9	0,5	-0,6	100,7	240,8	1,7	0,5	0,0
Mar	232,7	111,2	221,0	91,7	57,5	624,1	1,0	1,8	170,2	289,9	2,1	1,0	0,0
Abr	184,3	120,1	145,5	99,9	17,8	518,0	1,0	0,2	98,5	245,0	1,5	1,0	0,0
Mai	86,0	70,8	75,3	59,0	0,0	298,3	1,1	0,6	35,5	113,9	1,2	1,1	1754,4
Jun	34,6	36,2	22,5	22,7	0,9	177,5	2,2	5,3	11,0	43,0	1,0	2,2	0,0
Jul	18,8	19,7	14,9	18,7	0,0	111,1	1,9	6,1	3,4	28,6	1,0	1,9	3508,8
Ago	5,2	7,4	3,4	4,9	0,0	39,4	2,7	8,7	0,1	7,7	0,7	2,7	21052,6
Set	5,1	9,0	0,4	0,6	0,0	46,6	2,4	6,6	0,0	8,0	0,6	2,4	40350,9
Out	14,7	28,5	3,2	4,8	0,0	180,6	3,9	18,1	0,1	15,5	0,5	3,9	21052,6
Nov	13,4	17,4	5,7	8,4	0,0	65,0	1,2	0,3	0,0	21,1	0,8	1,2	26315,8
Dez	44,5	45,3	22,4	29,1	0,0	149,7	0,8	-0,8	8,7	74,7	1,0	0,8	1754,4

Fonte: Autor (2022).

**Tabela A14 - Propriedades estatísticas das vazões sintéticas para o Engenheiro Ávidos**

Mês	$\bar{X}$	<i>S</i>	<i>med</i>	<i>mad</i>	<i>min</i>	<i>max</i>	<i>Curt</i>	<i>q1</i>	<i>q3</i>	<i>C. Var.</i>	% zeros
Jan	0,49	0,75	0,21	0,25	0,00	13,98	25,96	0,07	0,63	0,65	0,12
Fev	1,32	2,26	0,53	0,58	0,00	49,67	16,57	0,19	1,25	0,59	0,00
Mar	4,60	7,42	2,15	2,36	0,00	116,82	17,87	0,89	4,99	0,62	0,00
Abr	8,22	14,75	2,37	2,68	0,00	224,21	13,94	1,08	8,75	0,56	0,00
Mai	3,74	6,12	1,44	1,94	0,00	39,45	10,94	0,30	4,22	0,61	0,15
Jun	0,89	1,40	0,41	0,44	0,00	12,05	8,31	0,13	0,81	0,63	0,02
Jul	0,38	0,65	0,11	0,15	0,00	4,56	6,96	0,02	0,43	0,58	0,63
Ago	0,06	0,11	0,02	0,03	0,00	1,25	20,49	0,00	0,06	0,52	2,72
Set	0,04	0,10	0,00	0,00	0,00	0,69	9,57	0,00	0,02	0,39	5,15
Out	0,05	0,11	0,01	0,01	0,00	1,02	19,63	0,00	0,06	0,48	4,69
Nov	0,04	0,08	0,01	0,01	0,00	0,51	5,88	0,00	0,04	0,52	4,85
Dez	0,09	0,15	0,03	0,04	0,00	3,19	28,12	0,01	0,11	0,57	1,90

Fonte: Autor (2022).

**Tabela A15 - Propriedades estatísticas das vazões sintéticas para o São Gonçalo**

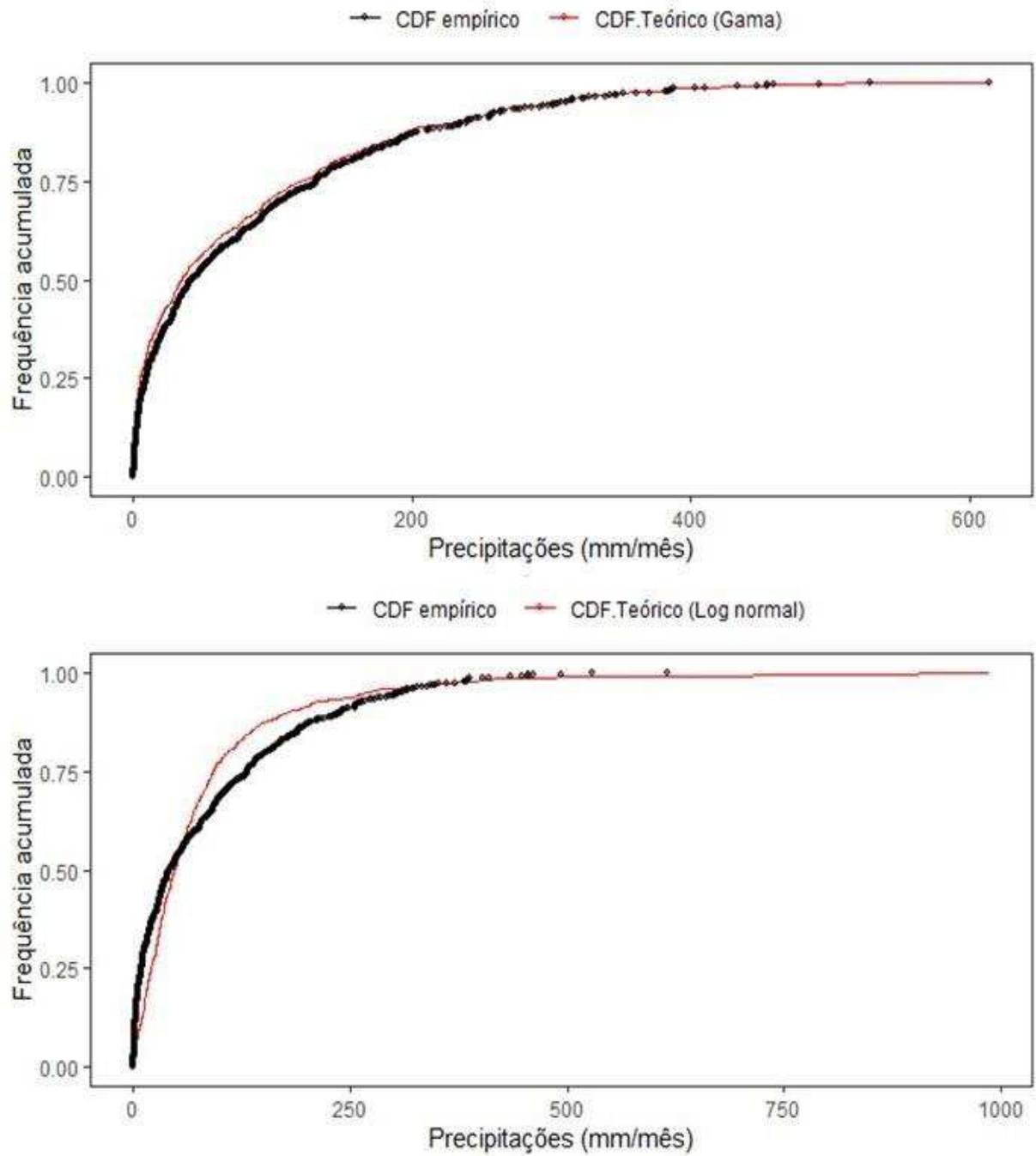
Mês	$\bar{X}$	<i>S</i>	<i>med</i>	<i>mad</i>	<i>min</i>	<i>max</i>	<i>Curt</i>	<i>q1</i>	<i>q3</i>	<i>C. Var.</i>	% zeros
Jan	0,18	0,32	0,05	0,06	0,00	2,20	3,21	0,02	0,14	0,58	0,67
Fev	0,32	0,41	0,16	0,19	0,00	6,69	9,84	0,05	0,49	0,77	0,07
Mar	1,17	1,87	0,61	0,55	0,00	29,42	29,93	0,31	1,28	0,62	0,00
Abr	2,31	3,97	0,68	0,68	0,00	53,98	12,34	0,34	2,58	0,58	0,00
Mai	1,32	2,18	0,51	0,65	0,00	14,07	11,35	0,12	1,46	0,61	0,48
Jun	0,29	0,46	0,13	0,13	0,00	4,27	8,68	0,05	0,25	0,62	0,65
Jul	0,12	0,18	0,04	0,04	0,00	1,47	5,53	0,02	0,16	0,67	0,52
Ago	0,02	0,02	0,01	0,01	0,00	0,14	2,98	0,01	0,03	1,06	2,09
Set	0,01	0,01	0,00	0,00	0,00	0,09	7,66	0,00	0,01	0,52	6,62
Out	0,01	0,01	0,00	0,00	0,00	0,12	13,20	0,00	0,01	0,48	7,02
Nov	0,01	0,01	0,00	0,00	0,00	0,05	4,23	0,00	0,01	0,57	6,51
Dez	0,02	0,03	0,01	0,01	0,00	0,48	24,03	0,00	0,02	0,63	3,84

Fonte: Autor (2022).

**Tabela A16 - Valores das métricas de desempenho dos modelos de geração estudados**

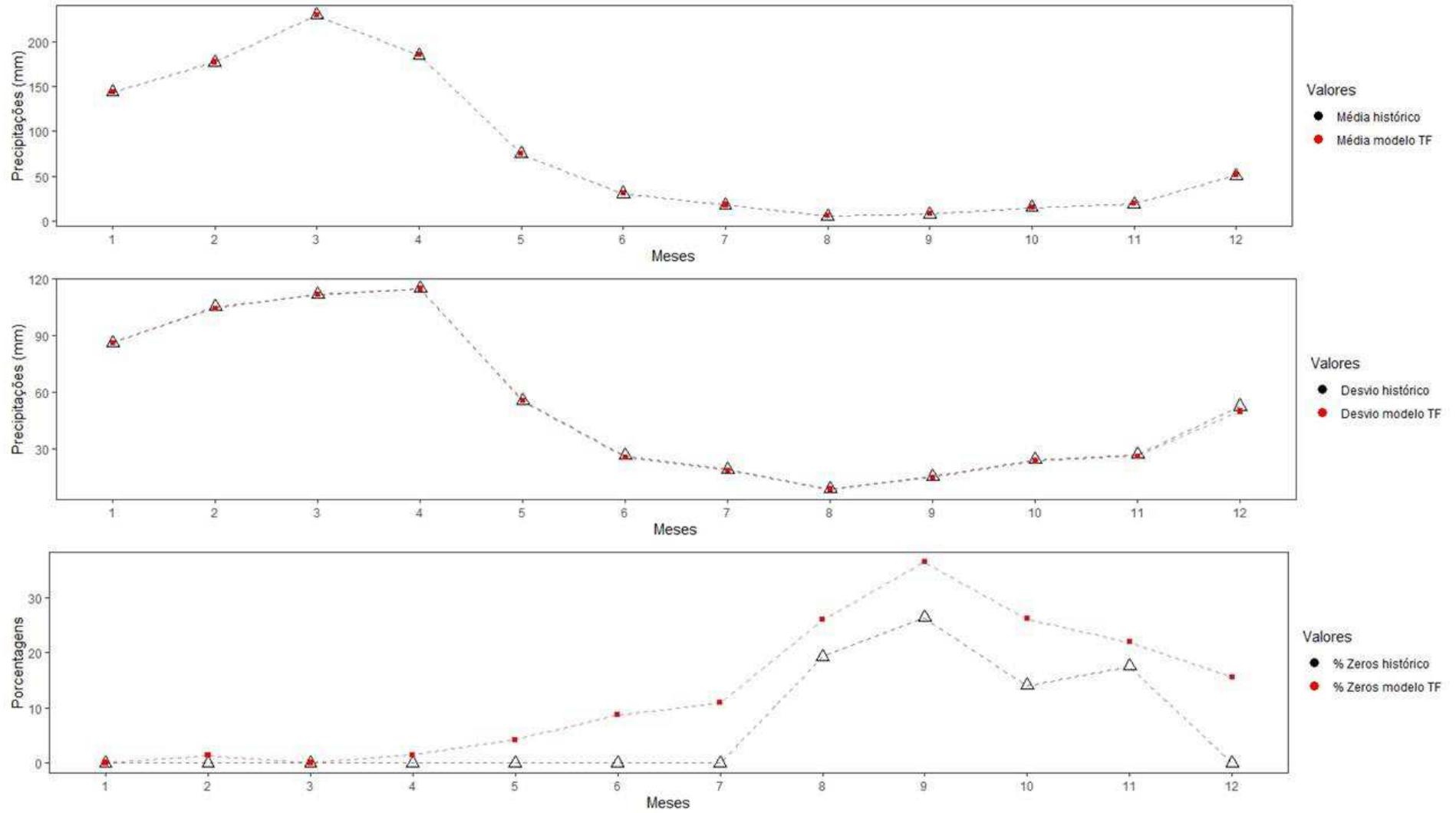
	$\bar{X}$		<i>S</i>		<i>q1</i>		<i>q2</i>		<i>Cvar</i>	
	TF	MDF	TF	MDF	TF	MDF	TF	MDF	TF	MDF
ME	-0,531	0,044	0,638	0,241	1,863	1,433	-2,233	-3,744	-0,038	-0,004
MAE	0,546	3,306	0,674	1,708	3,776	3,983	5,930	6,192	0,038	0,058
RMSE	0,680	5,279	0,958	2,347	6,299	5,524	8,150	10,162	0,047	0,068
MAPE	2,713	6,281	2,094	5,693	-	-	13,682	8,508	4,969	6,050

Fonte: Autor (2022).

**Figura A1** - CDF da amostra comparada com a da distribuição

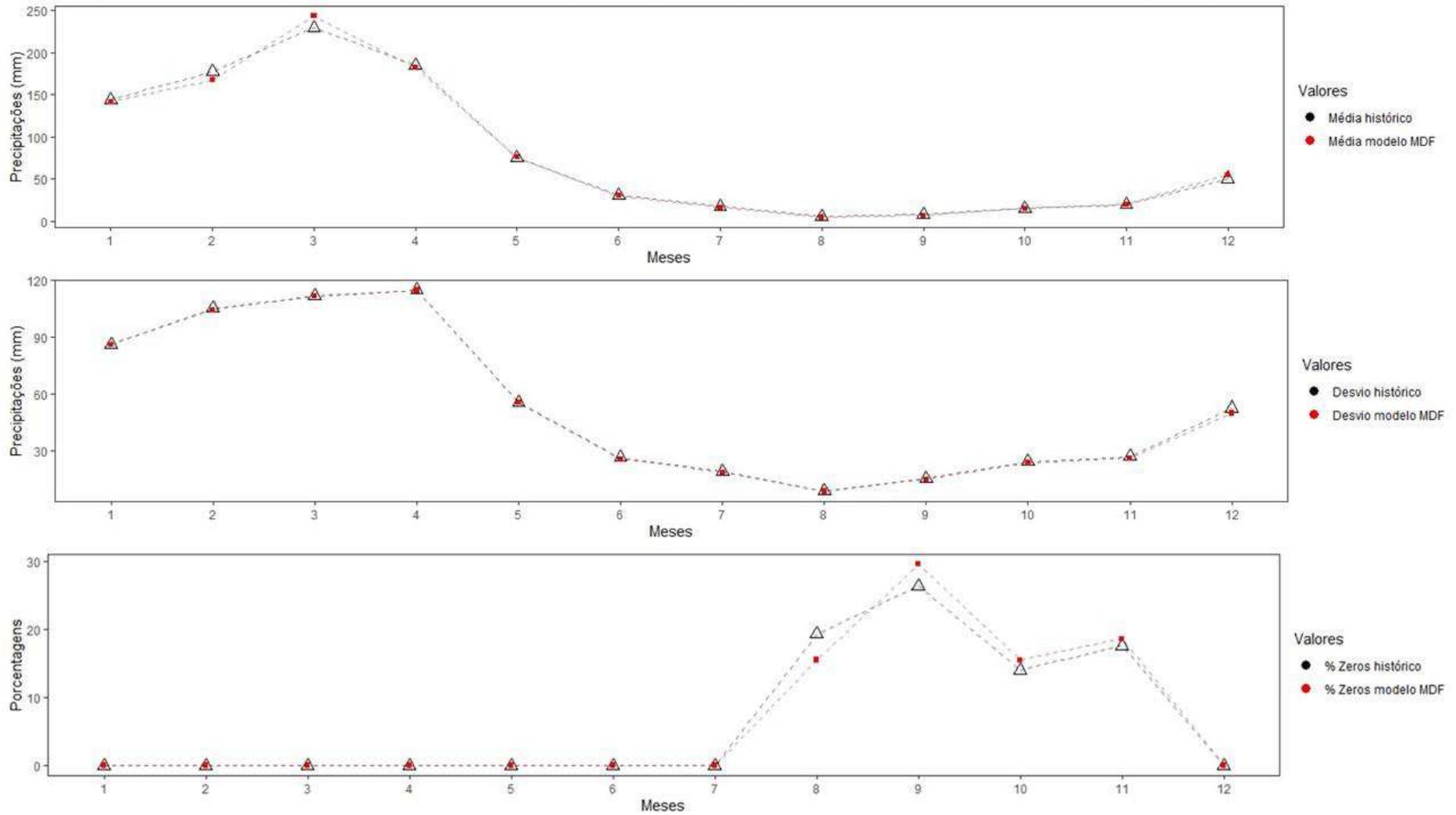
Fonte: Autor (2022).

**Figura A2 - Comparação mensal entre média, desvio padrão e percentagem de zeros das SH e sintéticas gerada pelo TF**

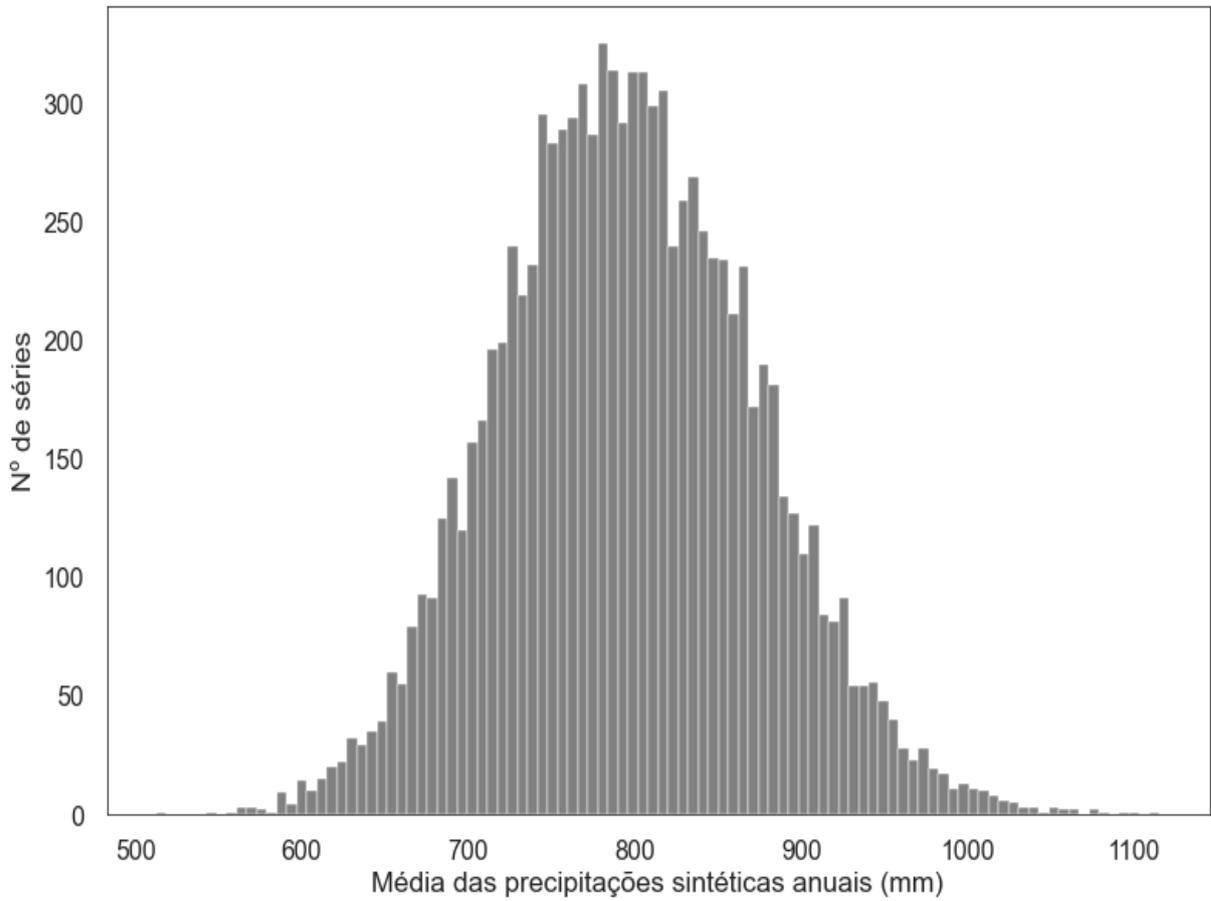


Fonte: Autor (2022)

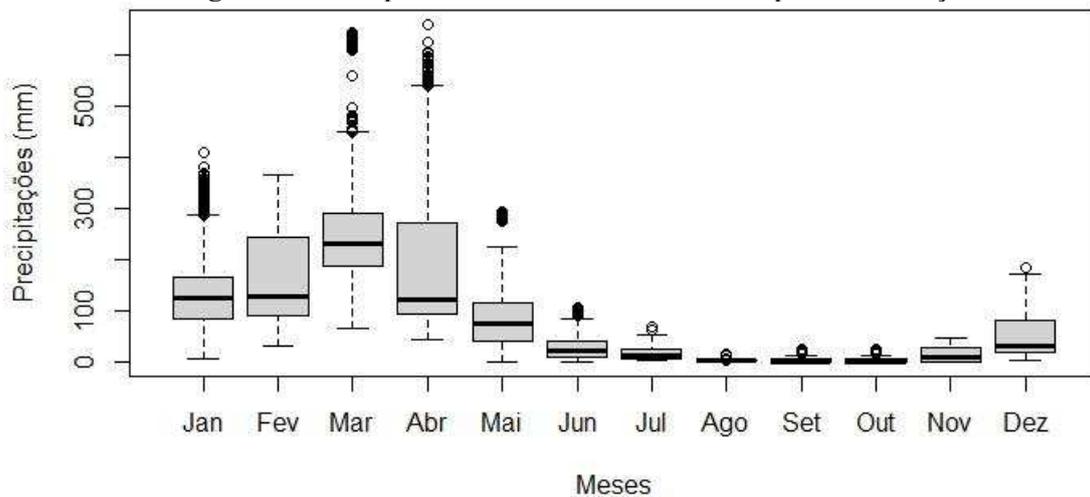
Figura A3 - Comparação mensal entre média, desvio padrão e percentagem de zeros das séries histórica e sintéticas gerada pelo modelo MDF



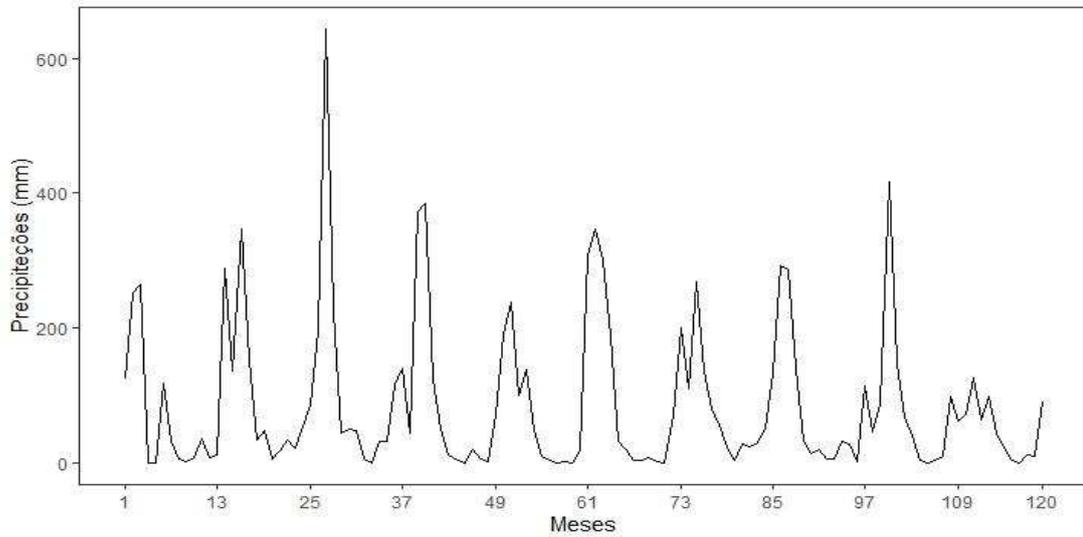
Fonte: Autor (2022).

**Figura A4 – Médias das precipitações sintéticas anuais**

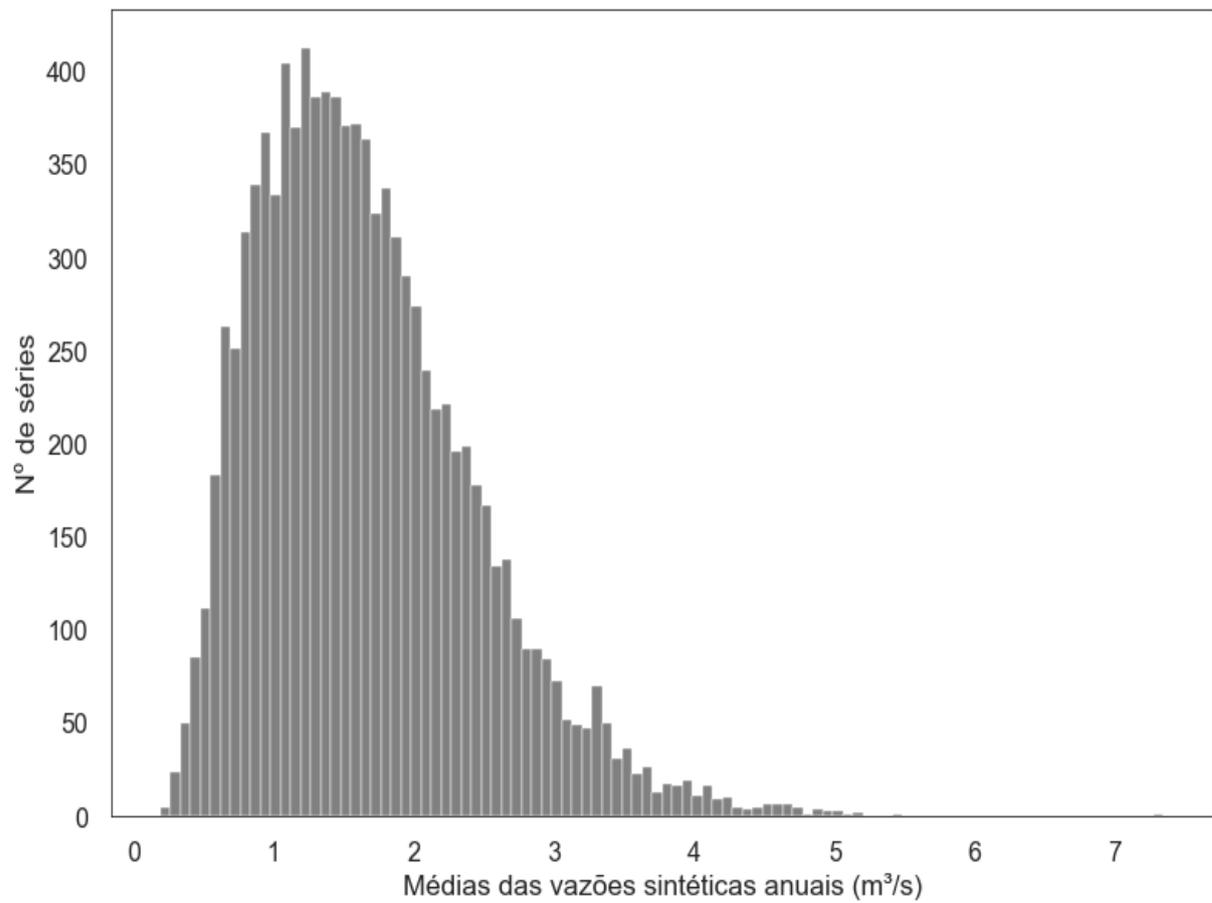
Fonte: Autor (2022)

**Figura A5 - Boxplot das séries sintéticas tratadas para São Gonçalo**

Fonte: Autor (2022).

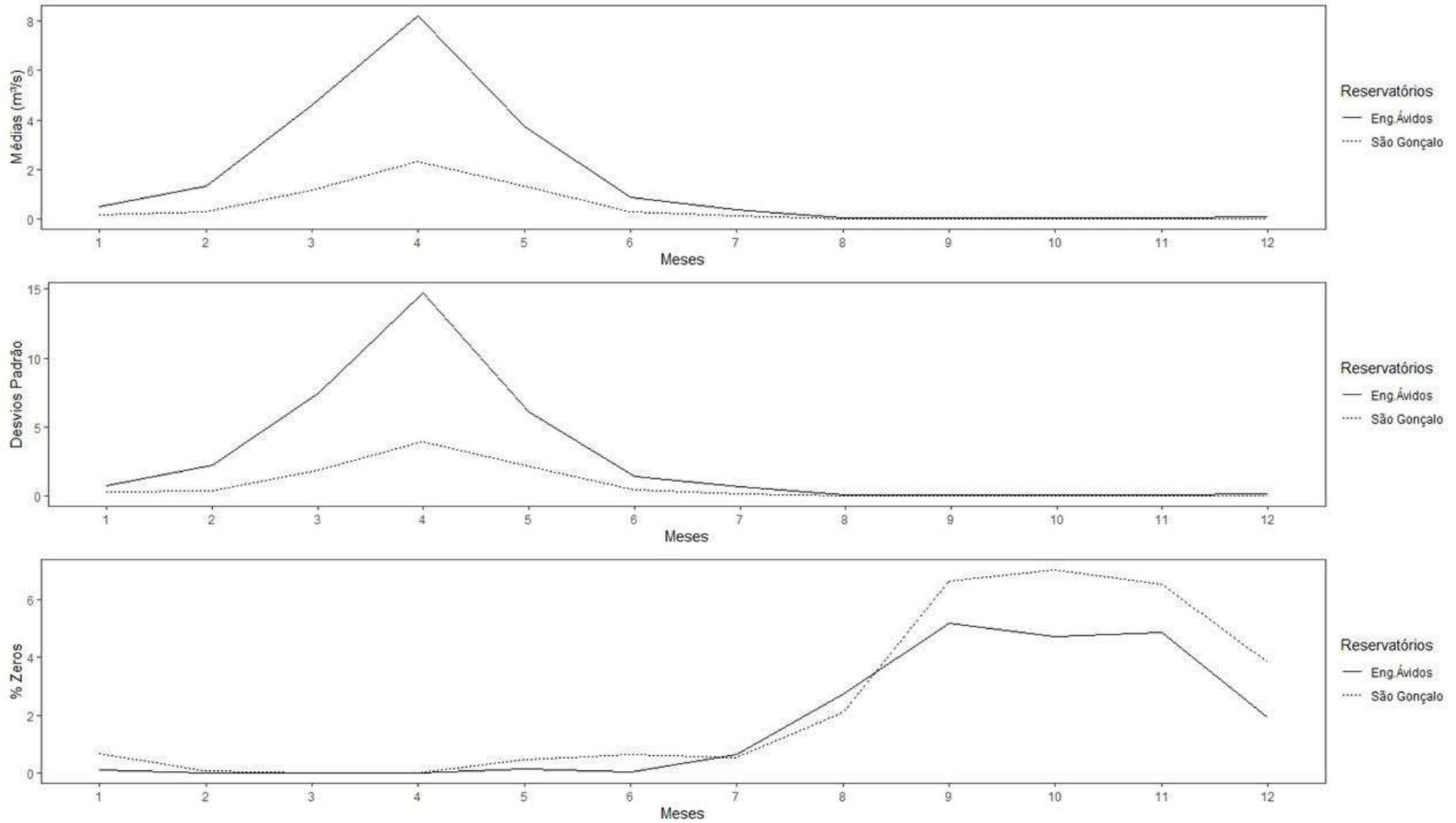
**Figura A6 -** Representação de uma série pluviométrica sintética do Reservatório São Gonçalo

Fonte: Autor (2022).

**Figura A7 –** Média das vazões anuais sintéticas para cada cenário

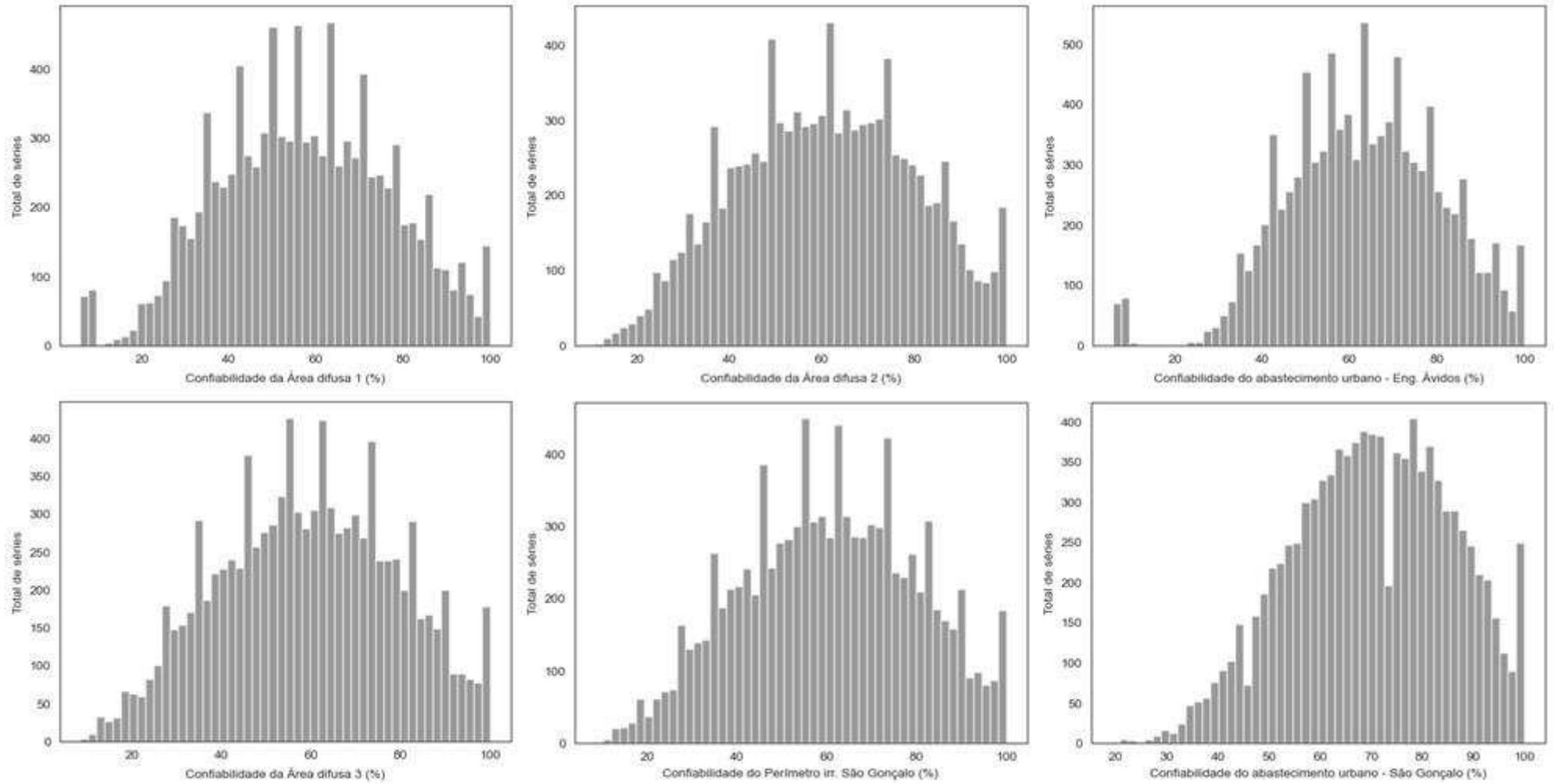
Fonte: Autor (2022)

Figura A8 - Representação de série sintética de vazão para o REA



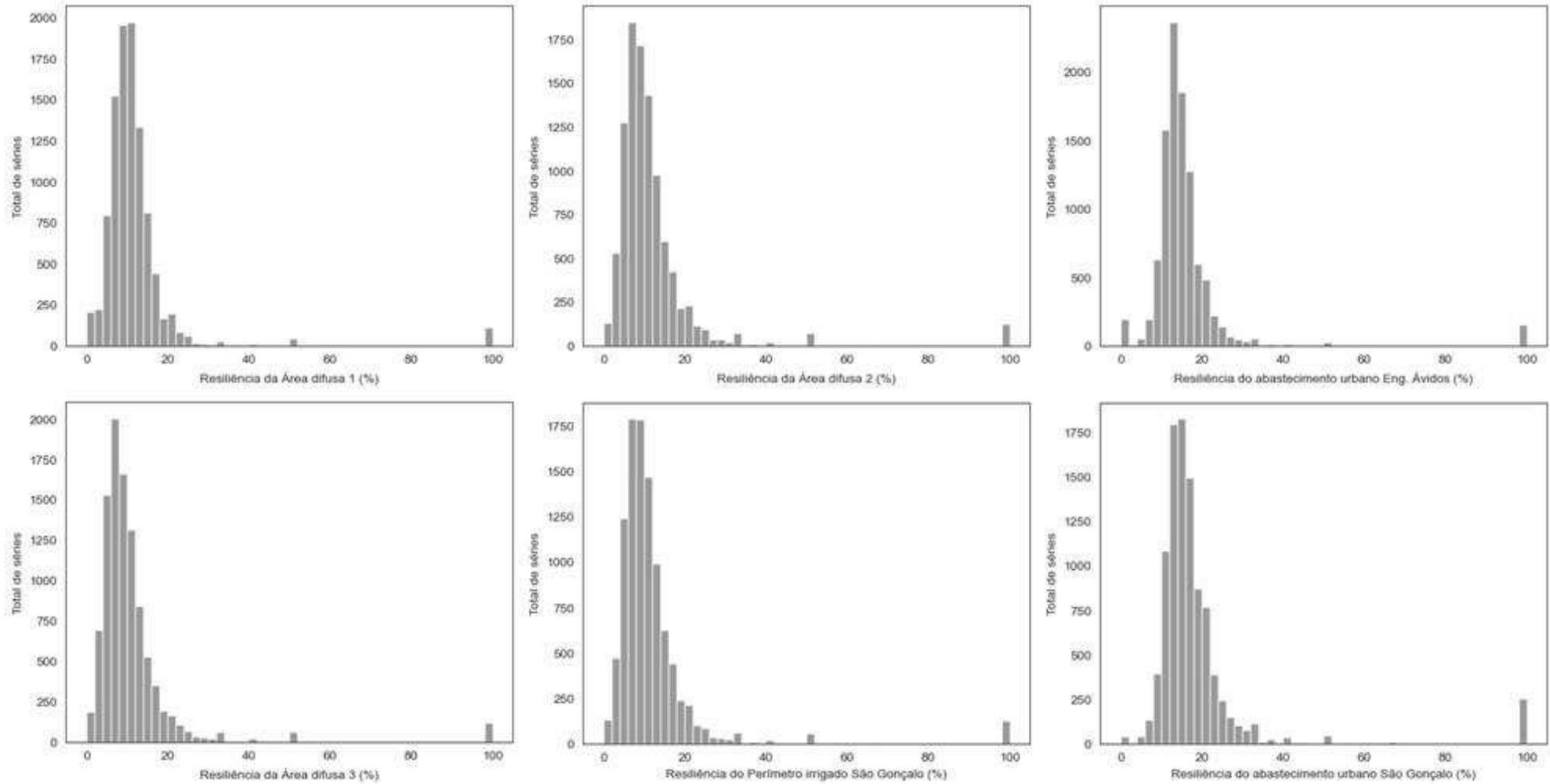
Fonte: Autor (2022)

**Figura A9 - Histogramas de distribuição das confiabilidades das demandas para o CO1**



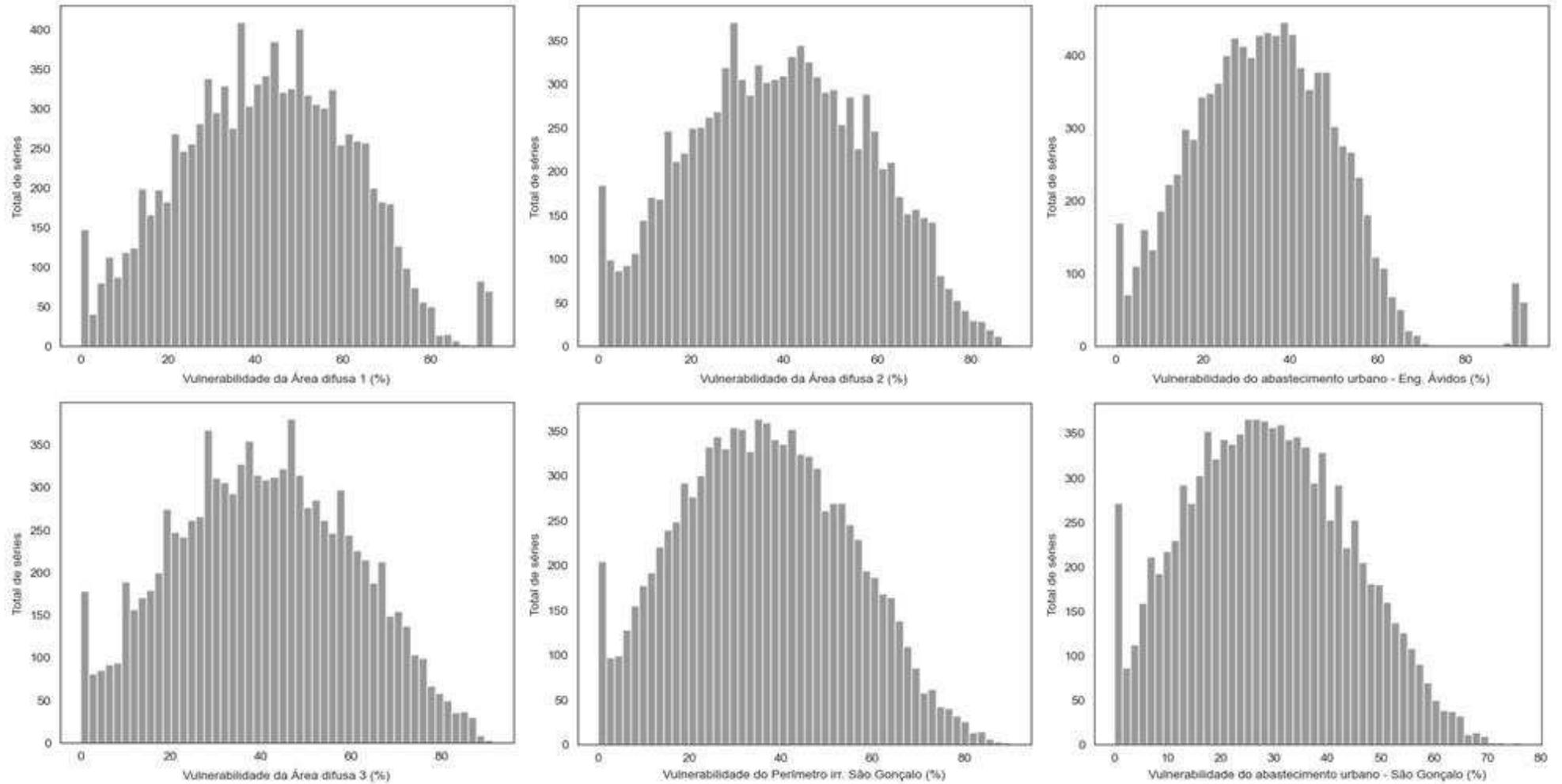
Fonte: Autor (2022)

**Figura A10** - Histogramas de distribuição das resiliências das demandas para o CO1



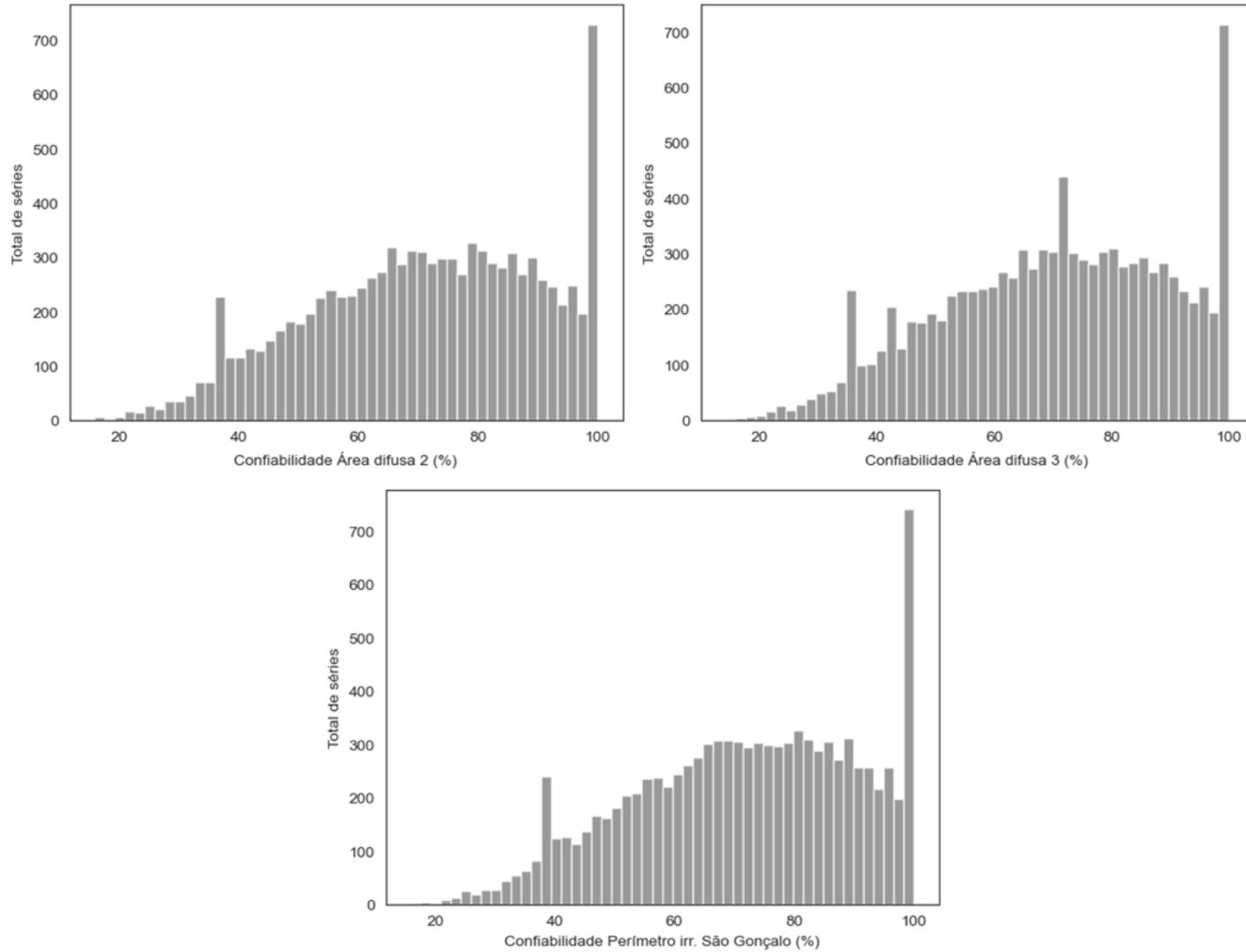
Fonte: Autor (2022)

**Figura A11 - Histogramas de distribuição das vulnerabilidades das demandas para o CO1**



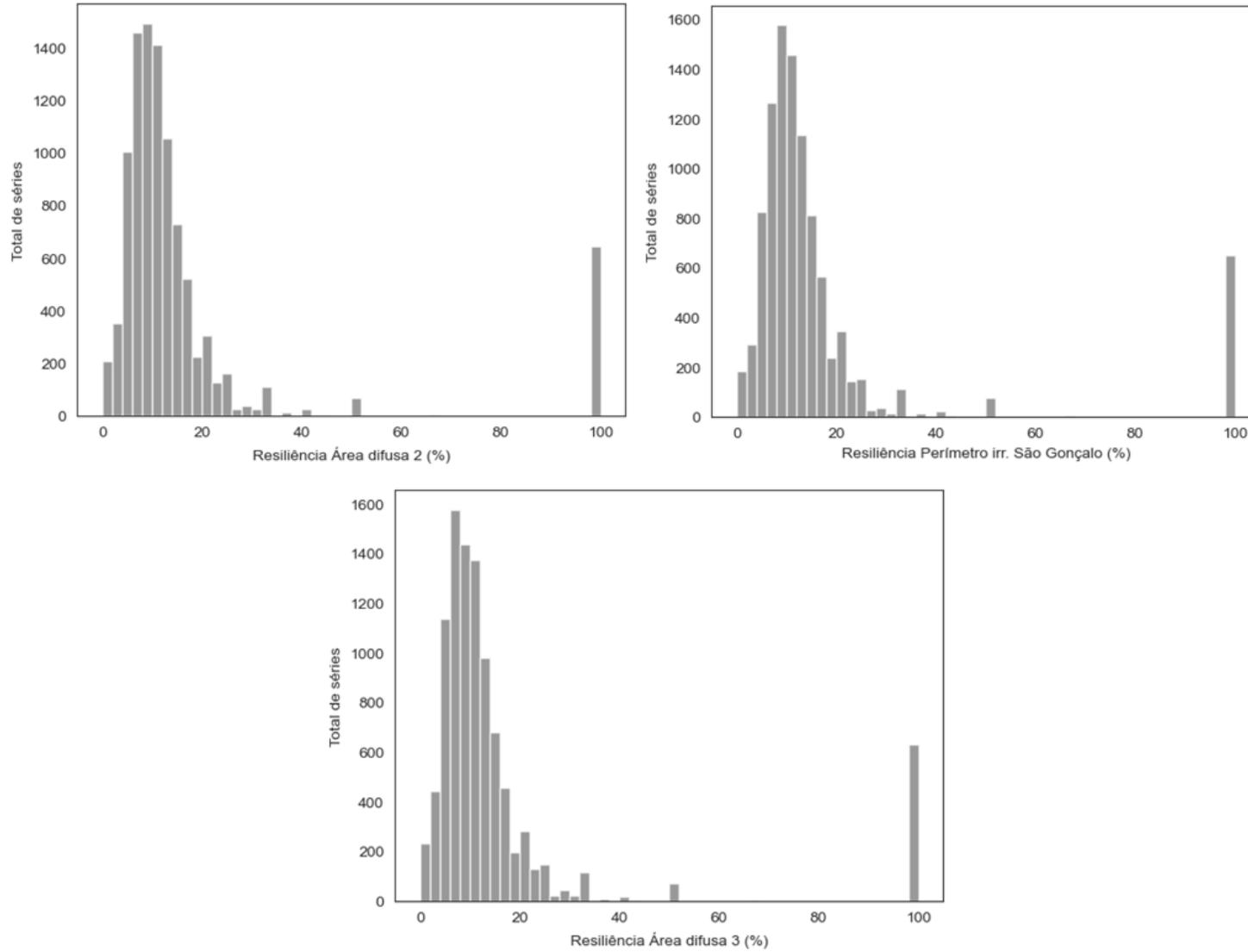
Fonte: Autor (2022)

**Figura A12 - Histogramas de distribuição das confiabilidades das demandas para o CO2**



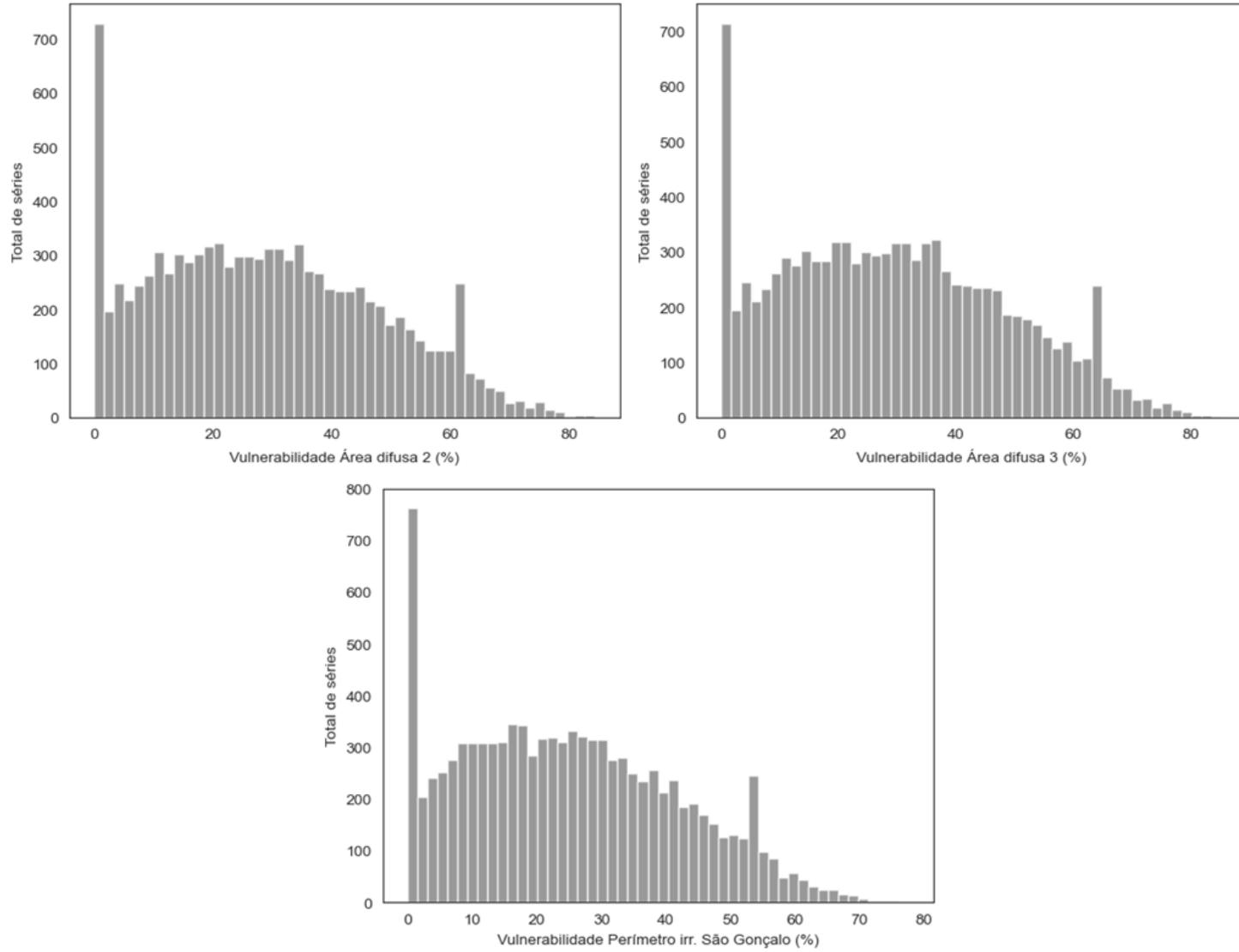
Fonte: Autor (2022)

**Figura A13** - Histogramas de distribuição das resiliências das demandas para o CO2



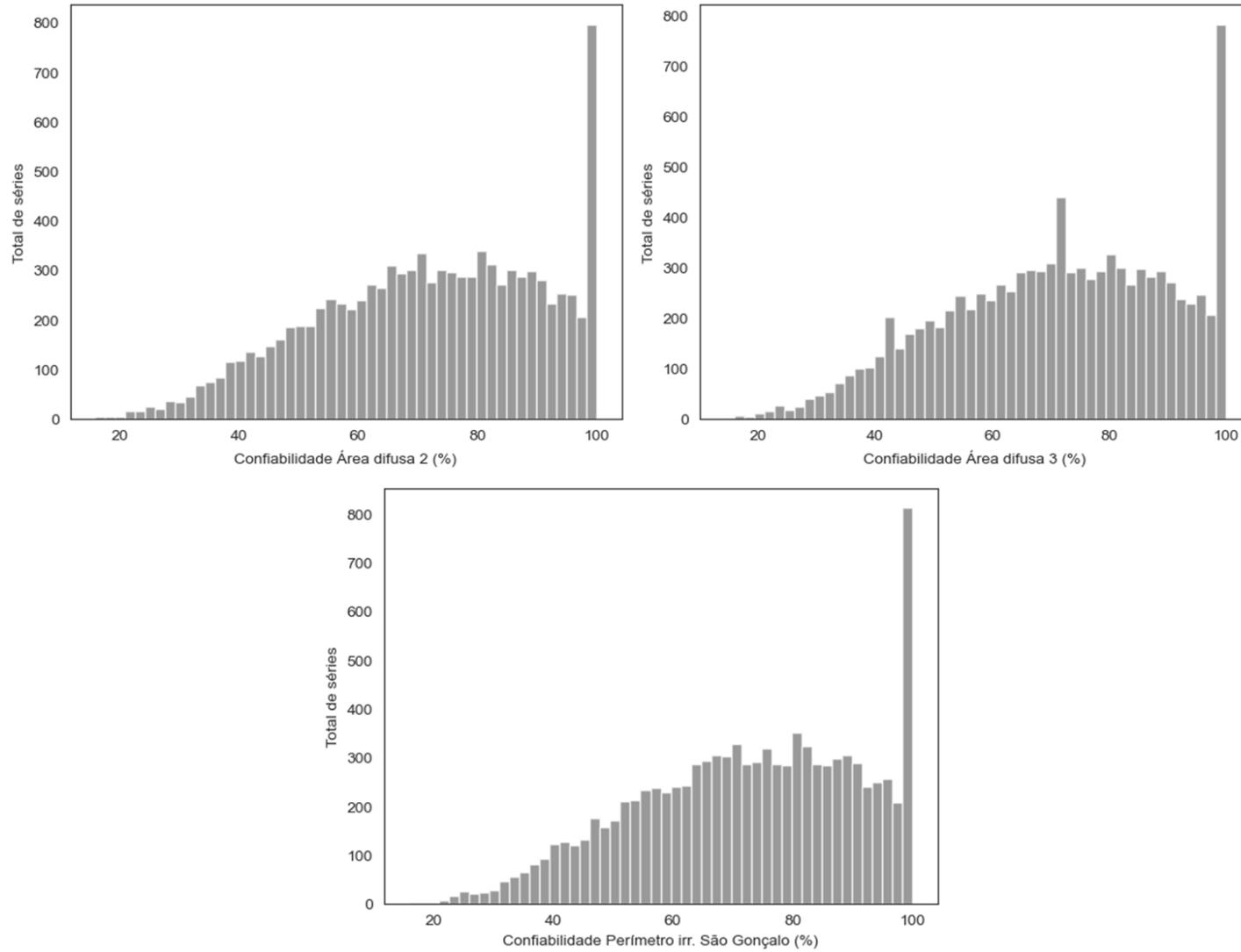
Fonte: Autor (2022)

**Figura A14 - Histogramas das vulnerabilidades das demandas rurais do CO2**



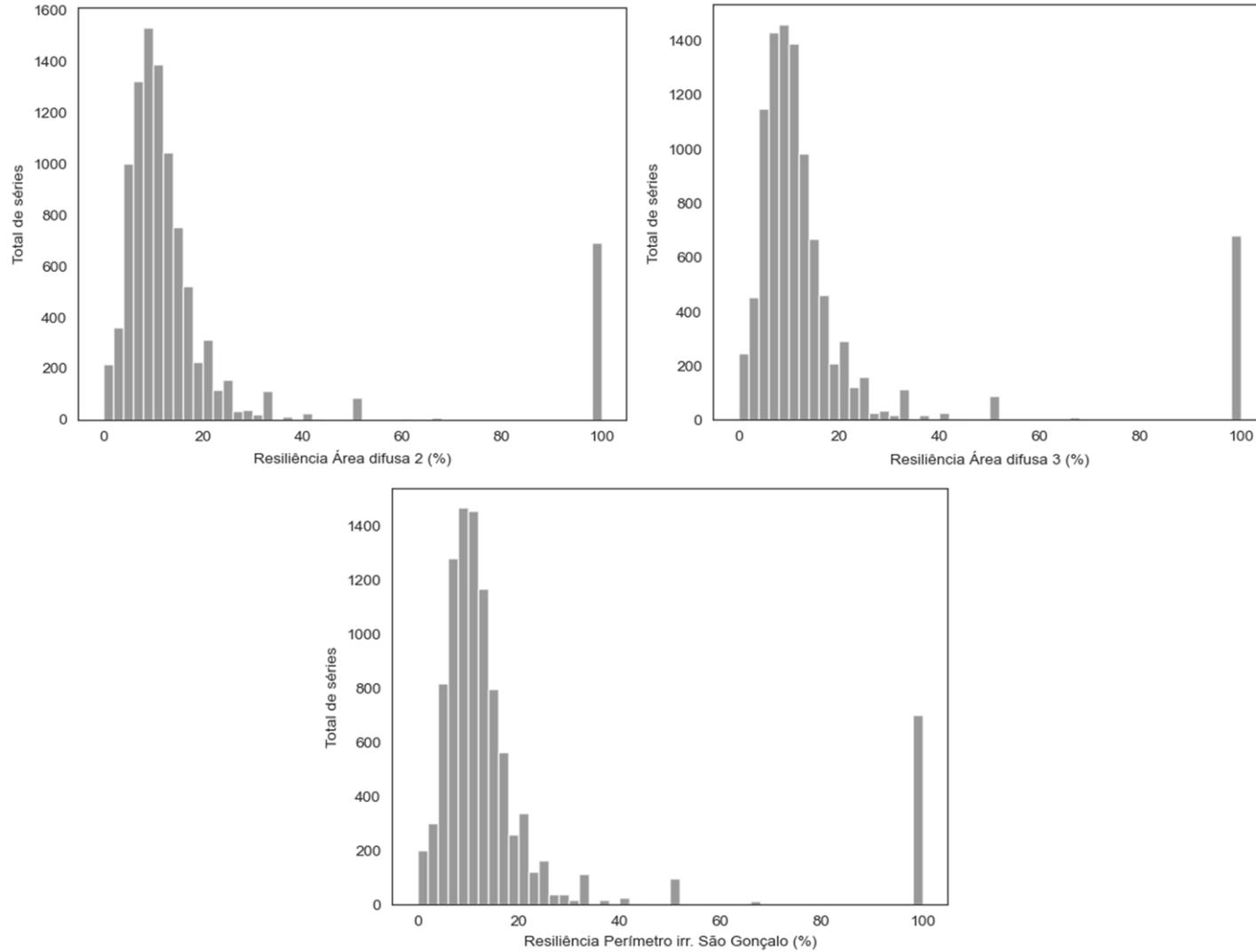
Fonte: Autor (2022)

**Figura A15 - Histogramas de distribuição das confiabilidades das demandas para o CO3**



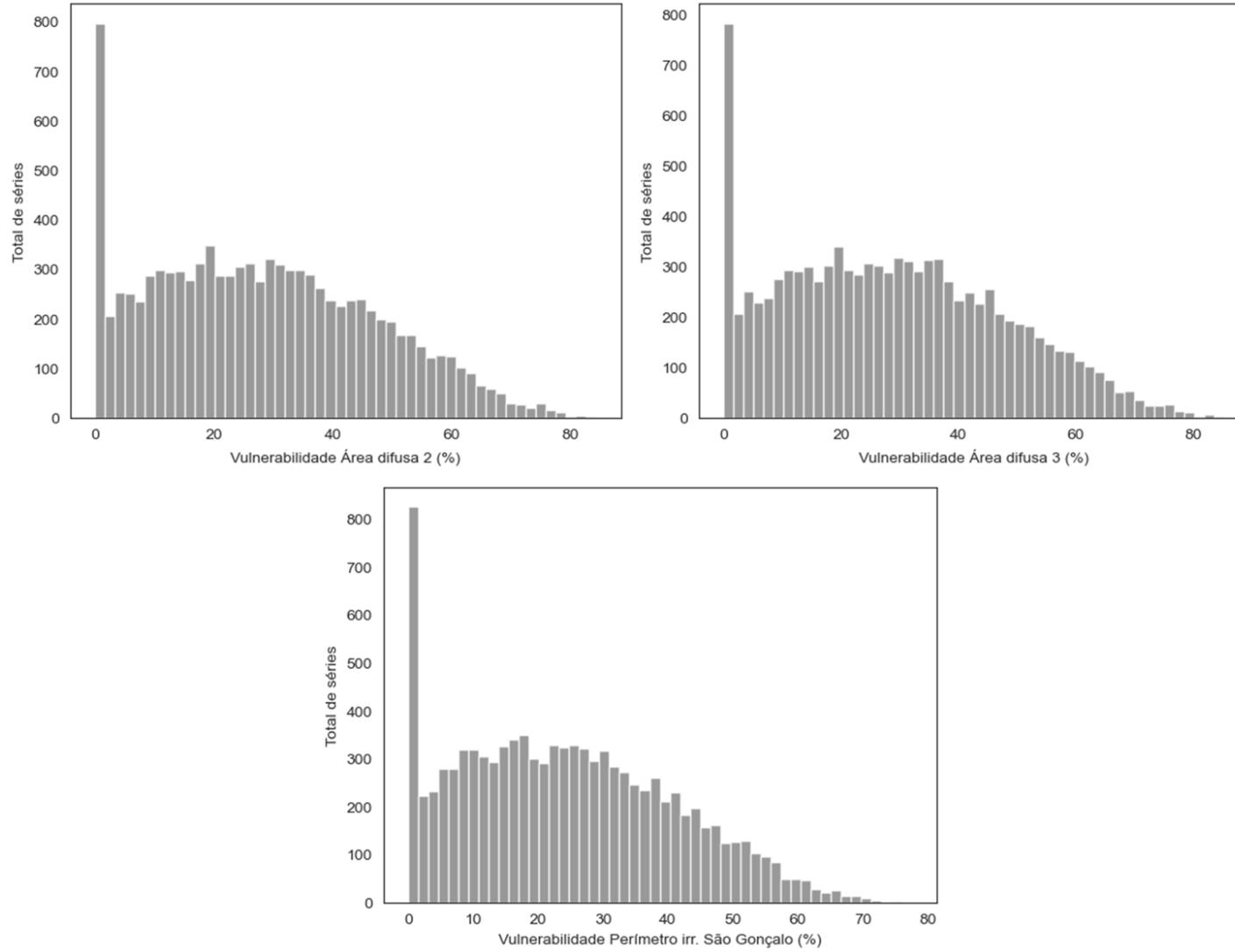
Fonte: Autor (2022)

**Figura A16 - Histogramas de distribuição das resiliências das demandas para o CO3**



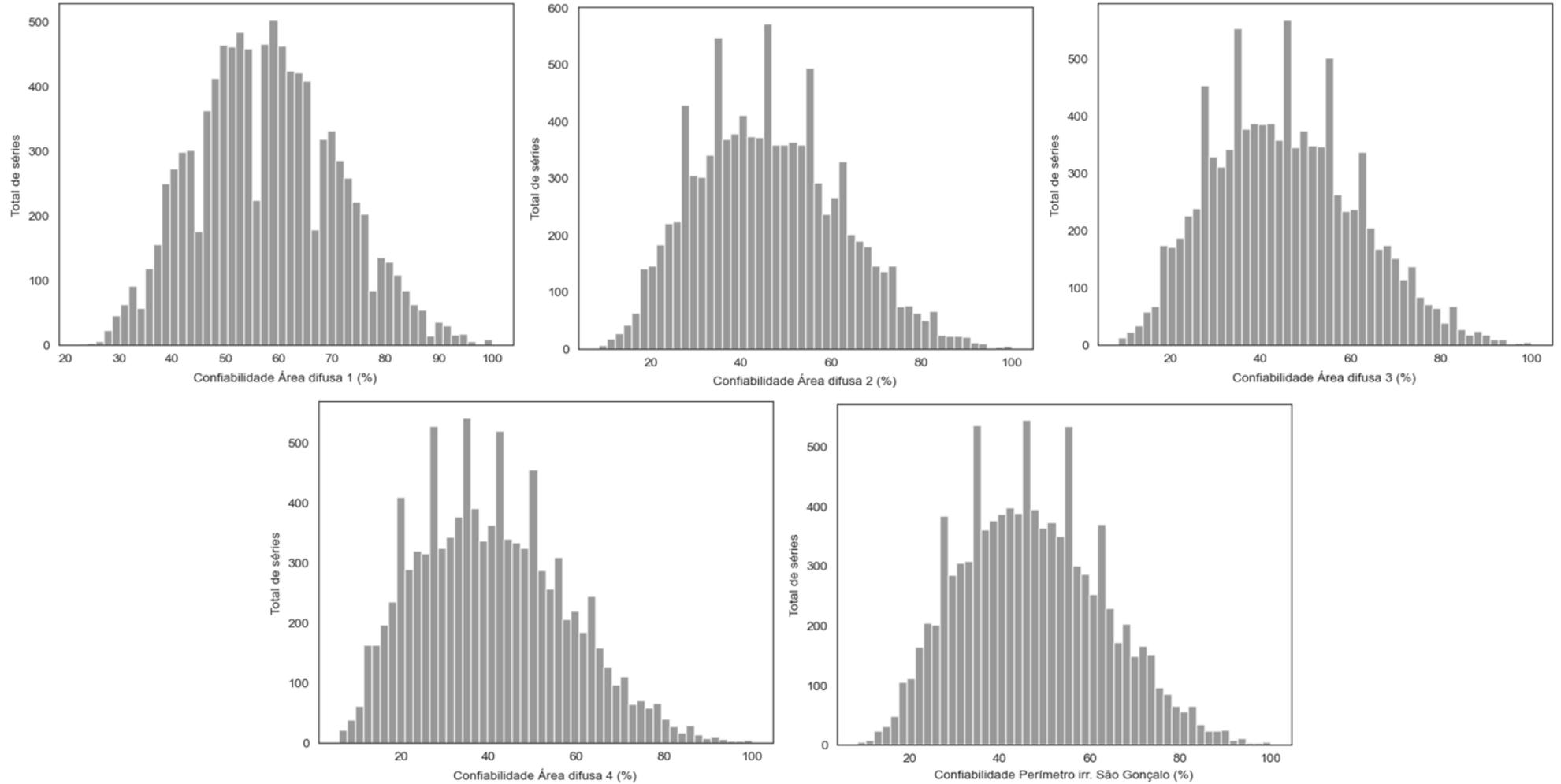
Fonte: Autor (2022)

**Figura A14 - Histogramas de distribuição das vulnerabilidades das demandas para o CO3**



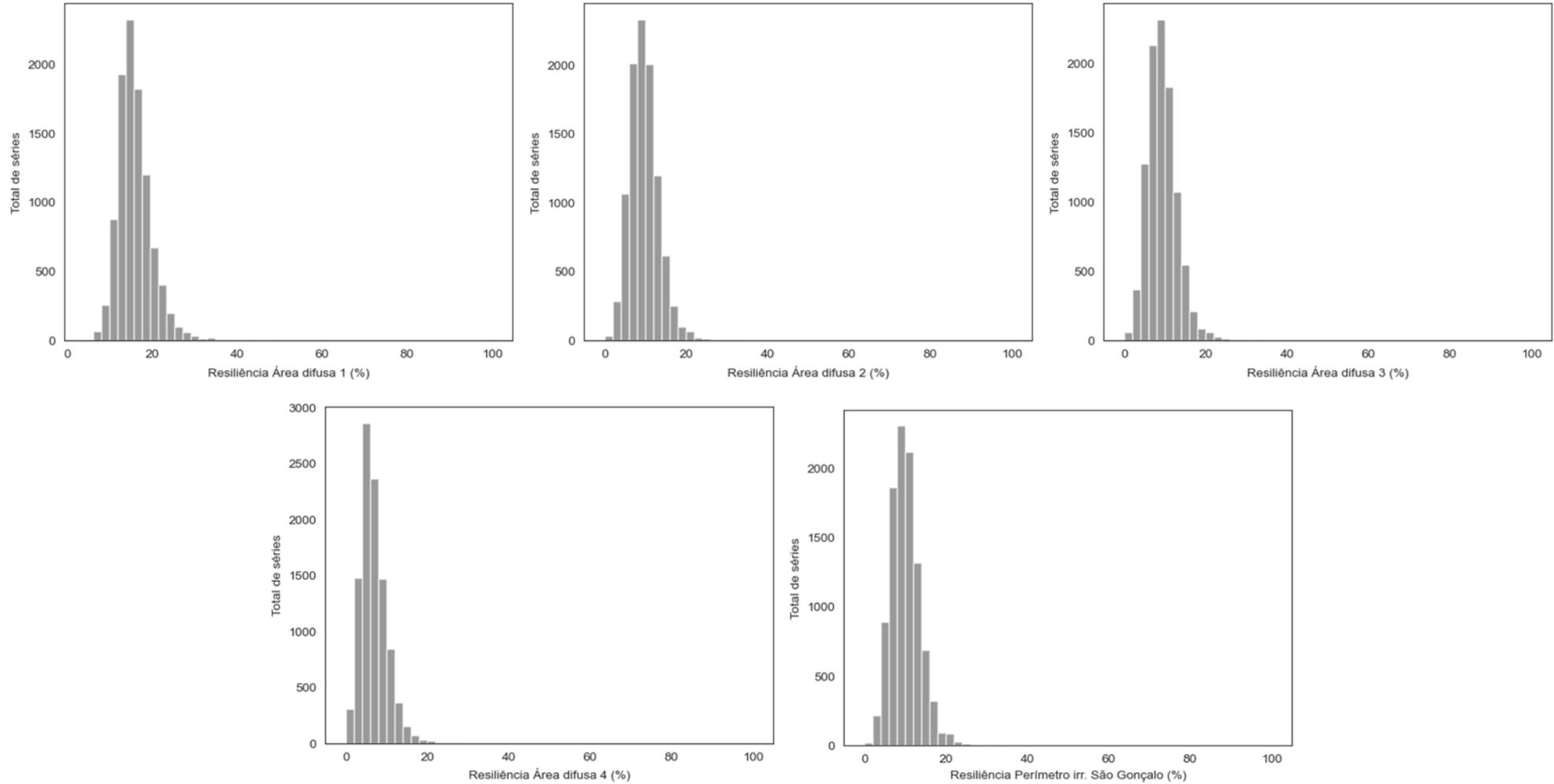
Fonte: Autor (2022)

**Figura A16 - Histogramas de distribuição das confiabilidades das demandas para o CO4**



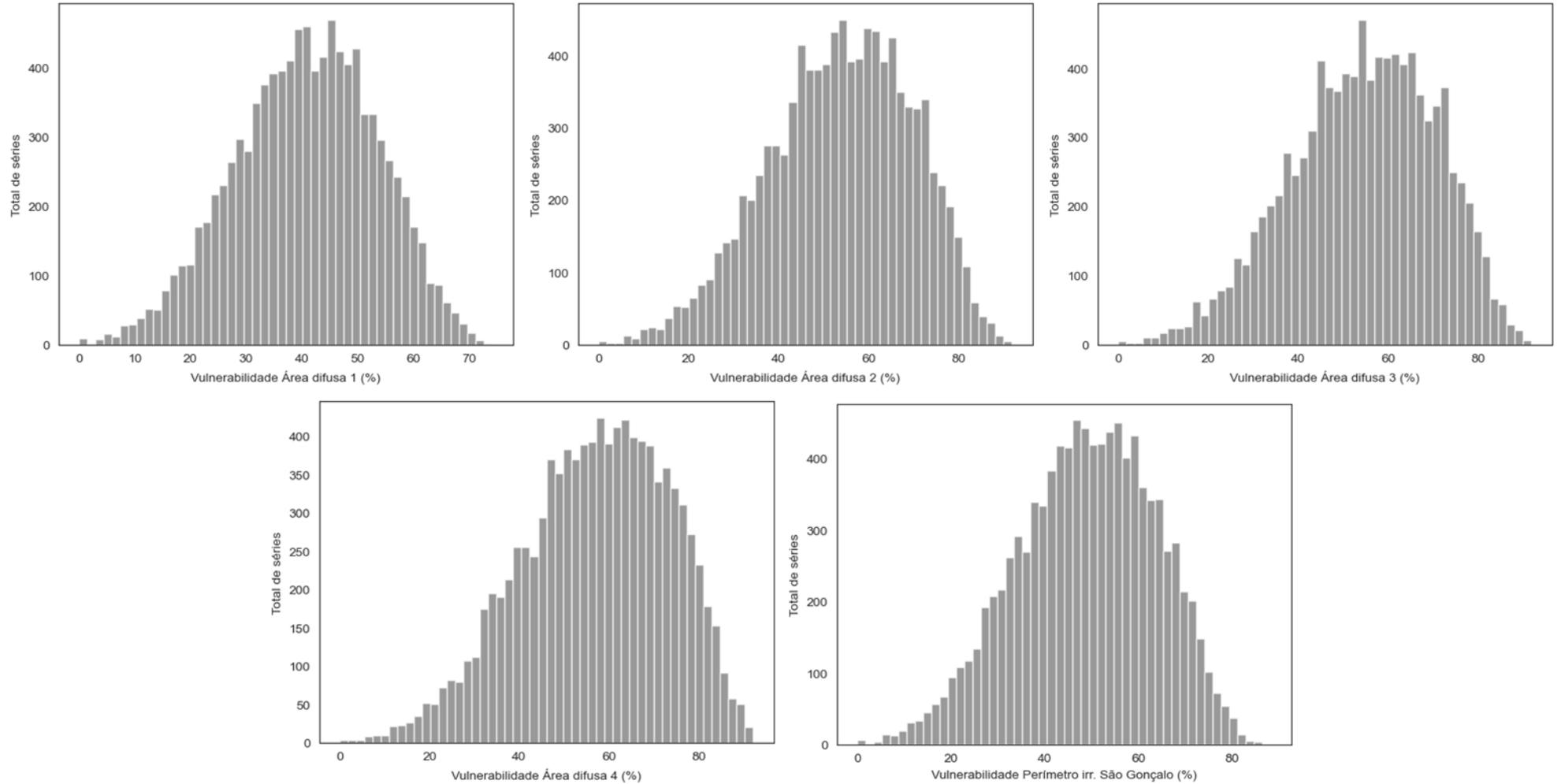
Fonte: Autor (2022).

**Figura A17 - Histogramas de distribuição das resiliências das demandas para o CO4**



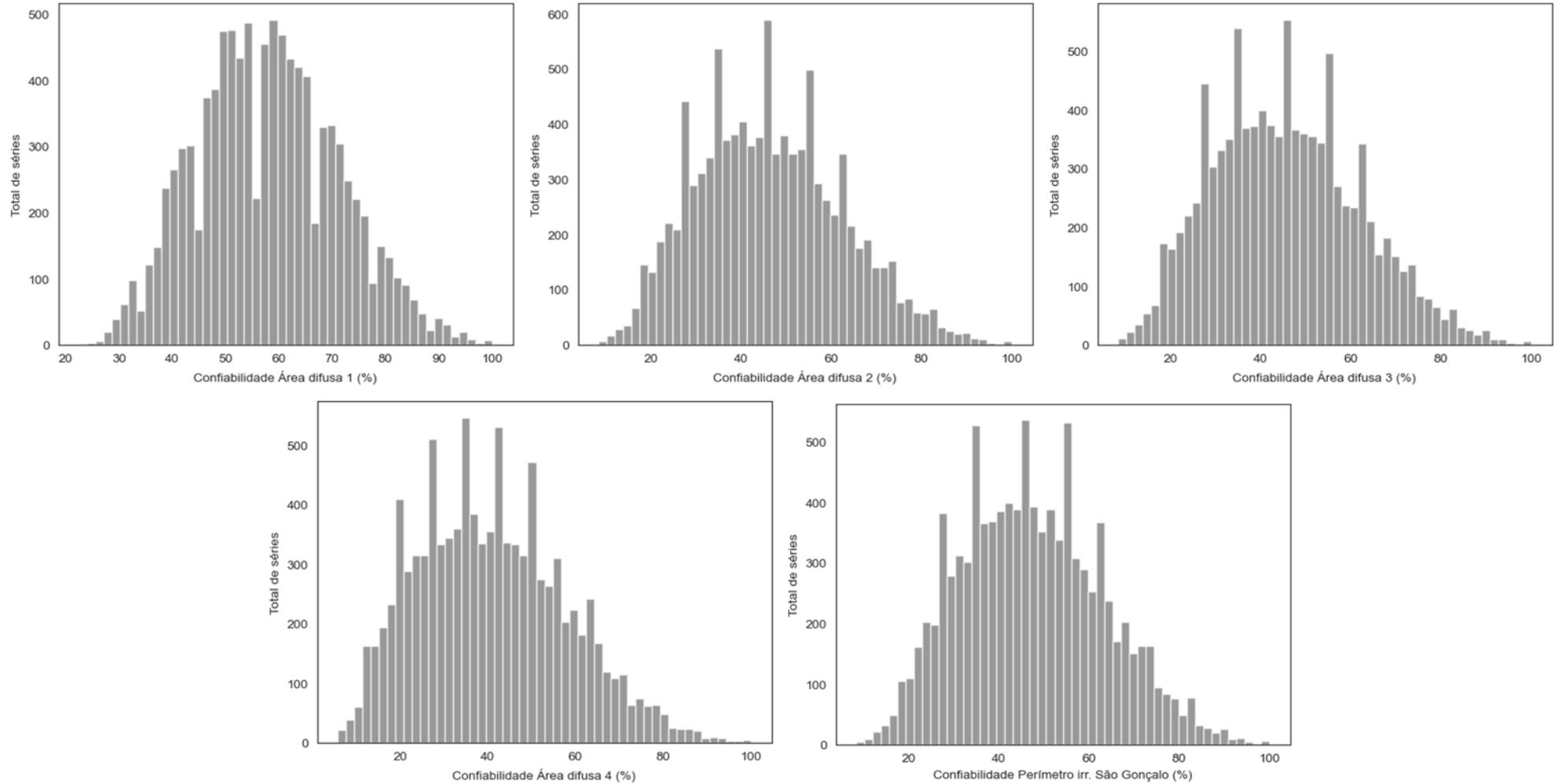
Fonte: Autor (2022).

**Figura A18** - Histogramas de distribuição das vulnerabilidades das demandas para o CO4



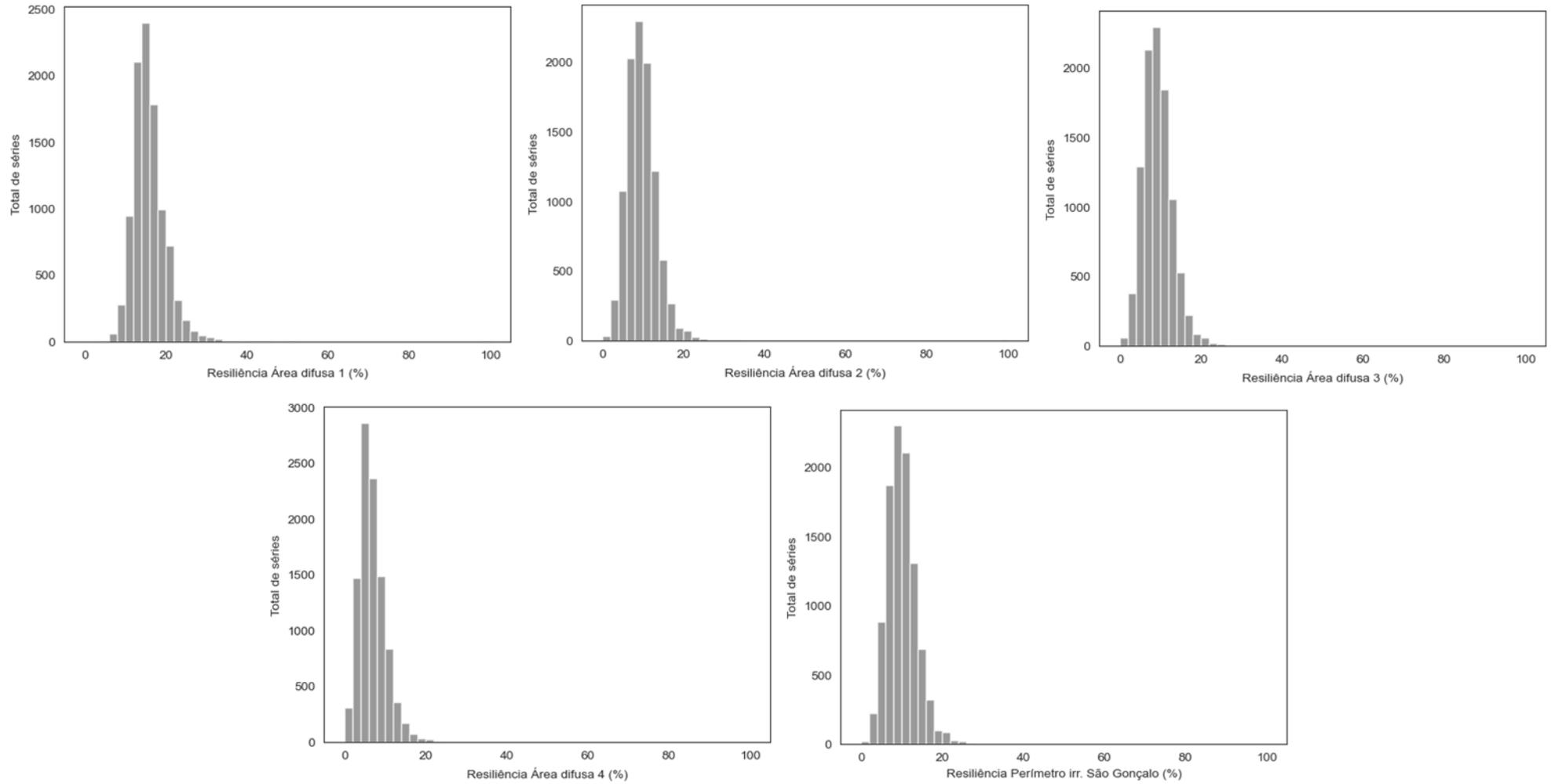
Fonte: Autor (2022).

**Figura A19 - Histogramas de distribuição das confiabilidades das demandas para o CO5**



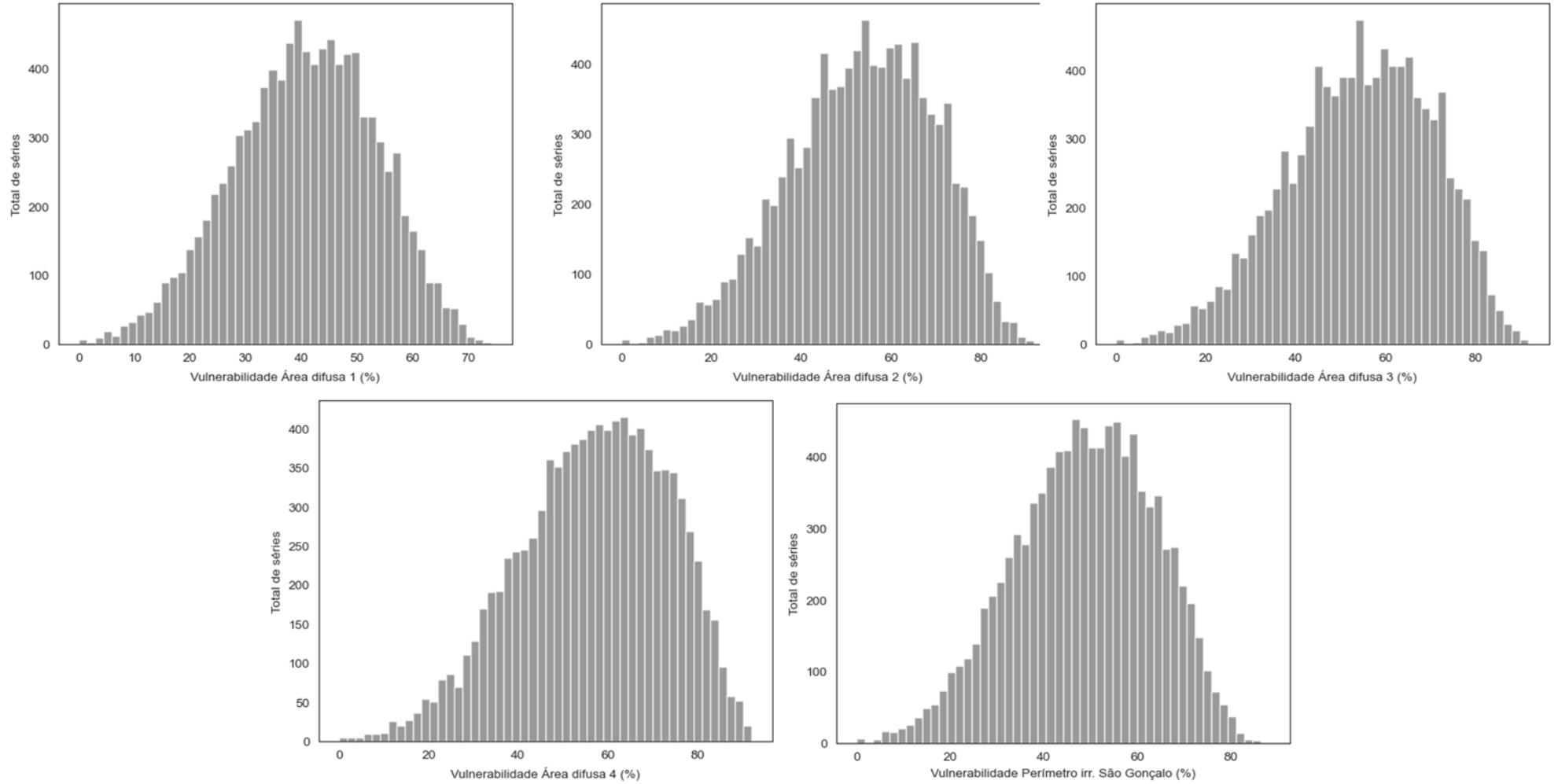
Fonte: Autor (2022).

**Figura A20 - Histogramas de distribuição das resiliências das demandas para o CO5**



Fonte: Autor (2022).

**Figura A21 - Histogramas de distribuição das vulnerabilidades das demandas para o CO5**



Fonte: Autor (2022).