



UNIVERSIDADE FEDERAL DO PARANÁ

ALANA LOUISE WERNECK LASSEN

MODELO DE PREVISÃO DE POLUIÇÃO DIFUSA POR TÉCNICAS DE
INTELIGÊNCIA ARTIFICIAL – ESTUDO DE CASO: BACIA DO RIO BARIGUI

CURITIBA

2021

ALANA LOUISE WERNECK LASSEN

MODELO DE PREVISÃO DE POLUIÇÃO DIFUSA POR TÉCNICAS DE
INTELIGÊNCIA ARTIFICIAL – ESTUDO DE CASO: BACIA DO RIO BARIGUI

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Engenharia de Recursos Hídricos e Ambiental do Programa de Pós-Graduação em Engenharia de Recursos Hídricos e Ambiental, Setor de Tecnologia da Universidade Federal do Paraná.

Orientador: Regina Tiemy Kishi, Dr-Ing.

Coorientador: Gabriel Henrique de Almeida Pereira, Dr.

CURITIBA

2021

L347m

Lassen, Alana Louise Werneck

Modelo de previsão de poluição difusa por técnicas de inteligência artificial – estudo de caso: bacia do Rio Barigui [recurso eletrônico] / Alana Louise Werneck Lassen - Curitiba, 2021.

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

Orientadora: Profa. Dra. Regina Tiemy Kishi

Coorientador: Prof. Dr. Gabriel Henrique de Almeida Pereira

1. Inteligência artificial. 2. Bacia hidrográfica - Paraná. 3. Água - Poluição. I. Universidade Federal do Paraná. II. Kishi, Regina Tiemy. III. Pereira, Gabriel Henrique de Almeida. IV. Título.

CDD 551.483



TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ENGENHARIA DE RECURSOS HÍDRICOS E AMBIENTAL da Universidade Federal do Paraná foram convocados para realizar a arguição da dissertação de Mestrado de **ALANA LOUISE WERNECK LASSEN** intitulada: **MODELO DE PREVISÃO DE POLUIÇÃO DIFUSA POR TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL - ESTUDO DE CASO: BACIA DO RIO BARIGUI**, sob orientação da Profa. Dra. REGINA TIEMY KISHI, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

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

CURITIBA, 30 de Julho de 2021.

Assinatura Eletrônica

17/08/2021 21:59:18.0

REGINA TIEMY KISHI

Presidente da Banca Examinadora

Assinatura Eletrônica

18/08/2021 08:56:23.0

DANIEL HENRIQUE MARCO DETZEL

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

16/09/2021 14:51:50.0

JOAQUIN IGNACIO BONNECARRÈRE GARCIA

Avaliador Externo (UNIVERSIDADE DE SÃO PAULO)

Assinatura Eletrônica

18/08/2021 06:58:31.0

CRISTOVÃO VICENTE SCAPULATEMPO FERNANDES

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

18/08/2021 09:35:18.0

MARCELO LUIZ DE SOUZA

Avaliador Externo (AGÊNCIA NACIONAL DE ÁGUAS)

Dedico esta dissertação ao meu marido que com paciência me apoiou, com fé me levantou e com amor me fez voar.

AGRADECIMENTOS

Às instituições que me deram apoio, ANA e UFPR, por acreditarem na pesquisa. As vidas das pessoas mudam por conta de vocês.

À Marcelle Vargas e equipe da UFPEL, que de bom grado me ajudaram na pesquisa e me deram suporte para utilizar a ferramenta isochrones.

À Caroline Kozak, que cedeu os resultados de sua pesquisa e em quem vejo uma grande pessoa e pesquisadora.

À minha família, por ser o suporte. Obrigada por serem o exemplo daquilo que quero oferecer aos outros, por terem me ensinado a acreditar na verdade e por estarem sempre presente.

À minha psicóloga Michele, que me fez enxergar no espelho o meu lado competente.

À minha equipe de trabalho no projeto Paranapanema, professora Regina, Gabriel, Eileen e Thiago, pela paciência de me verem crescer e por escutarem minhas ideias.

À minha orientadora, professora Regina Kishi por acreditar nas minhas ideias e ter trazido sua bondade à minha vida.

Ao meu coorientador, Gabriel Pereira, por acreditar no meu trabalho, estar sempre disponível e dividir comigo seu conhecimento.

Aos meus primeiros chefes na SANEPAR, Ely, Nicolás e Luciano, que me fizeram apaixonar pelos recursos hídricos e não mediram esforços para me ensinar.

Ao meu para sempre mentor, Cristóvão, por ser sempre mais. Mais que professor, mais que amigo, (muito) mais que média, mais que comum. Obrigada por confiar em mim e me dar apoio.

Ao meu marido, que é para sempre minha melhor versão no mundo.

À Deus, por me dar a fé, por responder todas as minhas orações e por quem todas as coisas são feitas. Pelo amor.

RESUMO

O enquadramento de corpos de água, sendo um importante instrumento da gestão dos recursos hídricos, é baseado no conhecimento da dinâmica da qualidade da água no tempo. São de grande suporte as estimativas de cargas poluidoras dos corpos hídricos e a identificação destas por via de emissão e fonte de poluição para priorizar as ações de recuperação e conservação da qualidade da água diretamente nas fontes. Assim, subsidiar a decisões dos gestores dos recursos hídricos da bacia é essencial a estimativa de cargas pontuais e difusas. Sendo as cargas emitidas por vias de poluição difusa de bastante relevância na bacia e entendendo que a emissão destas ocorre principalmente durante os eventos de precipitação, é necessário o entendimento do comportamento temporal destas emissões. Desta maneira, esta pesquisa busca estudar as cargas poluidoras para fins de gestão e a discretização temporal de cargas na bacia do rio Barigui, no estado do Paraná. Com base em dados monitorados de eventos, estuda-se a relação entre as concentrações de fósforo total (PT), nitrogênio total (NT) e carbono orgânico dissolvido (COD) com dados da bacia e da precipitação, de forma a estimar cargas difusas com dados de fácil acesso, utilizando-se da biblioteca LazyPredict para busca da correlação por 42 diferentes regressores, programados em linguagem Python. São testadas 69 variáveis em um modelo treinado e verificado por 191 amostras coletadas por amostrador automático em eventos de chuva em dois pontos distintos no rio Barigui. Os resultados mostram que cada parâmetro de qualidade da água performa um melhor coeficiente de determinação (R^2) com um regressor específico, sendo Bagging para PT com melhor performance de $R^2 = 0,935$, Extra Tree para NT com $R^2 = 0,967$ e ambos Decision Tree e Extra Tree para COD, com $R^2 = 0,689$. Os melhores resultados obtidos não utilizaram todas as variáveis disponíveis, sendo o modelo apresentado capaz de melhorar a performance dos regressores a partir da escolha de seus dados de entrada.

Palavras-chave: Poluição difusa; Inteligência Artificial; Gestão de Recursos Hídricos

ABSTRACT

The classification of water bodies as an important instrument for the water resources management is based on the dynamics of water quality over time knowledge. Pollution loads estimative of water bodies and emission and pollution source identification are a great support, to prioritize actions for the recovery and conservation of water quality directly at the source. Thus, supporting the decisions of the water resources managers in the basin is essential to estimate punctual and diffuse pollution loads. As the loads emitter by diffuse pathways are very relevant in the basing and understanding that their mission occurs mainly during precipitation events, it is necessary to understand the temporal behavior of these emissions. Therefore, this research aimed to study pollutant loads for management purposes and the temporal discretization of loads in the Barigui River Basin, in Paraná state. Based on monitored event data, the relationship between total phosphorus (TP), total nitrogen (TN) and dissolved organic carbon (DOC) concentrations was studied with basin characteristics and precipitation data, to estimate diffuse loads with easily accessible data, using the LazyPredict library to search the correlation with 42 different regressors, programmed in Python language. 69 variables were tested in a trained and verified model with 191 samples collected by automatic sampler during rainfall events at 2 different points in the Barigui River. The result showed that each water quality parameter performs a better coefficient of determination (R^2) with a specific regressor, being Bagging for TP, with $R^2 = 0.935$ in the best performance, Extra Tree for TN with $R^2 = 0.967$ and both Decision Tree and Extra Tree for COD, with $R^2 = 0.689$. The best results obtained did not use all available variables, and the presented model was able to improve the performance of the regressors from the choice of their input data.

Key-words: Diffuse pollution; Artificial Intelligence; Water Resources Management.

LISTA DE FIGURAS

FIGURA 1 - CICLO HIDROLÓGICO	10
FIGURA 2 - CURVA DE DISTRIBUIÇÕES TEMPORAIS DE PRECIPITAÇÃO	13
FIGURA 3 - CURVAS DE CAPACIDADES E TAXA DE INFILTRAÇÃO	14
FIGURA 4 - HIDROGRAMA TIPO.....	15
FIGURA 5 - MODELO SIMPLIFICADO DO BALANÇO HÍDRICO	18
FIGURA 6 - RELAÇÃO ENTRE OS INSTRUMENTOS DE GESTÃO HÍDRICA	25
FIGURA 7 - MODELOS MATEMÁTICOS	29
FIGURA 8 - BACIA DO RIO BARIGUI	46
FIGURA 9 - OUTORGAS NA ÁREA DE ESTUDO.....	47
FIGURA 10 - EVENTOS DE PRECIPITAÇÃO UTILIZADOS NESTE ESTUDO COM AMOSTRAS COLETADAS POR KOZAK (2016, 2020) - EVENTOS 2 E 3.....	50
FIGURA 11 - EVENTOS DE PRECIPITAÇÃO UTILIZADOS NESTE ESTUDO COM AMOSTRAS COLETADAS POR KOZAK (2016, 2020) - EVENTOS 4, 5 E 6.....	50
FIGURA 12 - EVENTOS DE PRECIPITAÇÃO UTILIZADOS NESTE ESTUDO COM AMOSTRAS COLETADAS POR KOZAK (2016, 2020) - EVENTOS 7, 8 E 9.....	51
FIGURA 13 - EVENTOS DE PRECIPITAÇÃO UTILIZADOS NESTE ESTUDO COM AMOSTRAS COLETADAS POR KOZAK (2016, 2020) - EVENTOS 10, 11, 12, 13 E 14	51
FIGURA 14 - ESTAÇÕES PLUVIOMÉTRICAS, POLÍGONOS DE THIESSEN E ISÓCRONAS NAS BACIAS EM ESTUDO	57
FIGURA 15 - CONCENTRAÇÕES REAL E MODELADA DE FÓSFORO TOTAL PELO MELHOR RESULTADO (SIMULAÇÃO 4500, $R^2 = ,935$, RSME = 0,309)	71
FIGURA 16 - CONCENTRAÇÃO REAL X PREVISTA EM CADA AMOSTRA.....	71
FIGURA 17 - ERRO RELATIVO DAS AMOSTRAS	72
FIGURA 18 - ERRO ABSOLUTO DAS AMOSTRAS	72
FIGURA 19 - MATRIZ DE CORRELAÇÃO ENTRE VARIÁVEIS DE ENTRADA	82
FIGURA 20 - CONCENTRAÇÕES REAL E MODELADA DE FÓSFORO TOTAL PELO MELHOR RESULTADO COM MENOS VARIÁVEIS	99

LISTA DE TABELAS

TABELA 1 - MODELOS REGRESSIVOS DA BIBLIOTECA LAZY PREDICT	43
TABELA 2 - OUTORGAS NAS BACIAS DE ESTUDO (FONTES PONTUAIS DE POLUIÇÃO).....	48
TABELA 3 - TESTES REALIZADOS	60
TABELA 4 - MELHORES RESULTADOS PELO INDICADOR R ² DAS SIMULAÇÕES PARA OS PARÂMETROS FÓSFORO TOTAL, NITROGÊNIO TOTAL E CARBONO ORGÂNICO DISSOLVIDO	64
TABELA 5 - MELHORES 10 RESULTADOS ENTRE TODOS OS TESTES (EXCETO T1, T2, T3 E T4) PARA TODOS OS PARÂMETROS	68
TABELA 6 - TESTES SIMULADOS A PARTIR DA MELHOR RESPOSTA (SIMULAÇÃO 4500, R ² = 0,935, RSME = 0,309) DOS TESTES INICIAIS DA TABELA 3.....	70
TABELA 7 - VARIÁVEIS UTILIZADAS NAS MELHORES SIMULAÇÕES PELO INDICADOR R ² PARA FÓSFORO TOTAL.....	74
TABELA 8 - VARIÁVEIS UTILIZADAS NAS MELHORES SIMULAÇÕES PELO INDICADOR R ² PARA NITROGÊNIO TOTAL.....	76
TABELA 9 - VARIÁVEIS UTILIZADAS NAS MELHORES SIMULAÇÕES PELO INDICADOR R ² PARA CARBONO ORGÂNICO DISSOLVIDO	78
TABELA 10 - QUANTIDADE DE VEZES QUE CADA COLUNA FOI UTILIZADA PARA GERAR AS 11 MELHORES SIMULAÇÕES PARA FÓSFORO TOTAL PELO INDICADOR R ²	91
TABELA 11 - QUANTIDADE DE VEZES QUE CADA COLUNA FOI UTILIZADA PARA GERAR AS 11 MELHORES SIMULAÇÕES PARA NITROGÊNIO TOTAL PELO INDICADOR R ²	93
TABELA 12 - QUANTIDADE DE VEZES QUE CADA COLUNA FOI UTILIZADA PARA GERAR AS 11 MELHORES SIMULAÇÕES PARA CARBONO ORGÂNICO DISSOLVIDO PELO INDICADOR R ²	95
TABELA 13 - SIMULAÇÕES REALIZADAS A PARTIR DAS VARIÁVEIS MAIS UTILIZADAS PARA GERAR OS MELHORES 11 RESULTADOS DE FÓSFORO TOTAL PELO INDICADOR R ²	97

SUMÁRIO

1	INTRODUÇÃO	5
1.1	JUSTIFICATIVA	7
1.2	OBJETIVO GERAL.....	8
1.3	OBJETIVOS ESPECÍFICOS	9
2	FUNDAMENTAÇÃO TEÓRICA	10
2.1	A QUANTIDADE DA ÁGUA.....	11
2.1.1	PRECIPITAÇÃO	12
2.1.2	INFILTRAÇÃO E ESCOAMENTO SUBTERRÂNEO.....	13
2.1.3	ESCOAMENTO SUPERFICIAL	15
2.1.4	EVAPOTRANSPIRAÇÃO.....	16
2.1.5	RETIRADA E LANÇAMENTO DE ÁGUA DE FONTES ANTROPOGÊNICAS 16	
2.1.6	MODELOS HIDROLÓGICOS	17
2.2	A QUALIDADE DA ÁGUA	19
2.2.1	POLUIÇÃO PONTUAL E DIFUSA	21
2.2.2	POLUENTES E INDICADORES	22
2.3	CONSIDERAÇÕES SOBRE A GESTÃO DE RECURSOS HÍDRICOS.....	24
2.4	MODELOS MATEMÁTICOS	29
2.4.1	INTELIGÊNCIA ARTIFICIAL	30
3	MATERIAIS E MÉTODOS	45
3.1	ESTUDO DE CASO.....	45
3.2	MÉTODO	51
3.2.1	Levantamento de dados disponíveis de concentração no rio em eventos de chuva 52	
3.2.2	Definição de variáveis referentes à bacia e aos eventos monitorados.....	52
3.2.3	Desenvolvimento de algoritmo para aplicação das técnicas de inteligência artificial 58	
3.2.4	Treinamento e validação dos modelos de inteligência artificial	58
4	RESULTADOS E DISCUSSÕES	63
5	CONCLUSÕES	100
6	REFERÊNCIAS BIBLIOGRÁFICAS	103

6.	APÊNDICE A – Programa em Python de Regressão Linear para Dados de Concentração Durante Eventos de Precipitação.....	111
7.	APÊNDICE B – Variáveis e Siglas	114

1 INTRODUÇÃO

A Agenda 2030 é um documento que visa a liberdade da raça humana da pobreza e da penúria, assim como a cura e proteção do planeta (ONU, 2015). Os Objetivos do Desenvolvimento Sustentável (ODS) integram a Agenda 2030, formando um conjunto de diretrizes mundial que contribui a um plano de ação para pessoas, para o planeta e para a prosperidade. O 6º ODS da Organização das Nações Unidas (ONU) é de “assegurar a disponibilidade e gestão sustentável da água e saneamento para todos”. A água, em quantidade e qualidade, dignifica a vida humana.

Sendo essencial para a vida, a água é protegida pela Constituição Brasileira de 1988 (BRASIL, 1988) e, conforme o inciso XIX do artigo 21, promulgou-se a Lei 9.433 (BRASIL, 1997) instituindo a Política Nacional de Recursos Hídricos e criando o Sistema Nacional de Gerenciamento de Recursos Hídricos. Tal lei objetiva, dentre outros, assegurar a disponibilidade de água, nos padrões de qualidade para uso e a utilização racional e integrada dos recursos hídricos. A fim de garantir tais objetivos, cinco instrumentos foram implementados por esta Lei:

1. Planos de Recursos Hídricos: fundamentam e orientam a aplicação da Política Nacional de Recursos Hídricos a longo prazo por bacia hidrográfica, por Estado e para o País. São compostos pelo diagnóstico da situação atual dos recursos hídricos, análise das alternativas futuras para crescimento demográfico, atividades produtivas e uso e ocupação do solo, balanço entre disponibilidade e demandas futuras dos recursos hídricos e suas possíveis complicações, metas de racionalização de uso, aumento e melhoria de quantidade e qualidade dos recursos disponíveis, medidas, programas e projetos futuros para alcançar metas, priorização das outorgas, diretrizes e critérios para a cobrança dos recursos e propostas de restrição de uso para proteção dos recursos;
2. Enquadramento dos Corpos de Água em Classes, segundo os usos preponderantes da água: instrumento de classes estabelecidas pela legislação ambiental para assegurar a qualidade das águas conforme o uso a elas destinados e diminuir os custos do combate à poluição, mediante ações preventivas permanentes e corretivas;

3. Outorga dos Direitos de Uso de Recursos Hídricos: assegura o controle quantitativo e qualitativo dos usos da água para os usos específicos de derivação/captação de água ou extração de água de aquífero subterrâneo, ambos para consumo final ou insumo de processo produtivo, lançamento de resíduos líquidos ou gasosos, aproveitamento de potencial hidrelétrico ou outros usos que alterem qualidade ou quantidade da água de um corpo d'água. As outorgas devem respeitar o enquadramento dos corpos d'água e preservar os usos múltiplos da água. A outorga simplesmente dá o direito ao uso da água, não sua alienação;
4. Cobrança pelo Uso de Recursos Hídricos: objetiva reconhecer a água como bem econômico (e não financeiro) para dar ao usuário uma indicação de seu valor, incentivar a racionalização desta e obter recursos financeiros para programas e intervenções nos planos de recursos hídricos. Os usos sujeitos a outorga serão cobrados e, os valores arrecadados, aplicados prioritariamente na bacia hidrográfica em que foram gerados. Os valores cobrados devem se basear, para a captação/derivação/extração de água, na quantidade, enquanto para o lançamento de efluentes e rejeitos, na quantidade e na qualidade;
5. Sistema de Informações sobre Recursos Hídricos: é um sistema de coleta, tratamento, armazenamento e recuperação de informações sobre recursos hídricos e fatores importantes à gestão, sendo concentrados no Sistema Nacional de Informações sobre Recursos Hídricos. O Sistema funciona de forma descentralizada quanto à obtenção de dados e informações, mas em um sistema unificado e disponível à toda a sociedade. O Sistema objetiva reunir, consistir e divulgar dados qualitativos e quantitativos dos recursos hídricos, atualizando-os e fornecendo subsídio para a elaboração dos Planos de Recursos Hídricos.

A atuação destes instrumentos de gestão em todo território serve para que seja reestabelecido uma consonância com o meio ambiente, uma vez que as cargas poluentes continuarão a ser geradas e a água sempre será consumida, sendo que ambas as necessidades deverão ser supridas no futuro.

1.1 JUSTIFICATIVA

O enquadramento e as outorgas nos corpos hídricos dependem, diretamente, da qualidade e quantidade de água disponível. Ambos os instrumentos de gestão dos recursos hídricos têm indicadores de qualidade para classificar e permitir os usos da água. Entretanto, a qualidade da água é amplamente discutida quanto à obediência dos padrões de qualidade no tempo.

Especificamente sobre o instrumento do enquadramento de corpos de água, este é regulamentado pela Resolução CONAMA 357/05 (BRASIL, 2005) que classifica os corpos hídricos de acordo com seu uso preponderante, estabelecendo padrões de qualidade para um conjunto de parâmetros. Entretanto, são recorrentes os questionamentos da concentração de poluentes como o melhor indicador de qualidade da água (MACHADO et al. 2019) uma vez que há variação temporal significativa para determinados parâmetros e as amostras de qualidade de água são coletadas usualmente em períodos não-chuvosos, tendendo a evidenciar a poluição das águas geradas por fontes pontuais e subestimar aquelas geradas pelo uso da terra que são carregadas durante os eventos de precipitação.

A fim de entender como a carga difusa se comporta no tempo e eventualmente formular um modelo de estimativa de carga difusa com base nos eventos de precipitação, julga-se necessária a utilização de dados de concentração no corpo d'água durante os eventos de precipitação, uma vez que são durante esses eventos que a parcela de carga é carregada em sua maior parte pela erosão e pelo escoamento superficial. Entretanto, a maioria dos dados de concentração das estações fluviométricas presente na bacia é referente a períodos sem chuva, quando as coletas das amostras de água são mais viáveis. Além disso, a quantidade de amostragem é periodizada, normalmente em coletas anuais, semestrais ou trimestrais para priorizar critérios logísticos do monitoramento em detrimento de critérios estatísticos no uso de estudos de modelagem, podendo não ser o suficiente para refletir a condição real do corpo hídrico para tais estudos.

Na bacia do rio Barigui, dentro da área em estudo, está localizada uma estação de qualidade da água cadastrada no Instituto Água e Terra (IAT), Boichininga – 65019520, em operação desde 1993. Esta estação tem 62 campanhas de dados de qualidade da água trimestrais, disponíveis entre os anos de 1993 e 2017, dentre as quais apenas 19 têm dados de Fósforo Total e 49 relacionados ao Nitrogênio (Nitritos,

Nitratos, Nitrogênio Amoniaco e Nitrogênio Kjeldahl). São 35 os registros da condição do tempo no momento da amostra, sendo apenas 7 deles em dias “chuvosos” (que não necessariamente indicam que a coleta foi realizada durante o evento da chuva, uma vez que a informação da condição do tempo é binária do dia, sendo preenchida como “bom” ou “chuvoso” apenas).

Podem e devem ser feitos questionamentos quanto à representatividade destas informações de qualidade da água quando são observadas pela perspectiva de gestão, como nos instrumentos de enquadramento e outorga, por exemplo, em que a classificação do corpo hídrico não é relacionada a métricas específicas temporais e espaciais.

A modelagem matemática da qualidade da água permite uma melhor compreensão da variação desta ao longo do tempo e do espaço, bem como dos impactos de lançamentos de cargas poluidoras e de possíveis melhoras trazidas pela implementação de medidas de despoluição hídrica (KNAPIK et al., 2014). Entretanto, as informações de saída de qualquer modelo hidrológico dependem de suas variáveis de entrada. É neste sentido que esta pesquisa pretende organizar temporalmente os dados de entrada, para que os resultados de simulações de qualidade da água estejam em escala temporal adequada para sua utilização, considerando a poluição hídrica por fontes difusas. Tais resultados trarão subsídios técnicos para discussões e definições de diferentes limites temporais de qualidade da água tanto para o enquadramento quanto para a outorga do uso da água, que atualmente não consideram o tempo como uma variável para captação, lançamento, uso ou monitoramento das águas, mas que têm potencial para otimizar tais ações e manter a saúde do corpo hídrico. De forma específica, esta pesquisa desenvolve uma metodologia para validar ou não a hipótese de que há uma relação matemática entre a concentração do poluente na água, sobretudo em eventos de precipitação, e as características da bacia e do evento de precipitação.

1.2 OBJETIVO GERAL

Desenvolver método de estimativa de concentração de cargas poluentes difusas no rio, a partir de dados do evento de precipitação e características da bacia.

1.3 OBJETIVOS ESPECÍFICOS

1. Desenvolver banco de dados de fácil acesso de possíveis variáveis que possam influenciar as concentrações de poluentes difusos no rio;
2. Aplicar e avaliar modelos de regressão por inteligência artificial que descrevam o comportamento das concentrações nos rios;
3. Analisar variáveis de maior influência na modelagem de cargas poluente a partir destes modelos de regressão por inteligência artificial;
4. Verificar se a correlação entre as variáveis de entrada pode ser um fator de influência nos modelos de regressão;
5. Verificar os comportamentos nas estimativas de cargas poluentes pelos modelos de regressão entre diferentes pontos de amostragem na bacia hidrográfica do rio Barigui.

2 FUNDAMENTAÇÃO TEÓRICA

A água é fundamental para a manutenção da vida, mas apenas 0,8% da água no planeta é doce e, dentre esse montante, apenas 3% está disponível superficialmente (SPERLING, 2017). Além desta porção de água ser pequena, a distribuição espacial não é homogênea, estando em abundância em alguns locais e em escassez em outras. A qualidade da água deve também ser um fator de preocupação, visando sempre a preservação do recurso para seu uso, uma vez que é limitante para suas finalidades.

Segundo Tucci (2015) e Sperling (2017), a água circula pela Terra carregando diferentes substâncias e estando disponível em diferentes quantidades no tempo. Durante o período de circulação fechada entre a superfície terrestre e a atmosfera, impulsionada principalmente pela energia solar, gravidade e rotação terrestre, chamado ciclo hidrológico, a água se apresenta em diferentes formas, sendo transferida pelos diferentes mecanismos, como na Figura 1, simplificada:



Fonte: USGS, 2019

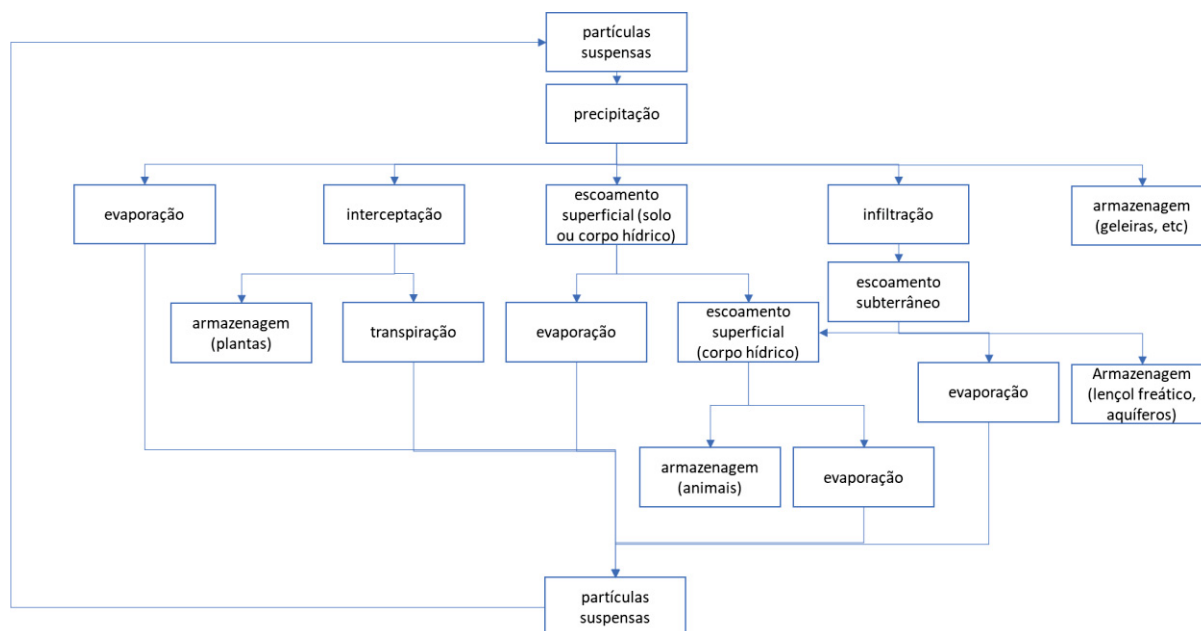
Ainda que o ciclo hidrológico só possa ser completado na escala do planeta todo, o elemento fundamental de análise hidrológica é a bacia hidrográfica, a partir da qual o ciclo hidrológico pode ser mais bem estudado em sua fase terrestre. A bacia hidrográfica é caracterizada como uma área, topologicamente delimitada por pontos de maiores altitudes, em que toda a precipitação que cai sobre ela converge a um mesmo ponto, denominado exutório. Pode ser considerada como um sistema em que o volume de água precipitado é sua entrada e a saída é o volume de água escoado no exutório, considerando-se também as perdas de água por evapotranspiração e profunda infiltração.

O movimento da água a permite realizar uma de suas principais funções: transporte de substâncias. É através da água das chuvas que muitos poluentes são carregados aos rios, contribuindo para a alteração na qualidade da água.

2.1 A QUANTIDADE DA ÁGUA

O volume de água no Brasil representa 12% do total do planeta. Apesar disso, a distribuição deste volume é heterogênea. Cerca de 80% encontram-se na bacia amazônica, sendo o restante distribuído em diferentes quantidades no restante do território (ANA, 2017). A disponibilidade hídrica, sendo a quantidade de água disponível para ser ofertada aos diversos usos, é entendida pelos processos em que a água é processada no espaço e, portanto, a quantidade de água disponível por determinado período de tempo depende da conjuntura de precipitação, condições do solo para infiltração e interceptação, movimento da água sobre e sob o solo, além da evapotranspiração da água, como na Figura 2.

FIGURA 2 – CICLO HIDROLÓGICO POR PROCESSOS



Fonte: A Autora (2021)

2.1.1 PRECIPITAÇÃO

A precipitação é toda quantidade de água que precipita em forma de neblina, granizo, saraiva, orvalho, geada, neve e, mais comumente, chuva. É uma das fases do ciclo hidrológico e tem quatro características principais: total, duração, distribuição temporal e distribuição espacial (TUCCI, 2015).

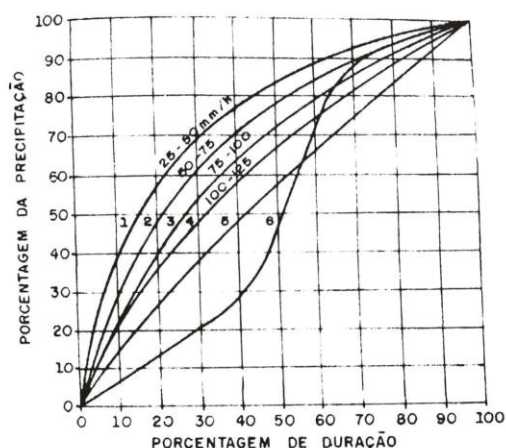
O total precipitado é calculado pela altura pluviométrica, sendo esta entendida como a espessura média da lâmina de água precipitada sobre a bacia admitindo nenhuma infiltração, evapotranspiração ou interceptação no local escoado. É quantificado em milímetro de chuva, definido como o correspondente ao volume de 1 litro por metro quadrado de superfície. A duração é o período de tempo em que a precipitação ocorre. A razão entre estas duas propriedades denomina a intensidade, sendo a precipitação por unidade de tempo, variável durante um evento.

Uma das propriedades da chuva, conforme citado, é a não homogeneidade na distribuição espacial desta, ou seja, a chuva cai em diferentes quantidades e tempo dentro de uma bacia hidrográfica. Quando medida por pluviômetros e pluviógrafos, as informações são relativas ao local em que foram coletadas. Para a geração de uma superfície de precipitação, ou seja, entender como a precipitação ocorreu em uma área, foram desenvolvidos diversos métodos para o cálculo da precipitação média. Os mais usuais são o Método de Thiessen, Método das Isoietas e os métodos

interpoladores, normalmente utilizados em sistema de informação geográfico (SIG), tais quais a Krigagem, o IDW (inverso da distância ponderada), Spline e Trend.

Um evento de precipitação, como citado anteriormente, tem como característica uma distribuição temporal variável. Isso significa que existem períodos durante um evento em que há maior precipitação que outros. A curva de distribuições temporais elaborada por Wilken (1978 citado por TUCCI, 2015) indica que, quanto menor a quantidade total precipitada, maior a porcentagem precipitada em uma menor porcentagem de duração do evento, conforme Figura 2.

FIGURA 2 - CURVA DE DISTRIBUIÇÕES TEMPORAIS DE PRECIPITAÇÃO



CURVAS 1 a 4 : Chuvas de 1 hora de duração
com diferentes intensidades
CURVA 5 : Chuvas de longa duração
(Frentes estacionárias)
CURVA 6 : Chuvas de 6 a 24 horas de
duração

Fonte: TUCCI (2015)

2.1.2 INFILTRAÇÃO E ESCOAMENTO SUBTERRÂNEO

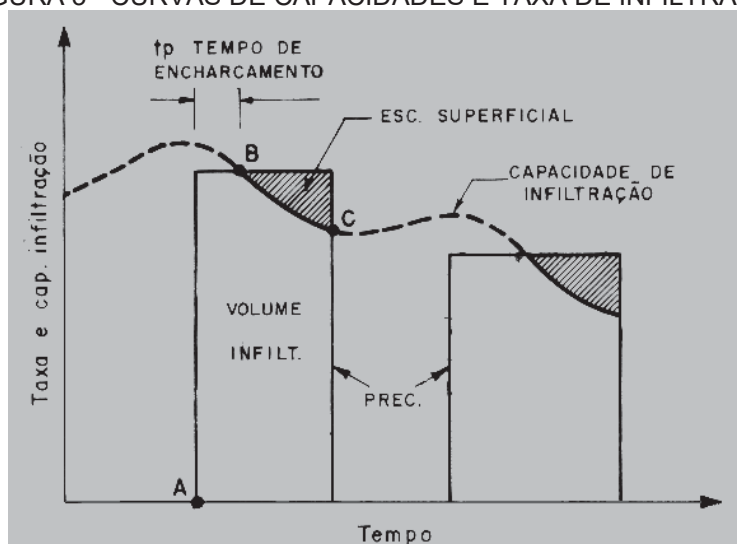
Infiltração é o movimento da água para dentro do solo marcado pela ação da gravidade e pelo potencial capilar. Quando a chuva atinge a superfície do solo, parte dela, ou toda, penetra no solo tanto pela força da gravidade (dirigida para baixo, tanto mais facilmente quanto maior for o diâmetro dos poros) quanto pelo efeito de capilaridade (maior quanto menor forem os poros). A água pode infiltrar até que se atinja a saturação do solo, que ocorre quando todos os poros do solo estão preenchidos com água, mas apenas na velocidade que a capacidade de infiltração permite. Estes dependem da duração da precipitação, da condição de saturação do solo antes do evento de chuva, da cobertura e uso da terra, assim como o tipo do solo.

A capacidade de infiltração é o máximo valor da intensidade da chuva que o solo pode absorver sem que inicie o escoamento superficial e é dada em mm/h. É variável no tempo por conta do umedecimento do solo e da passagem da água das camadas superiores às inferiores do solo.

A capacidade de infiltração pode ser calculada por diversos métodos, desde a medição em campo por infiltrômetros, ou com fórmulas empíricas tais quais as fórmulas de Horton, Philip, Green e Ampt e algoritmo de Berthelot, ou, ainda, estimativa a partir de chuvas e vazões observadas utilizando-se de hidrogramas.

O equacionamento geral da infiltração é uma representação matemática do movimento da água em solos não saturados (TUCCI, 2015) e pode ser descrito pela lei de Darcy. A Figura 3 relaciona os conceitos de infiltração, precipitação, escoamento superficial, taxa e capacidade de infiltração, as quais são base desta pesquisa.

FIGURA 3 - CURVAS DE CAPACIDADES E TAXA DE INFILTRAÇÃO



Fonte: TUCCI (2015)

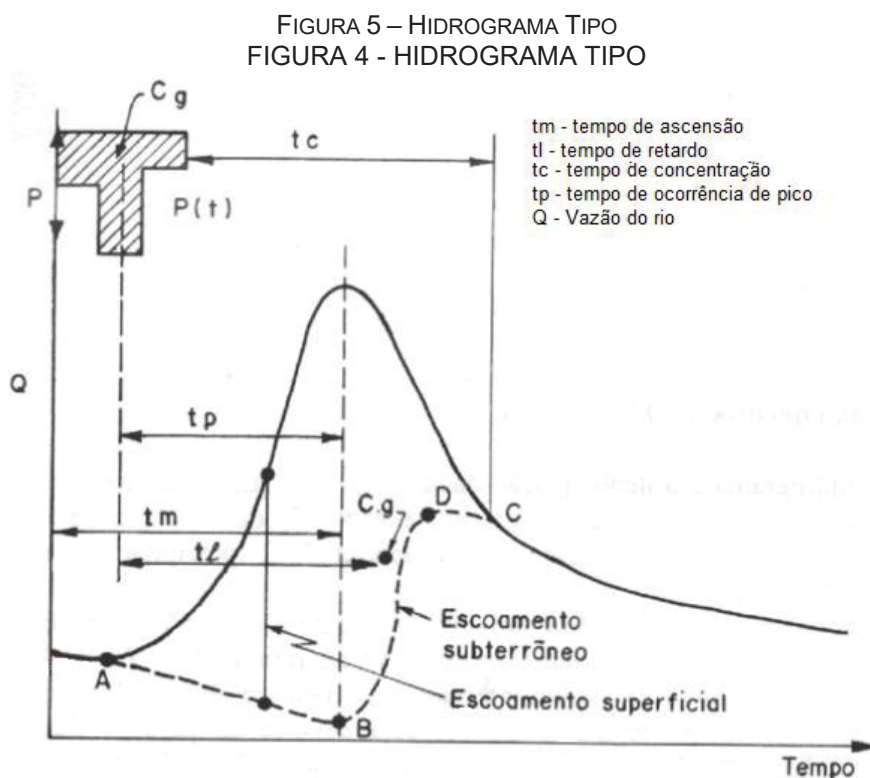
De forma simplificada, é possível perceber que a infiltração ocorre desde o início da precipitação até que se atinja o encharcamento das camadas superiores do solo, gerando em seguida o escoamento superficial. A infiltração continua ocorrendo conforme a água infiltrada é transportada para as camadas inferiores, mas é notória a diminuição da capacidade de infiltração.

A água infiltrada sofre um processo de redistribuição interna ao movimentar-se no solo que, quando cessado, conceitua a capacidade de campo pela quantidade de água retida pelo solo.

O escoamento subterrâneo está presente onde a água armazenada ultrapassa a zona de aeração e atinge a zona de saturação, em que se localizam os aquíferos. Os aquíferos podem ser confinados, quando a pressão é maior que a atmosférica e está limitado impermeavelmente em suas porções superiores e inferiores, ou não confinados. Estes reservatórios subterrâneos escoam lentamente aos rios e mares.

2.1.3 ESCOAMENTO SUPERFICIAL

O escoamento superficial, conforme apresentado anteriormente, inicia-se após a saturação da camada superior de solo, a qual ocorre devido à precipitação. Um evento de precipitação e suas consequências na vazão do rio são correlacionados por um hidrograma (Figura 4).



Fonte: TUCCI (2015)

O hidrograma de um ponto no curso d'água representa os fenômenos físicos que ocorrem na bacia. Assim que começa um evento de precipitação, não há aumento do nível do rio porque a chuva está infiltrando no solo, se este ainda não estiver saturado, como exemplificado na Figura 4. Assim que o nível do rio começa a subir, a

maior contribuição é proveniente do escoamento superficial, que tem seu início indicado pelo ponto A. Após atingir a cota máxima, a recessão no hidrograma é evidente, até o ponto de inflexão C, que representa o momento em que o escoamento superficial deixa de ocorrer na bacia e o escoamento subterrâneo torna-se o alimentador da vazão existente no corpo hídrico.

O formato de um hidrograma é influenciado pelas diversas características da bacia, como o tipo do solo, uso da terra e topografia, como também pelas modificações artificiais na hidrografia e no próprio evento de precipitação, com suas características de distribuição temporal, espacial, intensidade e até as condições de umedecimento do solo causada por precipitações anteriores.

A representação por modelos do escoamento superficial é, normalmente, limitada devido à variabilidade das condições físicas das bacias. Os modelos mais utilizados são os unidimensionais ou bidimensionais, classificados principalmente pela linearidade das equações que os representam e por serem empíricos ou conceituais.

2.1.4 EVAPOTRANSPIRAÇÃO

A evapotranspiração ocorre quando a água é convertida em vapor líquido pela energia solar e, então, para a atmosfera. As variáveis que interferem na capacidade de evapotranspiração são a radiação solar, temperatura do ar, vento, pressão de vapor e a resistência de fluxo das plantas.

A evapotranspiração pode ser medida diretamente por um lisímetro ou indiretamente por medidas de umidade do solo, métodos baseados em temperatura, métodos baseados na radiação ou método combinado.

A medida de evapotranspiração pode ser indicada em termos potenciais (ETP), que quantifica a água transferida à atmosfera por evaporação e transpiração considerando uma superfície extensa completamente coberta de vegetação de porte baixo e bem suprida de água, como em termos reais (ETR), considerando a quantidade de água transferida nas condições existentes de fatores atmosféricos e umidade do solo, segundo Tucci (2015).

2.1.5 RETIRADA E LANÇAMENTO DE ÁGUA DE FONTES ANTROPOGÊNICAS

Por ações humanas, a água é movimentada para ser utilizada e devolvida ao ambiente. Nesse sentido que existem as outorgas de recursos hídricos, a ser tratado no item 2.3, que autorizam a retirada de água, o lançamento ou o uso tanto nos corpos hídricos superficiais quanto nos subterrâneos.

A água pode ser utilizada no próprio corpo hídrico por construções de obras hidráulicas como barragens, retificações, canalizações, drenagens e travessias ou para serviços como dragagem minerária, limpeza e desassoreamento. Ainda que a água permaneça no local, tais atividades mudam o ciclo hidrológico natural por interferir nas condições do solo, atmosféricas e de qualidade da água.

A água também pode ser retirada de um corpo hídrico superficial ou subterrâneo para fazer o abastecimento humano, dessedentação animal, aquicultura, irrigação, entre tantos outros usos que justifiquem a captação deste recurso para uso, alterando a quantidade de água disponível no sistema e, principalmente, no local e a jusante da retirada.

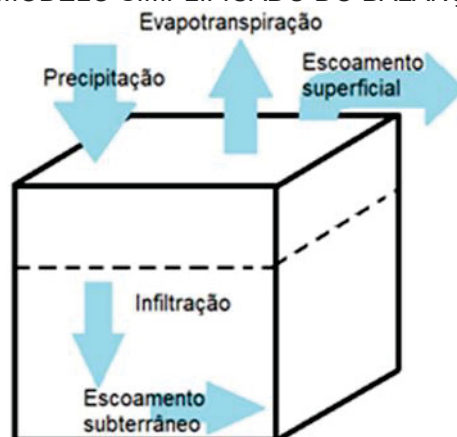
A água retirada, entretanto, volta ao sistema pelo lançamento deste recurso no rio, normalmente em forma de efluente sanitário ou industrial, alterando também a qualidade da água e a quantidade de água no sistema no local de despejo e a jusante deste ponto.

2.1.6 MODELOS HIDROLÓGICOS

Para compreender o ciclo hidrológico, cada etapa pode ser baseada em uma longa série de medição nas estações pluviométricas e fluviométricas. Dados de vazão para cada etapa não são medidos diretamente, sendo necessária uma separação das águas pós-precipitação, para que se entenda a quantidade de água no tempo. Nestas situações são utilizados modelos hidrológicos, que normalmente têm como base o balanço hídrico.

O conceito de balanço hídrico é base de um modelo hidrológico, que considera as entradas e saídas de água do sistema. De uma forma resumida, em uma bacia hidrográfica as entradas de água podem ser consideradas como a precipitação e o escoamento subterrâneo, enquanto as saídas podem ser consideradas como a evapotranspiração e os escoamentos superficial e subterrâneo. Há ainda a infiltração, que é um movimento da água dentro do sistema, conforme a Figura 5.

FIGURA 5 - MODELO SIMPLIFICADO DO BALANÇO HÍDRICO



Fonte: A Autora (2021).

Cada parcela de saída de um balanço hídrico transporta as cargas poluentes de uma maneira diferente e, por isso, precisam ser separadas e receberem abordagens diferentes. Podem ser ainda realizadas outras separações de escoamento, como por exemplo o escoamento superficial superior e sub-base, que têm comportamentos distintos e carregam diferentes tamanhos de partículas em diferentes velocidades. A abordagem a ser realizada depende de como estes transportes serão simplificados e representados.

Os modelos hidrológicos podem ser classificados como modelos determinísticos – aqueles em que a chance de ocorrência das variáveis envolvidas é ignorada, seguindo uma lei definida – ou estocásticos – conceito de probabilidade é introduzido na formulação do modelo. Podem ainda ser classificados em modelos empíricos – em que não há representação explícita dos processos físicos da bacia – ou conceituais – baseados nas equações que descrevem processo físico conceitual ou hipotético. Além disso, podem ser divididos em modelos concentrados – em que a área da bacia é considerada uma só, homogênea – ou distribuídos – em que a área total é dividida em unidades homogêneas irregulares ou regulares, considerando uma variabilidade espacial; este último adequado à bacia em questão (CLAKER, 1973 apud ALMEIDA e SERRA, 2017).

Atualmente existem diversos modelos hidrológicos, cada um com suas peculiaridades. São alguns deles: SWAT, MIKE-SHE, APEX, SCS, Stanford IV, ANSWERS, Tank, HEC-HMS, entre outros.

2.2 A QUALIDADE DA ÁGUA

A qualidade da água pode ser representada por meio de diversos parâmetros, que traduzem as suas principais características físicas, químicas e biológicas (SPERLING, 2017). Preponderantemente, estas características se apresentam em estados de sólidos em suspensão, sólidos dissolvidos e gases dissolvidos.

São parâmetros físicos da água: cor, turbidez, sabor, odor e temperatura.

São parâmetros químicos da água: pH, alcalinidade, acidez, dureza, quantidades de ferro, manganês, cloretos, nitrogênio, fósforo, oxigênio dissolvido (OD), matéria orgânica, micropoluentes inorgânicos e micropoluentes orgânicos.

Os parâmetros biológicos da água, por sua vez, são os microorganismos, tais quais bactérias, algas, fungos, protozoários, vírus e helmintos.

A alteração das características naturais da água, de forma a prejudicar o uso à qual está destinada, é chamada de poluição. A Política Nacional de Meio Ambiente, regida pela Lei 6938/81 (BRASIL, 1981), conceitua a poluição como a degradação da qualidade ambiental resultante de atividades que direta ou indiretamente: prejudiquem a saúde, a segurança e o bem-estar da população; criem condições adversas às atividades sociais e econômicas; afetem desfavoravelmente a biota; afetem as condições estéticas ou sanitárias do meio ambiente; ou, lancem matérias ou energia em desacordo com os padrões ambientais estabelecidos.

É de se notar, portanto, que a qualidade da água desejada depende do uso a que está destinada. Os padrões exigidos no Brasil são ambientados por quatro principais leis: Portaria 2.914 (BRASIL, 2011b) do Ministério da Saúde para o padrão de potabilidade, Resolução CONAMA 357 (BRASIL, 2005) do Ministério do Meio Ambiente para os padrões de corpos d'água, a Resolução CONAMA 430 (BRASIL, 2011a) para os limites de lançamento e a Resolução CONAMA 274 (BRASIL, 2000) de balneabilidade. Todas estas são complementadas por legislações estaduais.

Em atenção especial à Resolução CONAMA 357 (BRASIL, 2005), esta dispõe sobre a classificação dos corpos d'água e diretrizes ambientais para o enquadramento, assim como estabelece condições e padrões de lançamento de efluentes, sendo base para a implementação do instrumento de gestão de recursos hídricos do enquadramento, proposto pela Política Nacional de Recursos Hídricos, Lei 9433/97 (BRASIL, 1997).

O enquadramento separa as águas doces, salobras e salinas em classificações de acordo com o uso da água. Especialmente para as águas doces, foco desta pesquisa, são cinco:

- (i) Classe Especial: destinadas ao abastecimento para consumo humano, com desinfecção; à preservação do equilíbrio natural das comunidades aquáticas; e, à preservação dos ambientes aquáticos em unidades de conservação de proteção integral;
- (ii) Classe 1: destinadas ao abastecimento para consumo humano, após tratamento simplificado; à proteção das comunidades aquáticas; à recreação de contato primário; à irrigação de hortaliças que são consumidas cruas e frutas que se desenvolvam rentes ao solo e que sejam ingeridas cruas sem remoção de película; e, à proteção das comunidades aquáticas em Terras Indígenas;
- (iii) Classe 2: destinadas ao abastecimento para consumo humano, após tratamento convencional; à proteção das comunidades aquáticas; à recreação de contato primário; à irrigação de hortaliças, plantas frutíferas e de parques, jardins, campos de esporte e lazer, com os quais o público possa vir a ter contato direto; e, à aquicultura e à atividade de pesca;
- (iv) Classe 3: destinadas ao abastecimento para consumo humano, após tratamento convencional ou avançado; à irrigação de culturas arbóreas, cerealíferas e forrageiras; à pesca amadora; à recreação de contato secundário; e, à dessedentação de animais;
- (v) Classe 4: destinadas à navegação; e, à harmonia paisagística.

Para cada classe, são definidos padrões gerais mínimos de qualidade da água, que compreendem características de toxicidade, materiais flutuantes, óleos e graxas, gosto e odor, corantes, resíduos sólidos objetáveis, coliformes termotolerantes, DBO (Demanda Bioquímica de Oxigênio, utilizada para quantificação de matéria orgânica), OD (Oxigênio Dissolvido), turbidez, cor, pH, além de uma extensa lista de valores máximos para parâmetros orgânicos e inorgânicos. Os padrões de concentração de substâncias são concernentes à vazão de referência adotada para o corpo hídrico, em geral legisladas por Estado.

Nas bacias hidrográficas em que a condição de qualidade dos corpos de água esteja em desacordo com os usos preponderantes pretendidos, metas obrigatórias deverão ser estabelecidas (BRASIL, 2005).

Para a definição de tais metas e também o estabelecimento de ações que visem a melhora da qualidade da água, é necessário que se entenda como a poluição chega ao corpo hídrico e em qual quantidade. Neste sentido, são apresentadas a seguir as duas formas básicas em que a fonte de poluentes atinge o corpo hídrico: poluição pontual e poluição difusa.

2.2.1 POLUIÇÃO PONTUAL E DIFUSA

As fontes de poluição podem ser pontuais ou difusas. A poluição hídrica advinda de descarga pontual ou difusa pode ser de diferentes naturezas, chamadas de vias de poluição.

As fontes pontuais são caracterizadas por descargas localizadas, como os despejos de efluentes domésticos, sejam in natura ou após tratamento, ou despejos industriais.

A poluição de fonte pontual vem, em sua maioria, de lançamentos de estações de tratamento de esgoto (ETEs), lançamentos de esgoto doméstico bruto (domicílios que não têm seu esgoto tratado e/ou até mesmo não coletado) e do lançamento de indústrias. Em alguns locais e dependendo da escala de avaliação, os despejos de águas pluviais pelas galerias também são considerados pontuais, pois mesmo que a água provenha de um escoamento superficial de uma grande área distribuída, este escoamento é concentrado na galeria e despejado pontualmente no rio.

As fontes difusas são aquelas ocasionadas pelo transporte da poluição que se encontra dispersa em uma grande porção de espaço, por meio do escoamento das águas superficiais, das águas subterrâneas ou da deposição atmosférica, por exemplo. Os poluentes das fontes difusas são despejados ao longo do corpo hídrico, sendo mais difíceis de quantificar por não serem pontuais.

A poluição de fonte difusa está, na maioria das vezes, relacionada a precipitação e ao uso da terra, mas as vias de poluição são relacionadas ao ciclo hidrológico, tendo sua classificação dependente da importância de cada uma dentro da área de estudo. Em geral, as principais vias de poluição são a erosão (para os poluentes aderidos aos sólidos), o escoamento superficial (para os poluentes

dissolvidos), o escoamento subterrâneo (atrelado à infiltração) e a deposição atmosférica.

Em áreas rurais, a poluição difusa pode advir das atividades agrícolas, principalmente pelos fertilizantes e agrotóxicos aplicados no solo (GONÇALVES et al., 2000 citado por SODRÉ, 2012), ou pela pecuária. Em regiões urbanas, vários poluentes são carregados aos corpos d'água pelas águas de drenagem, com vazões de escoamento superficial mais altas devido à grande impermeabilização do solo.

2.2.2 POLUENTES E INDICADORES

A presença de substâncias químicas, físicas e biológicas na água é normal e até mesmo inevitável, porém a presença delas em excesso pode comprometer a saúde humana, além das atividades ligadas ao uso da água. Sendo assim, é essencial que seja feito o controle e acompanhamento da qualidade, em especial nas áreas de mananciais, assim como o tratamento apropriado para adequação aos padrões exigidos de potabilidade. Dentre os parâmetros físico, químicos e biológicos que têm recebido mais atenção no que se refere à gestão da qualidade da água, podem ser destacados:

Fósforo Total (PT): nutriente encontrado na água devido à drenagem urbana, atividade agrícola e, principalmente, ao lançamento de esgotos. Apesar de não integrar o padrão de potabilidade e não ser uma substância que torne necessariamente a água imprópria para consumo humano, a presença do fósforo é importante de ser estudada por ser, junto com o nitrogênio, o limitante do crescimento excessivo de plantas aquáticas e principal estimulante do fenômeno de eutrofização, que se caracteriza pelo crescimento em excesso de plantas aquáticas e algas que, quando em grandes quantidades, podem inviabilizar o uso da água. Os problemas predominantes causados pela eutrofização incluem os problemas estéticos, condições anóxicas, morte de peixes, além do aumento no custo e maior dificuldade no tratamento de água. A dificuldade no tratamento de água ocorre por conta do atendimento, principalmente, dos padrões organolépticos, devido à necessidade de remoção das algas em si, da cor, sabor e odor da água para consumo, exigindo um maior consumo de produtos químicos, além de demandar maior manutenção da estação de tratamento, em especial no que se refere às lavagens dos filtros.

Nitrogênio Total (NT): nutriente encontrado nas águas originária de áreas de drenagem e da decomposição de matéria orgânica, além de ser influenciada também pela aplicação de fertilizantes agrícolas, despejos industriais e esgotos domésticos. Juntamente com o fósforo, é um dos principais causadores da eutrofização. Pode causar sérios riscos à saúde humana, em especial das crianças que ainda mamam e de mulheres grávidas. Não existe um valor limite direto para NT na CONAMA 357.

Sólidos Suspensos Totais (SST): indicador que influencia, entre outros parâmetros, a turbidez da água, que por sua vez corresponde à redução da transparência da água. Portanto, a presença dos sólidos em suspensão influencia os padrões organolépticos de potabilidade como cor, odor e sabor.

Demanda Bioquímica de Oxigênio (DBO): indicador que revela a existência de matéria orgânica na água e um dos principais critérios para se avaliar a poluição hídrica de origem orgânica e não regradatária. Expresso em mg/L, a DBO representa o consumo de oxigênio necessário para a decomposição, por processos bioquímicos, dos compostos orgânicos presentes nos corpos hídricos.

Oxigênio Dissolvido (OD): representa, como o nome diz, a quantidade de oxigênio dissolvido nos corpos de água expressa em mg/L. Seus teores sofrem influência de diversos fatores como temperatura, turbulência, processo de fotossíntese e salinidade, além da ação antrópica. A importância do OD está relacionada à depuração dos poluentes, visto que os microrganismos envolvidos no processo a utilizam para a respiração. Desta forma, é um indicador importante para determinação da poluição hídrica.

Carbono Orgânico Dissolvido (COD): consiste na fração de carbono orgânico composta por vários compostos como aminoácidos, carboidratos, hidrocarbonetos, ácidos orgânicos, húmicos e fúvicos, tendo origem na decomposição de plantas e animais, assim como de suas fezes. Além disso, pode ser originário de componentes de efluentes que entram no sistema ou pela fixação do dióxido de carbono. Constitui-se como importante fonte de energia, tanto para os sistemas aquáticos, como para os sistemas terrestres, contribuindo com a acidez da água e afetando complexação, solubilidade e mobilidade de metais. Uma das principais formas de entrada do carbono orgânico dissolvido no sistema se dá pela precipitação, sendo também fortemente influenciada pelo tipo do solo, assim como seu uso. É um processo utilizado para estimar a concentração de matéria orgânica.

2.3 CONSIDERAÇÕES SOBRE A GESTÃO DE RECURSOS HÍDRICOS

A Política Nacional de Recursos Hídricos, Lei nº 9433/97 (BRASIL, 1997), instituiu cinco instrumentos de gestão de recursos hídricos, descritos anteriormente: Plano de Recursos Hídricos, Enquadramento dos Corpos d'Água, Outorga de Direito de Uso dos Recursos Hídricos, Sistema de Informações sobre Recursos Hídricos e Cobrança pelo Uso dos Recursos Hídricos.

O Relatório de Conjunturas de Recursos Hídricos do Estado do Paraná (IAT, 2020) afirma que há “necessidade de um maior controle sobre o uso da água em um cenário de avanço da degradação do ambiente natural”, sendo que esse controle é regido pelos instrumentos de gestão hídrica. No país, os instrumentos estão em diferentes fases de aplicação e fiscalização em cada uma das unidades da federação e diversas dificuldades vêm sendo encontradas.

Este item busca refletir brevemente sobre os instrumentos de gestão hídrica e a relação destes com a variabilidade intrínseca dos regimes hidrológicos, ou seja, “a flutuação da vazão do rio entre vazões de cheia e de estiagem” (ANA, 2019b), peça fundamental para a aplicação de tais ferramentas.

Ainda, “a Política Nacional de Recursos Hídricos (PNRH) tem como principal objetivo garantir à atual e às futuras gerações a necessária disponibilidade de água, em padrões de qualidade e quantidade adequados” (BRASIL, 1997), sendo esse o objetivo geral da aplicação destes instrumentos, em que entender as variabilidades espacial e temporal são a chave para uma boa gestão dos recursos.

O instrumento de enquadramento dos corpos d'água é essencial no sentido de estabelecer metas de qualidade e definir plano de ação para garantir seu cumprimento às futuras gerações. O enquadramento é um instrumento de planejamento, desenvolvido compatível com o Plano de Recursos Hídricos, que é elaborado em quatro etapas: diagnóstico, prognóstico, proposta de metas relativas às alternativas de enquadramento e programa de efetivação (ANA, 2020a). A eficácia de planejamento deste instrumento é refletida na elaboração dos instrumentos de outorga e cobrança, uma vez que precede e dá suporte a elaboração destes, sendo subsidiado pelo instrumento de sistema de informações (incluindo o monitoramento de quantidade e qualidade da água), conforme demonstrado na Figura 6.

FIGURA 6 - RELAÇÃO ENTRE OS INSTRUMENTOS DE GESTÃO HÍDRICA



Fonte: ANA, 2020a.

A classificação dos corpos hídricos é definida conforme os usos pretendidos, estabelecendo metas de qualidade da água. A resolução CONAMA 357/05 (BRASIL, 2005) determina que os “valores máximos estabelecidos para os parâmetros relacionados em cada uma das classes de enquadramento deverão ser obedecidos nas condições de vazão de referência” que, em termos temporais, situa limites de concentração de poluentes aceitáveis para o enquadramento. As vazões de referência, contudo, são diferentes nas unidades da federação, implicando em diferentes restrições do uso dos recursos hídricos.

Um grande questionamento, entretanto, está na medição e acompanhamento destes valores, competência esta do monitoramento. Uma vez que “para caracterização da qualidade atual da água, os dados de monitoramento realizados pelos órgãos gestores de recursos hídricos e de meio ambiente são a referência para identificar as conformidades e desconformidades com os usos existentes ou pretendidos” (ANA, 2020a) e com os padrões da Resolução CONAMA 357/2005 (BRASIL, 2005), a obtenção destes dados deveria contemplar diferentes características temporais e espaciais. As vazões de referência utilizadas na verificação do atendimento à classe utilizam dados contínuos, em sua maioria, de extensos períodos, englobando as variações temporais sofridas pelo corpo hídrico. As análises de qualidade da água, todavia, não acompanham essas variações. Além de não terem intervalos de tempo bem definidos, as condições às quais deveriam ser coletadas para serem representativas (como amostragem que contemple

sazonalidade, momentos de chuva de diferentes intensidades, condições extraordinárias, entre outros) não são existentes.

A aplicação de diferentes metodologias de atendimento às classes, como as apresentadas por Machado et al. (2019) de estatística simplificada ou de curva de permanência, serão representativas dependendo da qualidade e representatividade das informações que as alimentam. Se os dados de qualidade da água não são coletados em dias de chuva, por exemplo, a carga de poluição difusa mais expressiva é ignorada, gerando um cenário de condição do rio melhor do que o real. Ainda, os planos de ação previstos para o enquadramento podem não surgir o efeito previsto ou precisarão ser aplicados em maior escala para atingir o objetivo das metas progressivas e finais.

Analisando-se o objetivo do instrumento e considerando a grande variabilidade temporal das concentrações e vazões em um corpo hídrico, Machado et al. (2019) sugeriram a utilização de curvas de permanência de cargas máxima, calculadas pela multiplicação das vazões do corpo hídrico pelas concentrações limite de classe. Esta abordagem permite comparar as cargas geradas e transportadas que, afinal, compõe o objetivo final de alocar e diminuir tais cargas. Esta metodologia aplicada e analisada por Acosta et al. (2019, em desenvolvimento) na bacia do Alto Iguaçu, Paraná, identificou que as curvas de permanência considerando os limites de classe para concentração e para carga não se comportaram da mesma maneira, evidenciando que a forma de entender o problema de poluição precisa ser olhado de diferentes óticas para que as prioridades de ação não fiquem prejudicadas. Sendo o enquadramento um instrumento de estabelecimento de objetivos e metas com planos de ação objetivos, a forma de classificação dos corpos hídricos afeta o objetivo principal e, portanto, a eficiência de aplicação da ferramenta. A utilização de metodologias que considerem especificamente a variação temporal das cargas pode, pois, gerar percepções diferentes acerca de quais ações podem ser realizadas e a ferramenta do enquadramento ser aplicada com êxito.

Os causadores de poluição geram carga poluente e não concentração, pois a concentração é o entendimento da poluição já no corpo hídrico. A vazão do rio não é dependente e não atende, exceto em casos específicos, as determinações humanas. Desta forma, para contribuir com a despoluição hídrica, os planos precisam considerar

e elaborar ações aos atores de uma bacia hidrográfica que são os geradores da carga poluidora.

Porras (2014) demonstrou um estudo de estimativa de geração de cargas poluentes na bacia hidrográfica do Alto Iguaçu por via de poluição (caminhos em que a poluição é carregada até o corpo hídrico), indicando que a poluição difusa compõe de maneira expressiva a matriz de geração poluidora, resultado também observado no estudo recente de estimativa de geração de cargas poluentes na bacia do Paranapanema (ANA, 2021, em desenvolvimento). Tais resultados corroboram com o argumento de que a representatividade da poluição difusa nos mecanismos de classificação dos rios é essencial para que bons prognósticos e planos de ação sejam gerados, assim como afirmado por ANA (2020a).

Para a utilização da água em quantidade significativa por qualquer usuário para os usos a qual se destina, é necessária a obtenção da autorização de uso do recurso hídrico, uma vez que a água é um bem público e sua alocação é administrada pelo poder público. A outorga dos recursos hídricos tem finalidade de garantir e autorizar o usuário a utilizar a água que lhe foi outorgada, garantindo uma segurança jurídica e quantitativa e permitindo à administração pública gerenciar este bem público (ANA, 2019b). Tecnicamente, para garantir a quantidade de água, a outorga utiliza-se assim como o instrumento de enquadramento de vazões de referência, como por exemplo a $Q_{95\%}$ (vazão disponível em 95% do tempo sob visão estacionária), assumindo certo risco para o não cumprimento da garantia de quantidade de água. Ainda, a avaliação técnica de outorga relaciona-se com o instrumento de enquadramento dos rios, uma vez que a qualidade das águas conforme a classe do rio enquadrado (ou não) é analisada para a emissão da outorga, principalmente de lançamento de efluentes.

Em termos de poluição do rio, o instrumento de outorga não acompanha a variabilidade de sua qualidade no tempo, o que retorna à questão anterior, ainda que compreenda a variabilidade quantitativa da água. Se o enquadramento considera a classe a partir da qualidade da água em condições generalistas, quando um evento de chuva leva ao corpo d'água uma carga poluidora expressiva e a outorga de um lançamento de efluentes neste mesmo rio, por exemplo, continua constante dentro do seu limite, o rio poderá estar extrapolando os limites de sua classe durante um período. Uma questão relevante é entender qual o período suficiente para que surjam efeitos negativos no rio, como a alteração na biota ou impacto em outros usuários.

Outra questão relevante ao tema é a definição dos próprios mecanismos de mitigação de poluição. ANA (2020a) destaca que “em relação às cargas difusas, há dificuldades adicionais que vão desde a produção de estimativa confiável dessas cargas até o estabelecimento de ações que possam ter efetividade na sua mitigação”. A melhor compreensão da importância do tratamento da poluição difusa no instrumento do enquadramento e, subsequentemente, na outorga, pode levantar novas ideias de controle desta poluição. As outorgas de lançamento de efluentes e de captação podem vir a se tornar variáveis no tempo, conforme as condições climáticas, sendo outorgas sazonais ou dependentes de um índice variável mensal, por exemplo. Ainda, podem ser consideradas outorgas de lançamento de drenagem urbana, por exemplo, uma vez que estas podem ser contribuintes de alta carga e que hoje não são tratadas.

Diversas outras ideias de aplicação de mecanismos podem também ser adotadas quando na aplicação do instrumento de cobrança pelo uso dos recursos hídricos. Este instrumento objetiva dar ao usuário uma indicação do real valor da água, incentivar o uso racional da água e obter recursos financeiros para recuperação das bacias hidrográficas (ANA, 2019a). Em relação específica ao último objetivo, a aplicação deste instrumento em áreas contribuintes da poluição difusa pode vir a induzir um comportamento de redução da geração das cargas poluentes, de sustentabilidade ou de aplicação de mecanismos de tratamento desta poluição, por exemplo.

A exemplo da cobrança no setor elétrico que conta com tarifas variáveis, como a majoração em horário de ponta (maior demanda), as outorgas de lançamento pontual em grandes proporções poderiam sofrer cobrança quando o lançamento fosse realizado em momentos críticos determinados, por exemplo.

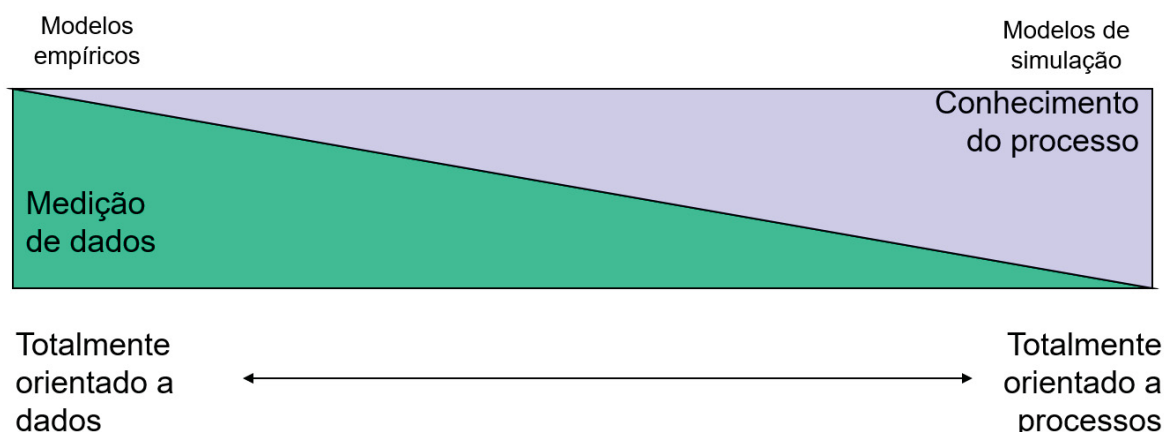
As ideias, ferramentas e metodologias serão desenvolvidas desta maneira apenas quando houver um controle mais avançado sobre a poluição efetiva nos corpos hídricos, considerando os poluidores pontuais e difusos, de forma representativa espacial e temporalmente. Ou seja, haverá um aperfeiçoamento dos processos quanto mais completo, e provavelmente complexo, foi o conhecimento acerca das fontes poluidoras e comportamento das cargas nos rios e reservatórios, principalmente acerca da variabilidade espacial e temporal.

2.4 MODELOS MATEMÁTICOS

Modelos matemáticos têm sido utilizados como ferramenta no planejamento e gestão de bacias hidrográficas podendo auxiliar em algumas atividades de diagnóstico e de previsão, bem como no acompanhamento de metas por meio da simulação das condições da qualidade da água de um corpo hídrico no que se refere a determinados parâmetros.

Modelos são representações de sistemas e processos, que dependem dos dados de entrada e são reflexos de como eles são construídos (processos considerados ou método de construção). Podem ser classificados de empírico, sendo totalmente orientados a dados, ou de simulação, totalmente orientados a processos, conforme ilustrado na Figura 7, podendo ter modelos híbridos entre os dois extremos.

FIGURA 7 - MODELOS MATEMÁTICOS



FONTE: A Autora (2021)

Os modelos numéricos com assimilação de dados são formados por conhecimentos de processos físicos e complementados com técnicas de assimilação de dados de campo. Modelos numéricos determinísticos são totalmente orientados a processos, em que o conhecimento do sistema está contido na forma de equações de processos físicos, químicos e/ou biológicos.

Diversos modelos são utilizados para simular e estimar a qualidade das águas. Os modelos usualmente empregados para simular a qualidade da água são o HEC-HAS (Hydrologic Engineering Center's River Analysis System), SIMCAT (SIMulation CATchment), TOMCAT (Temporal/Overall Model for CATchment, PC-QUASAR, MIKE-11, WASP (Water Analysis Simulation Program), AQUATOX, SWAT (Soil and Water Assessment Tools), QUASAR (QUALity Simulation Along Rivers),

QUAL2E, HSPF (Hydrological Simulation Program FORTRAN), CE-QUAL-W2, CE-QUALRIV1, HEC-5Q/QRRS e QUAL2K (TSCHIEDEL, 2013; FLECK et al., 2013; SILVA, 2009; TERCINI, 2014).

Ferreira et al. (2018) apresentam diversos modelos de qualidade da água (SWMM (Storm Water Management Model), SWAT (Soil & Water Assessment Tool), HSPF (Hydrological Simulation Program FORTRAN), AGNPS (AGricultural Non-Point Source Pollution Model), AnnAGNPS (Annualized Agricultural Non-Point Source Pollutant Model), CE-QUAL-W2, MIKE SHE, HEC-HMS, STORM e WEPP (Watershed Erosion Prediction Project)) que contemplam as cargas pontuais e difusas, comparando as informações básicas de cada modelo, as bases do equacionamento matemático, a aplicabilidade em SIG, os recursos computacionais e os parâmetros de qualidade da água simulados.

Modelos de inteligência artificial são usados para detecções de padrões, podendo relacionar causas e efeito. Estes modelos são construídos baseados em conceitos estatísticos, com base em dados (modelos empíricos), mas podem ter em sua estrutura alguns conceitos físicos, ou mesmo pré-condições estabelecidas (modelos híbridos).

A definição do melhor modelo a ser utilizado depende intrinsecamente do objetivo a qual será aplicado, uma vez que todos têm vantagens e limitações de uso, mas também da obtenção dos dados de entrada necessários que alimentam tais modelos, sendo esta uma das maiores dificuldades de aplicação dos modelos.

2.4.1 INTELIGÊNCIA ARTIFICIAL

A inteligência artificial objetiva desenvolver paradigmas ou algoritmos que utilizem de máquinas para realizar tarefas intelectuais (HAYKIN, 2001). Dentre os tipos de inteligência artificial, o *Machine Learning* (aprendizado de máquina) é a aprendizagem que se adapta a novas circunstâncias, detectando e extrapolando padrões (RUSSEL E NORVIG, 1995), em que a máquina pode aprender de três maneiras: aprendizado por reforço, aprendizado não supervisionado (por agrupamento ou redução dimensional) ou aprendizado supervisionado (por classificação ou regressão).

Para a predição numérica de alguma variável podem ser utilizadas técnicas de regressão, em que o aprendizado é supervisionado, ou seja, são fornecidos

exemplos de entrada e saída esperada e o modelo estuda as relações entre as variáveis numéricas de forma regressiva. Desta forma, com um novo conjunto de dados de entrada, o modelo pode prever seu resultado.

A classificação e regressão de um conjunto de dados é convencionalmente estatística, mas estudos mostram que técnicas como Redes Neurais Artificiais (RNA), árvores de decisão e *Support Vector Machines* podem se tornar mais precisas (FOODY e MATHUR, 2004 citado por PEREIRA, 2013).

A seguir estão descritos alguns tipos de aprendizagem de regressão utilizados para encontrar os melhores modelos regressivos que representam a realidade, especialmente utilizados pela biblioteca Lazy Predict. Uma biblioteca é um conjunto de módulos, que são programas escritos na linguagem Python que fornecem soluções padronizadas para problemas contínuos. A biblioteca Lazy Predict, especialmente, agrupa os diversos modelos regressivos (da biblioteca scikit-learn, de modelos de aprendizagem de máquina) a fins comparativos e de escolha quando um conjunto de dados é executado no *machine learning*. As definições estão baseadas nas informações sobre os modelos na biblioteca scikit-learn, disponíveis na internet.

2.4.1.1 Redes Neurais Artificiais (RNA)

A Rede Neural Artificial (RNA) é um processador maciçamente paralelo (em que grandes problemas computacionais são divididos em problemas menores, que são resolvidos em diversos processadores, ao mesmo tempo), constituído por unidades de processamento simples, que tem uma tendência natural para armazenar conhecimento empírico e torná-lo disponível para uso, segundo Pretopoulos, Arvanitis e Sigrimis (2012) apud Pereira (2013). É uma poderosa ferramenta para modelagem quando são desconhecidas as relações entre dados (LEK e GUÉGAN, 1999 citado por PEREIRA, 2013). É composta por diversas camadas organizadas e conectadas, normalmente classificadas em três grupos: camadas de entrada, camadas ocultas ou intermediárias e camada de saída.

A RNA pode apresentar um fluxo unidirecional, alimentada pelas camadas de entrada, sendo processadas nas camadas ocultas, até chegarem às camadas de saída, sendo conhecida como rede *feedforward*, em que elementos não-lineares (neurônios) são dispostos em camadas sucessivas, sendo a quantidade em cada camada variável. Conforme Heermann e Khazenie (1992) citado por Pereira (2013),

se poucos neurônios são colocados na camada oculta, a rede pode não conter o grau de liberdade necessário na análise. Porém, um grande número de neurônios, além de exigir desempenho computacional, pode perder a capacidade de generalização da rede, sem boas respostas para novos dados, sendo necessário um equilíbrio.

As RNAs objetivaram, em sua concepção, reproduzir o aprendizado humano e fazer conexões entre informações. São conceituadas atualmente como um “processador distribuído constituído de unidades de processamento simples, que têm a propensão natural para armazenar conhecimento experimental e torná-lo disponível para uso.” (HAYJIN, 2001).

A vantagem da utilização da RNA é produzir saídas, ou resultados, adequados às entradas que não foram utilizadas para treinamento, a chamada “generalização”. Ou seja, prever um resultado a partir de um conjunto de dados de entrada, sem que estes tenham sido inicialmente apresentados à rede, a partir de outros conjuntos de dados semelhantes ao ambiente a que a predição está inserida e com um conjunto de respostas conhecido. Por esse tipo de aprendizagem, problemas não-lineares podem ser compreendidos.

A rede neural artificial, sendo baseada no cérebro humano, é composta de neurônios e sinapses. O conhecimento é adquirido pelos neurônios por processos de aprendizagem e é armazenado pelas forças de conexão das sinapses, através dos pesos sinápticos.

Segundo Hayjin (2001), um neurônio é uma unidade de processamento de informação composto por três elementos básicos: um conjunto de sinapses (dados de entrada relacionados com seus pesos sinápticos), um somador (pondera os sinais de entrada pelas respectivas sinapses) e uma função de ativação (restringe a amplitude de saída de um neurônio). Matematicamente, um neurônio pode ser expresso pela equação 1:

$$y_k = \varphi \left(b_k + \sum_{j=1}^m w_{kj} x_j \right) \quad (1)$$

Onde:

y_k = sinal de saída do neurônio

$\varphi()$ = função de ativação

b_k = bias

w_{kj} = peso sináptico do neurônio

x_j = sinais de entrada

A rede neural de múltiplas camadas é vantajosa quando o conjunto de camadas de entrada é grande, pois pode extrair estatísticas de ordem elevada. Isso ocorre por conta da estrutura da rede, uma vez que a primeira camada é a de entrada, que fornece elementos para a ativação dos neurônios que resultarão na segunda camada, a camada oculta. Esta, por sua vez, servirá de entrada para a nova ativação de neurônios, repetindo-se pela quantidade de camadas ocultas da rede, até chegarem à camada de neurônios de saída.

Para que uma rede neural aprenda e melhore seu desempenho são aplicados processos de aprendizagem, onde a rede neural é estimulada por um ambiente (dados de entrada), sofre modificações nos seus parâmetros livres (informação passa pelos neurônios e camadas, chegando à camada final) e responde de uma maneira nova ao ambiente, de acordo com suas modificações.

Para que cada vez que as informações passem pelo processo, a rede aprenda e se modifique, a aprendizagem pode ser realizada de forma supervisionada ou não. A supervisão indica que há um resultado conhecido a ser obtido pela rede, portanto seu aprendizado é avaliado por um índice escalar de desempenho, enquanto a não-supervisão espera um sistema baseado em condições pré-estabelecidas para criar classes de respostas.

A aprendizagem supervisionada pode ter seu índice de desempenho atrelado à correção de erro, um dos processos de aprendizagem. A correção de erro compara a resposta da rede à resposta esperada, acionando um mecanismo de controle que realiza ajustes corretivos aos pesos sinápticos dos neurônios até que seja encontrado um índice de desempenho mínimo.

Na biblioteca Lazy Predict, o modelo que se utiliza da técnica de redes neurais é o *MLR Regressor* (Multi-layer Perceptron), que treina por backpropagation (a partir do erro na camada de saída, recalcula os pesos dos neurônios das camadas anteriores, ou seja, da última camada para a primeira) sem função de ativação no dado de saída, ou seja, utiliza a função identidade como função de ativação. Além

disso, usa o erro quadrático como função de perda. É um regressor que também admite a regressão multisaída, em que uma amostra pode ter mais de um objetivo.

Ha e Stenstrom (2003) utilizaram a rede neural para examinar a relação entre a qualidade da água durante eventos de precipitação e diversos usos da terra por classificação, com 10 variáveis de qualidade da água (173 amostras), 4 neurônios na camada oculta e 5 variáveis de uso da terra como saída. O estudo comparou três diferentes abordagens, sendo a percepção multi-layer, a função de base radial e a rede Bayesiana, atingindo 92,3% de classificação correta para teste e 94,5% para treino na percepção multi-layer.

2.4.1.2 Support Vector Machine

Outra técnica para análise de dados é o *Support Vector Machine* (SVM), que, diferentemente da RNA que pode ter várias soluções, converge para uma única solução ótima. O SVM é não-paramétrico, não assumindo uma distribuição estatística conhecida a priori dos dados.

O SVM é essencialmente um classificador supervisionado que tem por objetivo encontrar um hiperplano que separa um conjunto de dados em classes discretas, utilizando-se de processo iterativo e de exemplos de treinamento para ajustar este hiperplano (BURGES, 1998 citado por PEREIRA, 2013). Este hiperplano é ideal ou ótimo, otimizando a separação das classes e maximizando a distância entre elas, utilizado como fronteira de decisão (PEREIRA, 2013). São esperados ótimos resultados, mesmo em conjuntos de treino pequenos, uma vez que a base da abordagem do SVM é de que apenas as amostras que se encontram nas fronteiras das classes são necessárias para discriminação.

Os SVM determinam um hiperplano que diferencia duas classes entre os dados de entrada, a partir de n -dimensões, sendo robusto para outliers. Suas vantagens são a efetividade em espaços dimensionais elevados, efetividade em casos em que o número de dimensões é superior ao número de amostras, é eficiente em termos de memória por usar um subconjunto de pontos de treinamento na função de decisão e é versátil por usar diferentes funções de Kernel (função de decisão, definida como função de dois pontos no espaço, geralmente não linear, que pode ser interpretada como a distância entre os pontos de entrada). Suas desvantagens incluem evitar o *overfitting* (sobreajuste) ao escolher as funções de Kernel se o número

de recursos for muito maior do que o número de amostras, uma vez que a regularização é crucial, além de não fornecer estimativas de probabilidade diretamente, fazendo-as usando uma extensa validação cruzada quádrupla, conforme documentação da biblioteca scikit-learn.

O método de Support Vector Regression é uma extensão do Support Vector Machine (Máquina de Suporte Vetorizada) desenvolvido para classificação sendo um algoritmo de aprendizado de máquina supervisionado, com foco no treinamento e classificação de um conjunto de dados. O modelo depende apenas de um conjunto de dados de treinamento em que a função de custo ignora as amostras cuja previsão está próxima ao alvo.

A SVR tem três tipos de implementação, todos presentes na biblioteca Lazy Predict: SVR, NuSVR e Linear SVR. A *Linear SVR* é implementada mais rapidamente, mas considera apenas Kernel linear.

Liu e Lu (2014) aplicaram o SVM, assim como uma rede neural artificial, para desenvolver um modelo de previsão de nitrogênio total e fósforo total em bacias agrícolas com contribuição da poluição difusa. As variáveis consideradas foram de fluxo do rio, temperatura da água, tempo de viagem do fluxo, precipitação, oxigênio dissolvido e as próprias concentrações de fósforo total e nitrogênio total a montante em escalas temporais mensais, bimestrais e trimestrais em 12 pontos de amostragem. Além disso, foi utilizado algoritmo genético para otimizar a parametrização. Dentre os modelos rodados, SVM chegou em um erro com valores testados de até, no mínimo, RSME = 0,114.

2.4.1.3 Árvores de Decisão

A árvore de decisão é um método de aprendizagem supervisionada não-paramétrica que tem como vantagem mapear bem as relações não-lineares, sendo os modelos preditivos de alta precisão, estabilidade e facilidade de interpretação.

Os modelos baseados em árvores de decisão dividem a população ou amostra em conjuntos homogêneos com base em divisores/diferenciadores mais significativos das variáveis de entrada, prevendo o valor de uma variável por aprendizagem de regras de decisão simples, podendo ser entendida como uma aproximação constante por partes.

As árvores de decisão são vantajosas por sua simplicidade de entendimento e interpretação, além de serem facilmente visualizadas, por requererem poucos dados de entrada (apesar de não suportarem valores nulos), aceitarem dados numéricos e categóricos, lidarem com problemas de múltiplas saídas, serem um modelo de caixa branca, haver possibilidade de validação por meio de testes estatísticos, o que torna o modelo mais confiável e apresentam bom desempenho.

As árvores de decisão são passíveis de desvantagens, entretanto, uma vez que pode ocorrer o chamado *overfitting*, em que as árvores se tornam complexas demais e não generalizam o problema. Para este problema, são utilizados mecanismos de poda, definição do número mínimo de dados de entrada em um nó ou definição da profundidade máxima da árvore. Além disso, as árvores podem ser sensíveis a variações e tomarem decisões boas localmente, não retornando boas decisões globalmente, caso não sejam utilizadas árvores em conjunto. Por serem, por definição, formadas por aproximações constantes por partes, não são boas em extrapolação. Por fim, podem ser criadas árvores tendenciosas se algumas classes dominam, sendo necessário equilibrar o conjunto de dados de entrada anteriormente à aplicação da árvore.

Na biblioteca Lazy Predict, o modelo *Decision Tree Regressor* é aquele que utiliza o conceito de árvore de decisões de regressão unidimensional, ajustando uma curva senoidal com ruído, aprendendo regressões lineares locais que se aproximam da curva senoidal. O modelo *Extra Tree Regressor* também é uma árvore de decisões (que deve ser usada em método de conjuntos) e é construído por divisões aleatórias para separar as amostras de um nó em dois grupos.

Oliveira (2016) utilizou as árvores de decisão para modelar concentrações de fósforo total e nitrato na bacia do Rio das Velhas, Minas Gerais. Ela utilizou dados de vazão de 8 estações fluviométricas ao longo do rio principal, de concentração de 17 estações de qualidade da água de amostras mensais e dados de uso do solo de 8 categorias para classificar as variáveis. Os dados de entrada foram separados quanto aos períodos secos e chuvosos. A partir de uma avaliação do número de instâncias ótimas utilizadas na árvore de decisão, foi obtido um sucesso de classificação de 83% para nitrato e 82% para fósforo total. Entretanto, é enfatizado que o “monitoramento mensal das variáveis de qualidade não permite a construção de sequências temporais

consistentes”. Apesar disso, os resultados mostraram que os usos urbano e agropastoril estão associados às alterações dos parâmetros estudados na bacia.

2.4.1.4 Conjunto de Modelos

O objetivo do conjunto de modelos (*ensemble*) é combinar as previsões de vários regressores construídos com um algoritmo de aprendizagem para melhorar a generalização e robustez de um único estimador. Normalmente são dois tipos de conjuntos: (i) métodos de cálculo da média, em que os estimadores são construídos independentes e são calculadas as médias de suas previsões, o que, normalmente, gera um estimador melhor do que um único estimador por reduzir a variância; (ii) métodos de boost, em que os estimadores são construídos sequencialmente, tentando-se reduzir o viés do estimador combinado, construindo um bom conjunto pela combinação de modelos mais fracos.

Na biblioteca Lazy Predict são 9 diferentes regressores que utilizam da técnica de conjunto de modelos, conforme descritos a seguir.

Os estimadores tipo *Bagging* utilizam métodos de ensacamento, que dividem o conjunto de treinamento original em subconjuntos aleatórios, estimando-os por métodos de caixa preta, que têm suas previsões individuais agregadas para formar uma previsão final. Esse método é utilizado para reduzir a variância de um estimador ao utilizar da aleatoriedade em sua construção, reduzindo o *overfittinge*, portanto, funcionando melhor com modelos fortes e complexos. Por *default*, o estimador Bagging utiliza como base o regressor *Decision Tree*.

Os estimadores tipo *Random Tree* (árvore aleatória), que são os regressores *Random Forest* e *Extra-Trees*, utilizam técnicas de perturbação e combinação projetadas especificamente para árvores de decisão, ou seja, um conjunto diversificado de classificadores é criado por aleatoriedade, sendo a previsão do conjunto dada como a previsão média dos classificadores individuais. Para *Random Forest* (floresta aleatória), cada árvore no conjunto é construída a partir de uma amostra retirada com substituição (*bootstrap*) do conjunto de treinamento e a construção de cada nó utiliza a melhor divisão utilizando todos os recursos de entrada ou um subconjunto aleatório. Como consequência, as florestas aleatórias têm variação reduzida e, às vezes, um viés um pouco maior. Para *Extra-Trees* (árvores extremamente aleatórias), há ainda mais aleatoriedade no método ao se adicionar

limites aleatórios na construção dos nós, reduzindo a variância do modelo e aumentando o viés.

Estimadores tipo *AdaBoost* utilizam modelos fracos (ligeiramente melhores que suposições aleatórias) em versões modificadas dos dados, sendo estes ponderados por votos ou soma para produzir uma versão final. As modificações nos dados de cada iteração de reforço são realizadas pela aplicação de pesos para cada uma das amostras de treinamento. Os pesos são inicialmente definidos e, no processo de iteração, são modificados individualmente. Para aqueles em que a predição foi incorreta por conta do peso induzido, o peso é aumentado, enquanto para aqueles preditos corretamente, o peso é diminuído, sendo que os exemplos difíceis de prever recebem influencia cada vez maior até que se ajustem.

Estimadores do tipo *Gradient Tree Boosting* são uma generalização do boost para funções de perda diferenciáveis arbitrárias, ou seja, as árvores de decisão são incrementadas por um gradiente. O *Gradient Boosting Regressor* pode utilizar tamanhos de amostras pequenos e suporta várias funções de perda diferentes, sendo por default os mínimos quadrados. O *Hist Gradient Boosting Regressor* é uma árvore de aumento de gradiente, baseado em histograma, que pode ser mais rápido quando o número de amostras é maior que dez mil amostras e tem suporte integrado para valores ausentes. São também árvores de decisão com gradientes os modelos *XGB (Extreme Gradient Boosting)*, desenvolvido em 2016 com objetivo de acelerar a execução e melhorar a performance do modelo, e *LGBM (Light Gradient Boosted Machine)*, que aumenta o gradiente adicionando uma seleção automática e foca em aumento de gradientes maiores, acelerando o treinamento e melhorando o desempenho preditivo.

2.4.1.5 Modelo Linear

Os modelos lineares de regressão são aqueles em que o valor alvo (dado de saída) é expresso pela combinação linear dos recursos (dados de entrada), sendo que o procedimento de ajuste do modelo é minimizar uma função de erro, que compara os valores reais e preditos.

O modelo *Linear Regression* (regressão linear) ajusta um modelo linear de forma que a soma residual dos quadrados entre os alvos observados e previstos seja mínima.

O estimador *Ridge*, por sua vez, utiliza os mínimos quadrados ordinários, penalizando o tamanho dos coeficientes, para minimizar a soma residual dos quadrados ao utilizar um parâmetro adicional que busca robustez e colinearidade. Utiliza o termo de regularização L2 na função erro para evitar o *overfitting* (vício do modelo em uma amostra específica, fazendo o modelo ser incapaz de explicar outros dados). O estimador *Ridge CV* adiciona a validação cruzada ao embutir um novo parâmetro, utilizando a técnica de separação entre dados de treino e teste.

O estimador *Lasso* estima coeficiente esparsos, tendo a vantagem, a depender do problema, de tender a preferir soluções com menos coeficientes diferentes de zero, ou seja, reduzir o número de recursos da qual a solução é dependente. É similar matematicamente ao *Ridge*, adicionado um termo de regularização, que resolve a penalização dos mínimos quadrados, o L1. O estimador *Lasso CV* é equivalente ao *Ridge CV* em termos de validação cruzada.

O regressor Elastic Net é um modelo de regressão que utiliza as regularizações L1 e L2, sendo útil quando há vários recursos correlacionados entre si, mas com a dificuldade de determinar dois hiperparâmetros. O regressor *Elastic Net CV* utiliza a validação cruzada neste mesmo modelo.

O regressor *LARS (Least Angle Regression)* é um algoritmo de regressão de ângulo mínimo para dados de alta dimensão. O algoritmo, em cada etapa da regressão, encontra o recurso mais relacionado ao destino e, quando encontra correlação igual em vários recursos, ao invés de continuar ao longo do mesmo recurso, prossegue em uma direção equiangular entre os recursos. Sua aplicação é útil quando utilizada a validação cruzada, como *Lars CV*, ou outros tipos de solução por partes, apesar de ser sensível a efeitos de ruídos. É especialmente eficiente quando o número de recursos é significativamente maior que o número de amostras, sendo mais rápido quanto a seleção direta e tem a mesma complexidade dos mínimos quadrados comuns.

O estimador *Lasso LARS* é um modelo de laço implementado usando o algoritmo LARS e produz uma solução exata, linear por partes em função de seus coeficientes. O estimador *Lasso Lars CV* é também baseado no algoritmo LARS, que tem a vantagem de explorar valores mais relevantes do parâmetro alfa, sendo mais rápido que *Lasso CV* se o número de amostras for muito pequeno em comparação com o número de recursos e utiliza a validação cruzada.

Um estimador alternativo é o *Lasso LARS IC*, que usa o critério de informação de Akaike (AIC) e critério de informação de Bayes (BIC), uma forma de encontrar o valor ideal do parâmetro alfa alternativamente à validação cruzada, exigindo menos computacionalmente. Os critérios, entretanto, precisam de uma estimativa adequada dos graus de liberdade da solução e não é recomendável que tenham mais recursos que amostras.

Outro tipo de otimizador é o *Orthogonal Matching Pursuit* (OMP), que aproxima o ajuste de um modelo linear com restrições impostas ao número de coeficientes diferentes de 0, sendo um método de seleção de recurso avançado. O modelo busca a correspondência ortogonal, aproximando o vetor de solução ideal com o número fixo de elementos diferentes de zero, tendo como alvo um erro específico em vez de um número específico de coeficientes diferentes de zero. É considerado um algoritmo “guloso” por incluir, a cada etapa, o átomo mais altamente correlacionada com o residual atual, melhorando em cada iteração. O residual é recalculado usando a projeção ortogonal no espaço dos elementos previamente escolhidos. O *Orthogonal Matching Pursuit CV*, como indicado anteriormente, utiliza a validação cruzada.

A regressão Bayesiana inclui parâmetros de regularização no procedimento de estimativa, em que estes são ajustado aos dados disponíveis pela introdução de antecedentes não informativos sobre os hiperparâmetros do modelo. Ou seja, a saída do modelo é considerada como uma distribuição gaussiana para se obter um modelo totalmente probabilístico, em que há uma variável aleatória. Este tipo de abordagem se adapta aos dados disponíveis e pode ser usado para incluir parâmetros de regularização na estimativa. Entretanto, a inferência do modelo pode ser demorada. O regressor *Bayesian Ridge Regression* estima desta forma.

Os modelos lineares generalizados (GLM) entendem os modelos lineares de duas formas. Primeiramente, os valores previstos são ligados a uma combinação linear das variáveis de entrada por uma função de ligação inversa. Depois, a função de perda quadrada é substituída pelo desvio de unidade de uma distribuição na família exponencial, tornando uma média ponderada quando os pesos da amostra são fornecidos. Os modelos de distribuição exponencial podem ser de distribuição normal, Poisson, Gamma ou Gaussiana inversa. A escolha da distribuição depende se os valores alvo forem contagens, frequências relativas, positivos e distorcidos, ou pela análise

da cauda de distribuição. Na biblioteca Lazy é utilizada a regressão Linear Generalizada (*Generalized Linear Regression*) ou a regressão *Tweedie Regressor*, que permite modelar qualquer uma das distribuições usando o parâmetro de potência apropriado. As distribuições Poisson e Gamma podem ser obtidas pelo regressor *Tweedie* mas também por seus próprios regressores, *Poisson Regressor* e *Gamma Regressor*, respectivamente.

A regressão *SGD* (gradiente descendente estocástico) é utilizado quando o número de amostras é grande, sendo permitido o aprendizado online/out-of-core e utilizando diferentes funções de perda (convexa) e diferentes penalidades.

Por fim, os algoritmos passivo-agressivos, modelados pelo regressor *Passive-Agressive Regressor*, são algoritmos para aprendizagem em larga escala, sem necessidade de taxa de aprendizado, mas com um parâmetro de regularização.

Os modelos de regressão robusta ajustam um modelo de regressão na presença de dados corrompidos, ou seja, outliers ou erros no modelo. Na biblioteca do LazyPredict, o regressor *RANSAC* ajusta um modelo de subconjuntos aleatórios para considerar como inlier (dado discrepante dos demais, mas que se encontra no intervalo da distribuição aceitável dos valores medidos corretamente, não sendo considerado um “outlier”) a partir do conjunto completo e produz um resultado razoável com certa probabilidade, esta dependente do número de iterações. Já o regressor *Huber* aplica uma perda linear às amostras classificadas como outlier, sendo essa classificação determinada por um limite, e sua diferença é de atribuir um peso menor aos outlier ao invés de os ignorar.

2.4.1.6 Vizinhança

São métodos de aprendizagem baseados na vizinhança mais próxima, encontrando um número predefinido de amostras de treinamento mais próximas em distância do novo ponto e prever a partir deles. O número de amostras pode ser constante ou variar com base na densidade local de pontos e a distância pode ser qualquer medida métrica. São métodos não generalizantes e não paramétricos.

Os regressores baseados em vizinhança podem ser usados nos casos em que as variáveis são contínuas em vez de variáveis, sendo calculada pela média dos vizinhos mais próximos. O regressor *KNeighbors Regressor* aprende com base nos

vizinhos mais próximos a cada ponto, em que um valor inteiro é especificado e os pesos dados aos pontos são uniformes.

2.4.1.7 Estimadores Compostos

Os estimadores compostos são transformadores, sendo que o regressor *Transformed Target Regressor* transforma os alvos (dado de saída) antes de ajustar um modelo de regressão (modelo linear, por exemplo). As previsões são mapeadas de volta ao espaço original por meio de uma transformação inversa.

2.4.1.8 Processo Gaussiano

Os processos Gaussianos (GP) são métodos genéricos de aprendizagem para problemas de regressão probabilística, sendo vantajosos por interpolarem as observações na previsão, terem predição probabilística para ajustar ao calcular intervalos de confiança empíricos e são versáteis por poderem especificar diferentes kernels. Entretanto, os modelos não são esparsos, usando todas as informações das amostras para realizar as previsões, além de perderem eficiência em espaços dimensionais elevados, quando há muitos recursos. O regressor *Gaussian Process Regressor* utiliza uma média, normalmente igual a zero ou como a média entre os dados de treinamento, e uma covariância, pelas funções de Kernel.

2.4.1.9 Estimadores Fictícios

Os estimadores fictícios (*dummy estimators*) são úteis para obter um valor de linhas base das métricas de avaliação da qualidade das previsões de um modelo em previsões aleatórias. Tais avaliações podem ser baseadas nos métodos de pontuação do estimador, em um parâmetro de pontuação ou em funções métricas.

O regressor *Dummy Regressor* compara um estimador com regras práticas simples, implementando estratégias e ignorando os dados de entrada. As regras podem ser de previsão da média, da mediana, do quantil dos alvos de treinamento ou uma constante.

2.4.1.10 Crista de Kurlmel

A regressão *Kernel Ridge* combina a regressão *Ridge* com o truque de kernel, aprendendo uma função linear no espaço induzido pelo respectivo kernel e os dados.

Para kernels não lineares, é correspondente a uma função não linear no espaço original. A forma de aprendizagem é igual à de SVR, mas a função de perda utilizada é a de erro quadrática. O ajuste pode ser feito em formato fechado e o modelo não é esparso.

A biblioteca Lazy Predict, portanto, modela 42 regressores diferentes para cada conjunto de dados de entrada e teste, conforme Tabela 1:

TABELA 1 - MODELOS REGRESSIVOS DA BIBLIOTECA LAZY PREDICT

Forma de Aprendizado	Modelo
Rede Neural Artificial	MLP Regressor
Support Vector Machine	SVR
	NuSVR
	Linear SVR
Conjunto de Modelos	Bagging Regressor
	Random Forest Regressor
	XGB Regressor
	Gradient Boosting Regressor
	Extra Trees Regressor
	Hist Gradient Boosting Regressor
	LGBM Regressor
Árvore de Decisão	Decision Tree Regressor
	Extra Tree Regressor
Vizinhança	Kneighbors Regressor
Estimadores Compostos	Transformed Target Regressor
Processo Gaussiano	Gaussian Process Regressor
Estimador Fictício	Dummy Regressor
Crista de Kernel	Kernel Ridge
Modelo Linear	Poisson Regressor
	Huber Regressor
	Gamma Regressor
	Ridge
	Bayesian Ridge
	Ridge CV
	Linear Regression
	Lasso
Lasso CV	

	Lasso Lars IC
	Lasso Lars CV
	Lars
	Lars CV
	Lasso Lars
	Tweedie Regressor
	Generalized Linear Regressor
	Elastic Net
	Elastic Net CV
	RANSAC Regressor
	SGD Regressor
	Passive Agressive Regressor
	Orthogonal Matching Pursuit
	Orthogonal Matching Pursuit CV

Fonte: A Autora (2021)

3 MATERIAIS E MÉTODOS

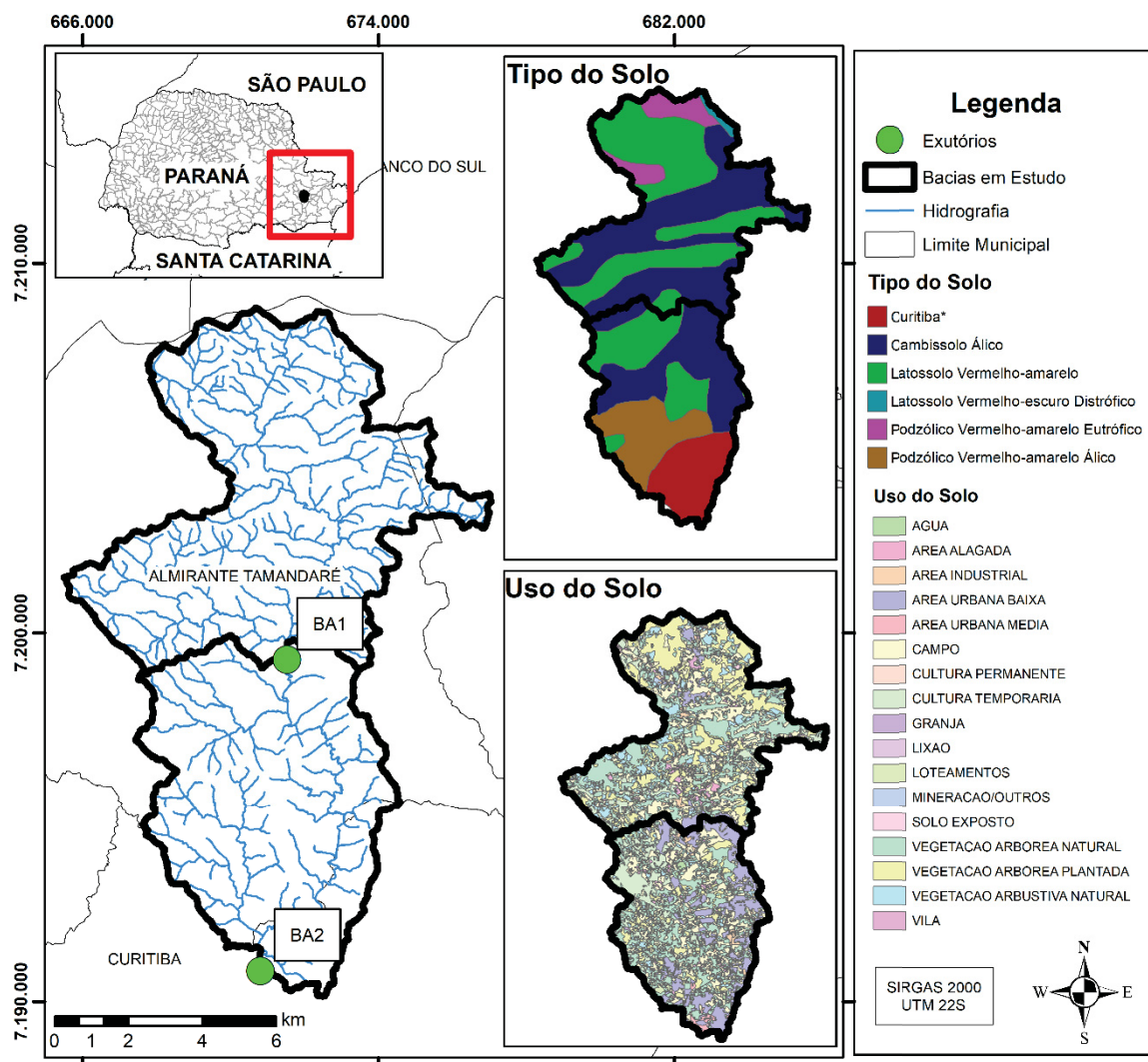
Para verificar a hipótese formulada de que existe uma relação matemática entre a concentração de poluente no rio durante um evento de chuva e as características da bacia e do próprio evento, foram utilizadas técnicas de inteligência artificial, conforme descrito no item 3.2. O estudo de caso foi conduzido na bacia do Rio Barigui, descrito no item 3.1, devido a disponibilidade de dados de eventos de chuva, onde amostras foram tomadas durante todo o evento, ou seja, na ascensão e na recessão do hidrograma de cheia, durante intervalos definidos pela diferença de nível do rio.

3.1 ESTUDO DE CASO

A Bacia do Rio Barigui, localizada no estado do Paraná, na bacia do Alto Iguaçu, foi utilizada como estudo de caso por conta dos dados coletados e disponibilizado sem Kozak (2016, 2020), referentes a concentrações de diversos parâmetros de qualidade da água coletados durante eventos de chuva através de amostrador automático.

A bacia hidrográfica do Rio Barigui é uma sub-bacia localizada na Bacia do Alto Iguaçu no Paraná nas cidades de Almirante Tamandaré, Curitiba e Araucária, com cerca de 267 km² de área de drenagem e 67 km de extensão. Para o estudo foram avaliados dois pontos de monitoramento, BA1 e BA2, conforme Figura 8. O ponto BA1, localizado nas coordenadas UTM 671556E e 7199322N é o exutório da sub-bacia hidrográfica com área de drenagem de 58 km² aproximadamente e comprimento do rio de 14,8km. Já o ponto BA2, a jusante do ponto BA1, se encontra nas coordenadas UTM 670683E e 7191046N com área de drenagem de cerca de 104 km² e comprimento do rio de 32,8 km.

FIGURA 8 - BACIA DO RIO BARIGUI



Fonte: A Autora (2021)

A classificação do uso da terra pelo IAT (Instituto Água e Terra) para o ano de referência de 2012 nesta bacia contempla dezessete usos, em dados conforme indicado na Figura 8.

Quanto ao tipo de solo, a bacia se divide em seis tipos pela classificação da EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária): Latossolo vermelho-amarelo, Cambissolo Álico, Podzólico Vermelho-amarelo eutrófico, Latossolo Vermelho-escuro Distrófico, Podzólico Vermelho-amarelo Álico e uma parcela urbana, denominada “Curitiba” nos dados, conforme Figura 8.

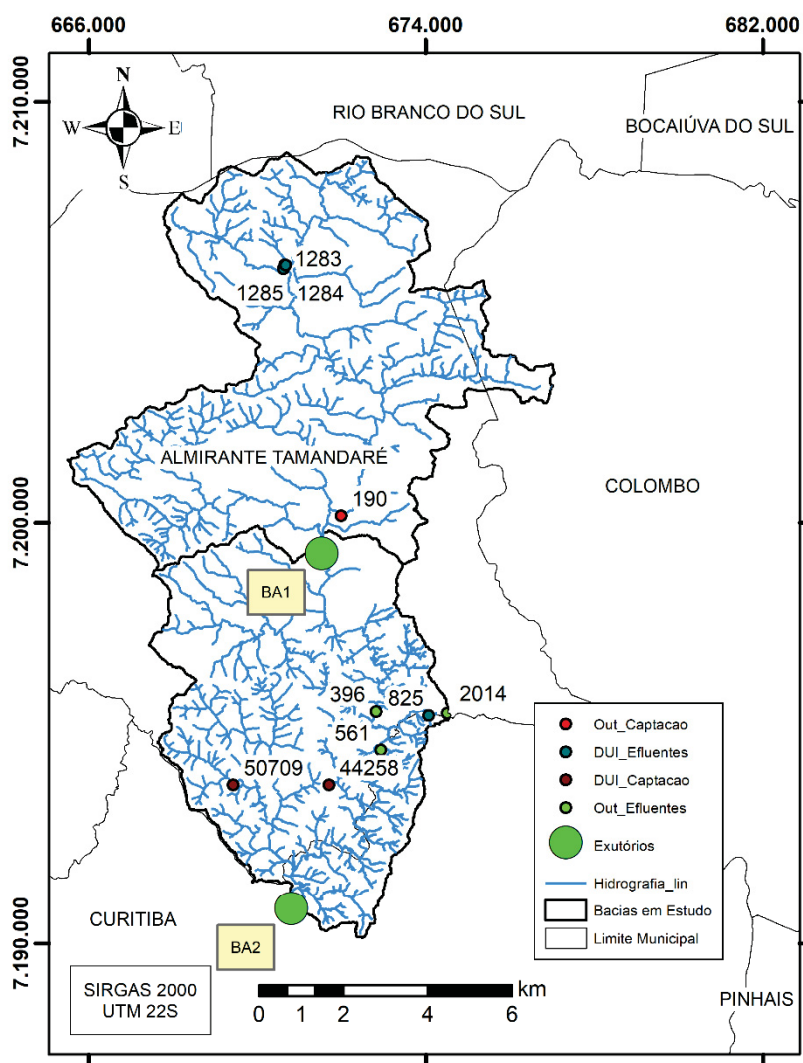
Quanto às fontes pontuais de poluição, a Figura 9 indica as outorgas de captação (em verde), dispensa de captação (em vermelho escuro), lançamento de

efluentes (em verde) e dispensa de lançamento nas bacias (em vermelho), sendo descritas na Tabela 2, conforme dados disponibilizados pelo IAT em site público.

A montante da bacia BA1 são outorgadas apenas captações de água, que impactam no regime hidrológico na retirada da água, sendo a captação da empresa de saneamento SANEPAR, de vazão 527,6 m³/h, a mais expressiva na bacia. Os impactos da poluição pontual nesta bacia, portanto, não devem ser relevantes.

Na bacia BA2, estão presentes outras outorgas que impactam na qualidade e quantidade de água. Entre BA1 e BA2 existe uma ETE com outorga de 418,5 m³/h no rio Barigui. Esta ETE possui 85% de eficiência e seu tratamento é feito por um reator anaeróbico UASB seguido de processos físico-químicos, com DBO máxima outorgada de 60 mg/L. Ainda está presente a indústria AMBEV, que tem seu ponto de lançamento logo a jusante da ETE no rio Barigui, podendo lançar até 120 m³/h com DBO máxima de 50 mg/L, além da indústria metalúrgica GUSSO, com 6m³/h.

FIGURA 9 - OUTORGAS NA ÁREA DE ESTUDO



Fonte: A Autora (2021)

TABELA 2 - OUTORGAS NAS BACIAS DE ESTUDO (FONTES PONTUAIS DE POLUIÇÃO)

Outorga	Bacia	Código PON	Solicitante	Uso	Rio	Vazão (m³/h)	DBO (mg/L)
Captação	BA1 BA2	190	COMPANHIA DE SANEAMENTO DO PARANÁ	Saneamento	Rio Barigui	527,6	-
Dispensa captação	BA2	44258	KAPAZI INDÚSTRIA E COMERCIO DE CAPACHOS LTDA	Indústria	-	1,8	-
	BA2	50709	VALE DO SOL ARGAMASSAS E LOCAÇÃO DE CAÇAMBAS – ME	Indústria	Córrego sem nome	0	-
Lançamento Efluentes	BA2	2014	METALURGICA GUSSO LTDA	Indústria	Córrego Sem Nome	6	30
	BA2	396	COMPANHIA DE SANEAMENTO DO PARANÁ	Saneamento	Rio Barigui	418,5	60
	BA2	561	AMBEV S.A.	Indústria	Rio Barigui	120	50
Dispensa Lançamentos	BA1 BA2	1285	ÁGUA MINERAL TIMBÚ LTDA – EPP	Indústria	Córrego sem nome	0	0
	BA1 BA2	1283	ÁGUA MINERAL TIMBÚ LTDA – EPP	Indústria	Córrego sem nome	0	0
	BA1 BA2	1284	ÁGUA MINERAL TIMBÚ LTDA – EPP	Indústria	Córrego sem nome	0	0
	BA2	825	NATUPHITUS INDÚSTRIA E COMÉRCIO DE COSMÉTICOS LTDA	Indústria	Córrego Sem Nome	0	0

Fonte: A Autora (2021)

As amostras de concentração obtidas por Kozak (2016, 2020) foram coletadas pelo amostrador automático SBN, desenvolvido por Braga (2013), com o intuito de entender o processo da poluição difusa e seus efeitos no corpo hídrico. O amostrador funciona por amostragem de intervalo de fluxo medindo a contribuição do evento de precipitação no corpo hídrico, sendo considerado um amostrador inteligente por detectar o início do evento pela variação crescente de nível do rio, coletando amostras representativas para o objetivo que foi desenvolvido. Ainda, é importante ressaltar que a coleta das amostras passa por um filtro, que pode ser um modificador nas concentrações de alguns poluentes. O amostrador é capaz de armazenar 24 frascos de amostras de capacidade de 1L, sendo esta uma das principais limitações do aparelho, assim como a preservação das amostras e os parâmetros de qualidade da água dos quais é possível analisar que sejam significativos para a poluição difusa.

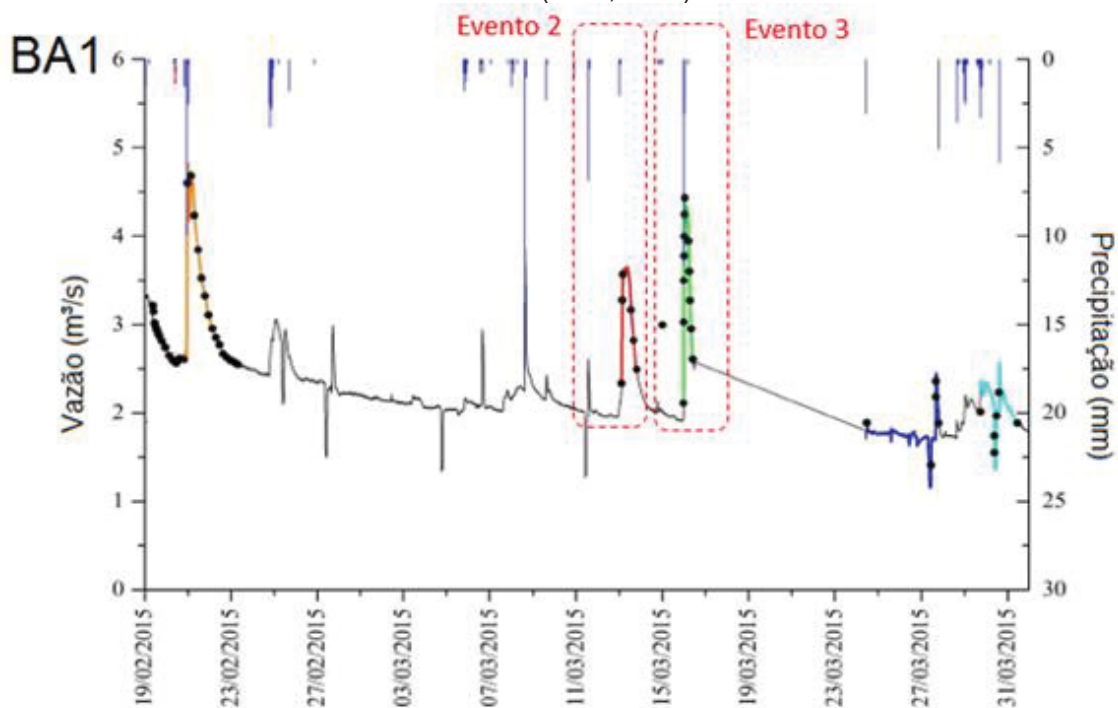
Para este estudo, foram analisadas apenas as amostras de fósforo total, nitrogênio total e carbono orgânico dissolvido, ainda que tenham sido analisados diversos outros parâmetros por Kozak (2016, 2020).

A amostragem da água foi realizada pelo amostrador automático ISCO, comandado por H50-XL Logger programado para funcionar no amostrador SBN desenvolvido, com adaptações por ainda estar em desenvolvimento, conforme indicado por Kozak (2016, 2020) e desenvolvido por Braga (2013). A amostragem é iniciada quando há uma variação mínima no nível do rio de 5cm em no máximo 10 minutos, sendo que o intervalo de leitura do equipamento é de 2,5 minutos. A próxima amostra é coletada caso a diferença de nível, seja crescente ou decrescente, seja superior a 5cm no intervalo de 10 minutos.

Os pontos BA1 e BA2, onde foram coletadas as amostras, foram escolhidos por terem estações pluviométricas de monitoramento automático instaladas em local próximo, sendo estas consideradas como estação no ponto de monitoramento neste estudo. Esta pode ser uma limitação de informação, uma vez que a informação nem sempre tem representatividade espacial, dado que a chuva não é homogênea na área de drenagem de uma bacia.

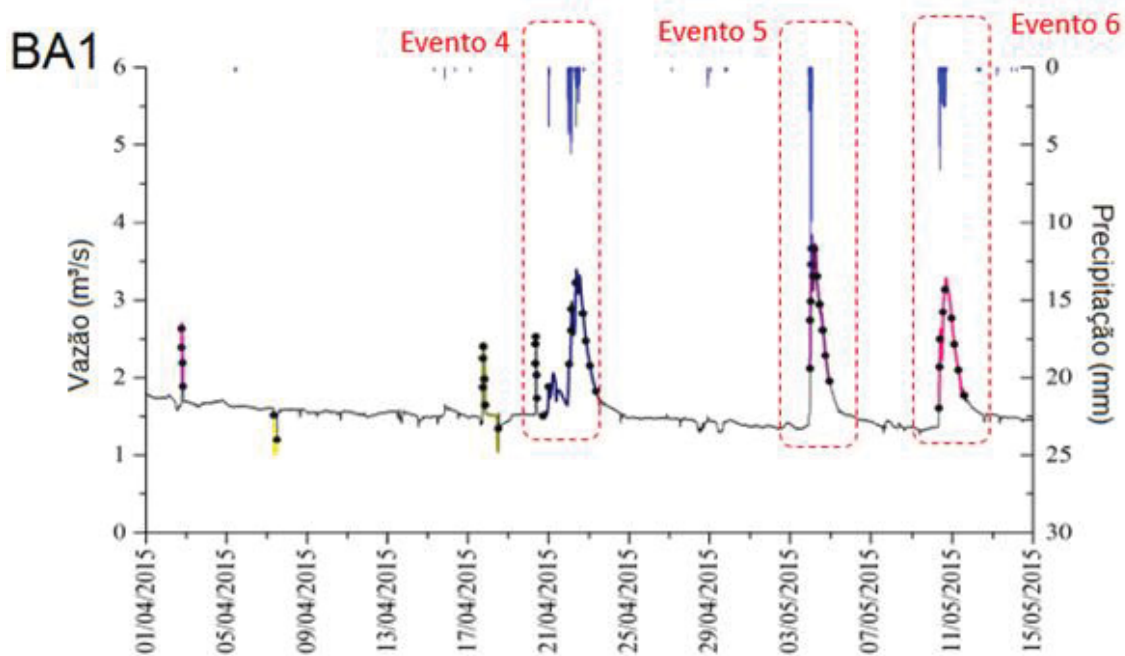
Para este estudo, foram considerados 16 eventos (campanhas de coleta). As amostras dos eventos 1, 15 e 16 foram coletadas em dias sem chuva, chamados "baselines". Os outros 13 eventos estão indicados nas Figura 10, Figura 11, Figura 12 e Figura 13, em que os pontos indicam o momento da coleta das amostras. Ao todo, foram obtidas informações de 191 amostras com dados concomitantes dos três poluentes, utilizadas neste estudo a fim de entender os fatores de influência das concentrações dos poluentes durante os eventos de precipitação, uma vez que Kozak (2016) indica que precipitações similares podem resultar em diferentes valores de concentração média do evento (EMC) devido as suas diferentes características e à resposta do corpo hídrico. Além disso, como base para o desenvolvimento da metodologia deste estudo, considerou-se a análise de que "a diluição de poluição ou efeitos de concentração dependem dos dias secos antecedentes, intensidade da chuva, volume transportado, uso e ocupação da bacia, entre outros" (KOZAK, 2016).

FIGURA 10 - EVENTOS DE PRECIPITAÇÃO UTILIZADOS NESTE ESTUDO COM AMOSTRAS COLETADAS POR KOZAK (2016, 2020) - EVENTOS 2 E 3



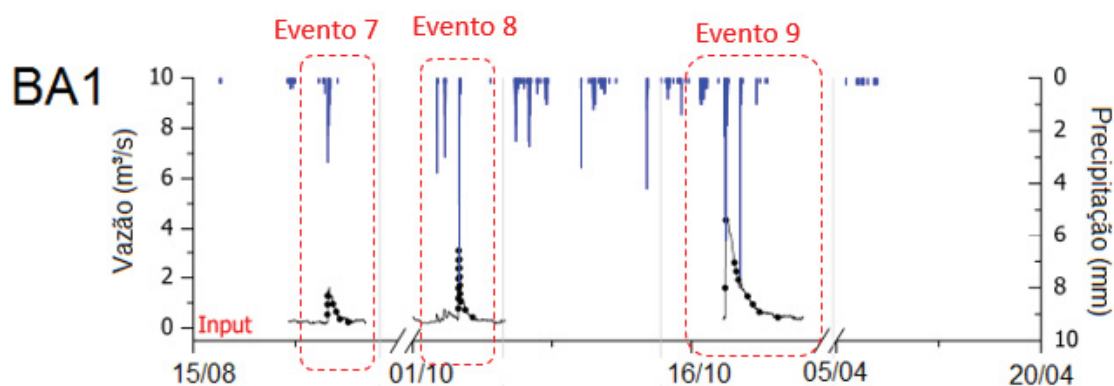
Fonte: Adaptado de KOZAK (2016, 2020)

FIGURA 11 - EVENTOS DE PRECIPITAÇÃO UTILIZADOS NESTE ESTUDO COM AMOSTRAS COLETADAS POR KOZAK (2016, 2020) - EVENTOS 4, 5 E 6



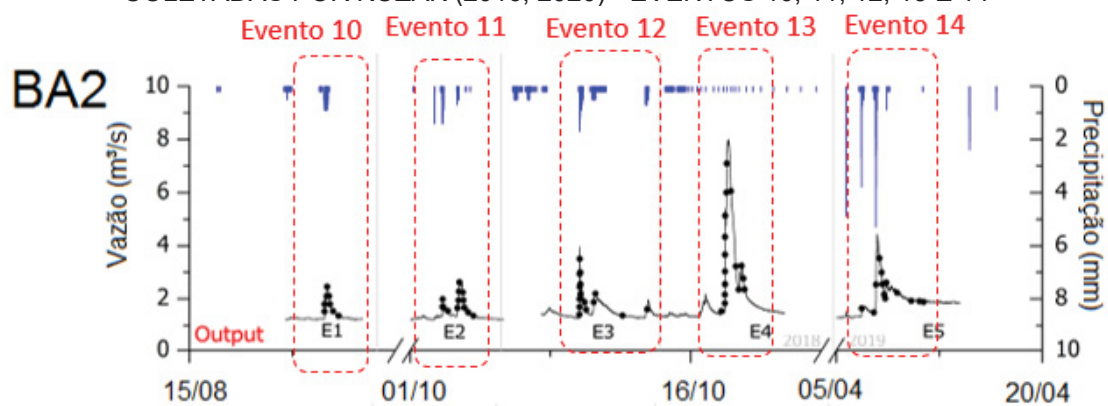
Fonte: Adaptado de KOZAK (2016, 2020)

FIGURA 12 - EVENTOS DE PRECIPITAÇÃO UTILIZADOS NESTE ESTUDO COM AMOSTRAS COLETADAS POR KOZAK (2016, 2020) - EVENTOS 7, 8 E 9



Fonte: Adaptado de KOZAK (2016, 2020)

FIGURA 13 - EVENTOS DE PRECIPITAÇÃO UTILIZADOS NESTE ESTUDO COM AMOSTRAS COLETADAS POR KOZAK (2016, 2020) - EVENTOS 10, 11, 12, 13 E 14



Fonte: Adaptado de KOZAK (2016, 2020)

3.2 MÉTODO

A metodologia empregada neste estudo para verificar a hipótese formulada de que existe uma relação matemática entre a concentração de poluente que é carregado ao rio durante um evento de chuva e as características da bacia hidrográfica e do próprio evento de chuva, abordando os aspectos do ciclo hidrológico seguiu os seguintes passos:

- (i) levantamento dos dados disponíveis de concentração no rio;
- (ii) definição das variáveis referentes à bacia nos eventos monitorados, conforme o ciclo hidrológico;
- (iii) desenvolvimento de algoritmo para aplicação das técnicas de inteligência artificial;
- (iv) treinamento e validação dos modelos de IA.

3.2.1 Levantamento de dados disponíveis de concentração no rio em eventos de chuva

Conforme discutido na introdução, os dados de monitoramento convencional disponíveis nos bancos de dados institucionais são discretos e escassos, não sendo possível uma representatividade da poluição difusa de eventos de chuva.

Desta forma, esta pesquisa foi conduzida na bacia do rio Barigui, utilizando-se dos dados de concentração por Kozak (2016, 2020), que realizou coletas durante eventos de precipitação.

3.2.2 Definição de variáveis referentes à bacia e aos eventos monitorados

Tendo em vista que os dados obtidos por Kozak (2016, 2020) são referentes a concentração de diferentes poluentes no rio durante eventos de chuva, eles caracterizam as concentrações e cargas provenientes da poluição difusa, mais especificamente dos poluentes carregados pelo escoamento superficial e pela erosão, além da eventual carga pontual que esteja ocorrendo no período. A partir desta consideração, foram analisadas quais são as informações mais importantes para gerar um modelo de previsão de concentração de poluentes durante um evento de precipitação.

Com base no ciclo hidrológico e nos fatores intervenientes que definem a distribuição temporal e espacial da vazão e concentração no rio, foi elaborada a base de dados da bacia hidrográfica em estudo, com dados de precipitação de postos pluviométricos na bacia nos períodos de estudo, características fisiográficas, informações georreferenciadas do tipo e uso da terra e evapotranspiração média, para o período dos dados de qualidade da água nos eventos de precipitação coletados e analisados por Kozak (2016, 2020).

A partir dessa base de dados, foram definidos os dados de entrada a serem analisados pelas técnicas de inteligência artificial, os quais foram considerados como potenciais informações relevantes para a formação de um modelo de previsão de concentração. Estes dados buscam refletir todas as etapas do ciclo hidrológico, entendendo que a concentração é função de como a água se comporta no tempo e no espaço, como apresentado na Figura 5. Assim, os dados estudados totalizam 69 variáveis e são descritos nos subitens a seguir:

3.2.2.1 Ponto de coleta

Indicativo do local (variável espacial) onde foram realizadas as coletas, podendo ser BA1 ou BA2.

3.2.2.2 Evento

Indicativo do evento (variável temporal) em que foram coletas as amostras, variando de 1 a 16.

3.2.2.3 Tempo de concentração (minutos)

Tempo (variável temporal), em minutos, desde o início da chuva detectada na estação pluviométrica próxima ao ponto de coleta e o momento da coleta, limitado ao tempo total da bacia de contribuição do ponto, sendo de 193 minutos para BA1 e 364 minutos para BA2. Tais tempos de concentração máximos foram calculados pela equação 2 de Kirpich, descrita a seguir, que considera o comprimento do talvegue (L, em km) e a declividade do talvegue (S, em m/m) para calcular o tempo de concentração (T_c , em minutos):

$$T_c = 3,898L^{0,77}S^{-0,385} \quad (2)$$

A partir do tempo de concentração calculado para cada bacia e do modelo digital de elevação (MDE) com resolução 30x30, foi utilizada a ferramenta computacional Watershed Isochrones desenvolvida no Laboratório de Simulação Hidrológica e Processamento de Dados, veiculado ao curso de graduação em Engenharia Hídrica e do Programa de Pós-Graduação em Recursos Hídricos da Universidade Federal de Pelotas (UFPEL), para obter as isócronas das bacias, ou seja, as curvas de mesmo tempo de concentração da bacia, ferramenta também utilizada por Luz et al. (2014).

3.2.2.4 Tempo desde a chuva inicial (minutos)

Tempo (variável temporal), em minutos, desde o intervalo de tempo anterior ao primeiro registro de precipitação da estação pluviométrica localizada próximo ao ponto e o momento da coleta da amostra.

3.2.2.5 Concentração de NT, PT e COD (mg/L)

Concentração dos poluentes amostrados por Kozak (2016, 2020).

3.2.2.6 Precipitação Acumulada – dentro do tempo de concentração (mm)

Quantidade de precipitação acumulada entre o momento da coleta da amostra e o tempo de concentração anteriormente ao momento da coleta, referente à estação pluviométrica mais próxima ao ponto de coleta. Esta variável busca apresentar a quantidade de água que pode ter contribuído ao escoamento superficial e, portanto, ao carreamento do poluente que produziu a contribuição da concentração do poluente no rio. Pode ser entendido também, para quando o tempo de concentração for menor que o máximo da bacia, como a máxima precipitação a carrear a carga poluidora e alterar a concentração do rio entre o momento anterior à chuva e o momento da coleta, ou seja, a contribuição da poluição difusa. Para os tempos de concentração máximos ou superiores ao máximo, pode ser entendida como o carreamento da poluição difusa daquela parcela de tempo.

3.2.2.7 Precipitação Acumulada – desde chuva inicial (mm)

Quantidade de precipitação acumulada desde o início da chuva e o momento da coleta da amostra, indicando a máxima precipitação possível a gerar escoamento superficial e carrear a poluição difusa, referente à estação pluviométrica mais próxima ao ponto de coleta.

3.2.2.8 Precipitação Antecedente Acumulada de 1, 3, 5 e 7 dias (mm)

Quantidade de precipitação acumulada entre o momento da coleta da amostra e exatamente 1, 3, 5 ou 7 dias anteriores, refletindo uma das variáveis para entender o estado de saturação do solo, que define a separação do escoamento subterrâneo e superficial, referente à estação pluviométrica mais próxima ao ponto de coleta.

3.2.2.9 Vazão no rio (m³/s)

Vazão na estação fluviométrica próxima ao ponto de coleta, obtida pela leitura da régua da estação e transformada em vazão pela curva-chave da estação, dado já processado pelo IAT (Instituto Água e Terra).

3.2.2.10 Condição do hidrograma (ascendente ou descendente)

Indicador binário do comportamento do hidrograma e, portanto, do rio. Se está subindo, indicador 2, se está descendo (recessão), indicador 1. Tem função de indicar se o rio está subindo ou descendo, uma vez que o período de recessão é caracterizado por um longo e lento decréscimo de vazão até que o momento em que o escoamento de base é atingido novamente (DAVIE, 2008). Além disso, é na ascendência o momento em que o escoamento superficial pode estar trazendo maior quantidade de material da bacia, principalmente pelo efeito first-flush, em que, após longos períodos sem chuva, a poluição se acumula e é mobilizada pelo início da próxima precipitação (BRAGA, 2013).

3.2.2.11 Evapotranspiração real (mm/dia)

Quantidade de água média evapotranspirada na bacia. A informação é da Agência Nacional de Águas, disponibilizada pelo aplicativo SSEBop BR (ANA, 2020c), que calcula a evapotranspiração média diária no Brasil com base em sensoriamento remoto e aplicação do balanço de energia (por dados de temperatura do ar).

3.2.2.12 Área da bacia (km²)

Área da bacia (variável espacial), em quilômetros quadrados, referente à bacia de drenagem no tempo de concentração, definida pelas isócronas, para cada amostra.

3.2.2.13 Declividade média (m/m)

Declividade média (variável espacial), em metros/metros, da bacia de drenagem no tempo de concentração para cada amostra, que pode interferir diretamente na quantidade de poluente carregado pelo escoamento superficial.

3.2.2.14 Tipo de solo (%)

Tipo de solo, em porcentagem na bacia. Dado disponibilizado pela EMBRAPA, sendo dois tipos de Latossolo (vermelho-amarelo e vermelho-escuro distrófico), um Cambissolo (álico), um Podzólico (vermelho-amarelo álico) e uma parcela de área sem informação de tipologia de solo, localizada na parte altamente adensada de Curitiba e, provavelmente, muito impermeável, denominada “Curitiba”. Variável

diretamente associada com a separação do escoamento entre subterrâneo e superficial.

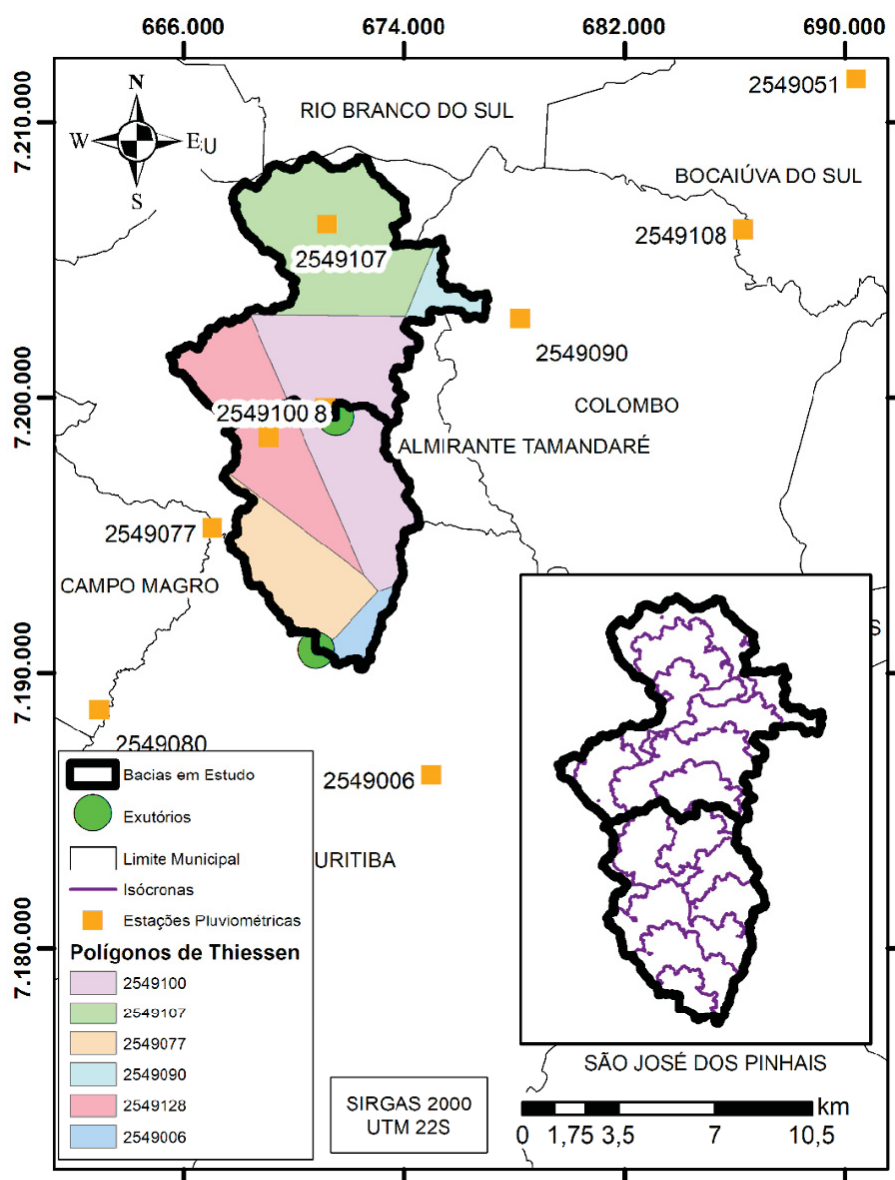
3.2.2.15 Uso da Terra (%)

Uso da Terra, em porcentagem na bacia, conforme dado disponibilizado pelo Instituto Água e Terra (IAT). Variável diretamente associada com a separação do escoamento entre subterrâneo e superficial, além de intercepções.

3.2.2.16 Precipitação Acumulada Antecedente de 1, 3, 5 e 7 dias (mm) – Thiessen

Quantidade de precipitação acumulada entre o momento da coleta da amostra e exatamente 1, 3, 5 ou 7 dias anteriores, média na bacia calculada pelo método de Thiessen com dados das estações fluviométricas (2 dados diários), das estações 2549100, 2549107, 2549128, 2549077, 2549090 e 2549006. Para a formação dos polígonos de Thiessen, entretanto, foram utilizadas mais três estações que não tiveram influência dentro das bacias de estudo, sendo elas: 2549080, 2549051 e 2549108.

FIGURA 14 - ESTAÇÕES PLUVIOMÉTRICAS, POLÍGONOS DE THIESSEN E ISÓCRONAS NAS BACIAS EM ESTUDO



Fonte: A Autora (2021)

3.2.2.17 Precipitação Acumulada Antecedente de 1, 3, 5 e 7 dias (mm) da estação fluviométrica (utilizada para Thiessen) - 2549100, 2549107, 2549128, 2549077, 2549090 e 2549006.

Quantidade de precipitação acumulada entre o momento da coleta da amostra e exatamente 1, 3, 5 ou 7 dias anteriores, referente a cada uma das estações fluviométricas.

3.2.3 Desenvolvimento de algoritmo para aplicação das técnicas de inteligência artificial

Para buscar variáveis (descritas no item 3.2.2) que apresentaram melhor relação com os dados de concentração (item 3.2.1), foi desenvolvido um algoritmo para automatizar a busca, uma vez que o número de simulações seria extenso e de impossível realização manual. O algoritmo foi escrito em linguagem Python e utilizou a biblioteca LazyPredict, que possui 42 diferentes regressores.

O algoritmo faz combinações entre as variáveis definidas no item anterior, em que cada variável tem um número correspondente e o programa escolhe números aleatórios dentre os possíveis para cada combinação. Desta forma, o algoritmo busca os melhores modelos que as represente juntamente com a concentração da substância no rio através de vários regressores, quando comparados os critérios de busca.

Com base nesses resultados, as variáveis mais significativas (que fazem parte das melhores relações apontadas pelos regressores, serão as variáveis que comporão o modelo.

3.2.4 Treinamento e validação dos modelos de inteligência artificial

3.2.4.1 Dados de entrada

Os dados de entrada consistiram em uma matriz com 191 linhas, amostras correspondentes ao número de dados de concentração de fósforo total, nitrogênio total e carbono orgânico dissolvido, e 69 colunas, sendo 68 variáveis independentes e uma variável dependente (concentração). Todos os testes realizados pelo programa estão indicados na Tabela 3.

O programa considerou, inicialmente, 80% das amostras escolhidas aleatoriamente para treino, enquanto as demais 20% foram utilizadas para verificação. As amostras foram escolhidas aleatoriamente, mas serão tratadas neste trabalho como escolhidas “pseudo-aleatoriamente” pois foi fixado um número arbitrário (42) que induz o programa a separar os dados de treino e verificação sempre da mesma forma, para que pudessem ser comparadas as análises de cada poluente.

3.2.4.2 Critério de busca

Para encontrar o melhor resultado de regressão para cada poluente, seria necessário rodar o programa com todas as possíveis combinações de dados de variáveis independentes (colunas), para que fossem encontradas as variáveis que produzissem o melhor resultado de predição de concentração. Seriam necessárias, logo, $3,6 \cdot 10^{94}$ (67 fatorial) rodadas do LazyPredict, que se torna inviável pelo tempo de processamento, uma vez que cada processamento dura entre 1 e 2 segundos e, portanto, seriam necessários bilhões de anos para obter o resultado. Visando contornar esta dificuldade, decidiu-se por inserir no programa duas aleatoriedades (número de variáveis a serem testadas e a escolha dessas variáveis) que permitissem testar algumas combinações de variáveis dentre o universo de 67 fatorial combinações possíveis. Ainda que não tenham sido testadas todas as combinações, é esperado que a análise de algumas delas consigam descrever uma tendência ou, ainda, bons resultados de combinação de variáveis.

Para cada poluente, foram realizadas 6 rodadas diferentes de testes, para testar o número de combinações necessárias para atingir modelos com bons resultados. Os testes foram: rodagem de 20 combinações diferentes considerando todas as variáveis possíveis; 1000 combinações com todas as variáveis; 4500 combinações com todas as variáveis; rodagem de 20 combinações diferentes considerando todas as variáveis exceto concentração de outros poluentes; 1000 combinações com variáveis exceto concentração; 4500 combinações com variáveis exceto concentração. Cada combinação, um número aleatório de variáveis a serem escolhidas era gerado e, então, essas variáveis eram escolhidas aleatoriamente para comporem os modelos.

A fim de analisar o que o tamanho da separação das amostras de teste e verificação influenciariam no resultado, apenas para o Fósforo Total foram realizadas as rodadas de análise: T1 – 10% de amostra de teste e 90% de treino; T2 – 30% da amostra de teste e 70% da amostra de treino.

Para analisar a influência da aleatoriedade da escolha de quais amostras seriam treino e verificação, foram realizadas apenas para o Fósforo Total as rodadas de análise: T3 – mudança do número arbitrário de aleatoriedade de 42 para 3; T4 – mudança do número arbitrário de aleatoriedade de 42 para 94.

Ainda, foram realizados testes de acordo com os resultados analisados da matriz de correlação de dados e combinação de dados selecionados, com intuito de observar se somente esses componentes já seriam o suficiente para obter bons resultados, sem que se dependa exclusivamente de uma escolha das variáveis independentes da inteligência artificial.

A comparação entre resultados utilizou o indicador R^2 (coeficiente de determinação), e para os melhores resultados também foi calculado o RMSE (raiz do erro quadrático médio). Os testes realizados estão indicados na Tabela 3.

O nome de cada teste indica as principais variáveis utilizadas por eles. O número é referente a quantidade de simulações realizadas (quantidade de combinações de dados de entrada, que são definidos aleatoriamente dentre aqueles possíveis para a simulação). As letras são referentes a quais dados de entrada são passíveis de serem escolhidos aleatoriamente para a simulação, conforme indicado abaixo.

TABELA 3 - TESTES REALIZADOS

Descrição do Teste	Teste/ Quant. de simulações	Variáveis utilizadas	Treino/ Teste	Def. Pseudo Aleatória
Simulações com número aleatório de variáveis utilizadas e variáveis escolhidas aleatoriamente	20	Potencialmente todas, exceto concentração do parâmetro em análise	80% / 20%	42
	1000			
	4500			
	20sc	Potencialmente todas, sem variáveis de concentração		
	1000sc			
	4500sc			
	t1 / 20	Potencialmente todas, exceto concentração do parâmetro em análise	90% / 10%	3 94
	t2 / 20		70% / 30%	
	t3 / 20			
	t4 / 20			
	1000sce	Potencialmente todas, sem variáveis de concentração e estações contribuintes de Thiessen (variáveis 46 a 69)	80% / 20%	42
1000scet	Potencialmente todas, sem variáveis deconcentração e estações contribuintes			

		de Thiessen (variáveis 46 a 69) e tipo de solo (variáveis 19 a 24)		
	1000sceu	Potencialmente todas, sem variáveis deconcentração e estações contribuintes de Thiessen (variáveis 46 a 69) e uso da terra (variáveis 25 a 41)		
	1000scetu	Potencialmente todas, sem variáveis deconcentração e estações contribuintes de Thiessen (variáveis 46 a 69), tipo de solo (variáveis 19 a 24) e uso da terra (variáveis 25 a 41)		
	1000scuteth	Potencialmente todas, sem variáveis deconcentração e estações contribuintes de Thiessen (variáveis 46 a 69), precipitação de Thiessen (variáveis 42 a 45), tipo de solo (variáveis 19 a 24) e uso da terra (variáveis 25 a 41)		
Modelo rodado com variáveis específicas, escolhidas estrategicamente	Utodas	Todas, exceto concentração do parâmetro em análise		
	Utodas-sc	Todas, sem variáveis deconcentração		
	Utodas-sce	Todas, sem variáveis de concentração e estações contribuintes de Thiessen (variáveis 46 a 69)		
	Utodas-sceu	Todas, sem variáveis de concentração e estações contribuintes de Thiessen (variáveis 46 a 69) e uso da terra (variáveis 25 a 41)		
	Utodas-scet	Todas, sem variáveis de concentração e estações contribuintes de Thiessen (variáveis 46 a 69) e tipo de solo (variáveis 19 a 24)		
	Utodas-scet	Todas, sem variáveis de concentração e estações contribuintes de Thiessen (variáveis 46 a 69), tipo de solo (variáveis 19 a 24) e uso da terra (variáveis 25 a 41)		
	Utodas-simples1	Tempo de concentração, tempo desde chuva inicial, vazão no rio,		

		evapotranspiração real, área da bacia, declividade média e precipitação antecedente de Thiessen de 1, 3, 5 e 7 dias		
	Utodas-simples2	Vazão no rio, evapotranspiração real, área da bacia, declividade média e precipitação antecedente de Thiessen de 1 dia		
Modelos rodado com as variáveis que geraram o melhor resultado para R ²	Melhor R	(depende do resultado)		
Modelo rodado com as variáveis mais utilizadas para obtenção dos melhores resultados	Variáveis mais expressivas	(depende do resultado)		

Fonte: A Autora (2021)

4 RESULTADOS E DISCUSSÕES

As simulações descritas na Tabela 3 foram realizadas em linguagem Python, conforme APÊNDICE A – Programa em Python de Regressão Linear para Dados de Concentração Durante Eventos de Precipitação, com o objetivo de encontrar a melhor relação entre a concentração dos parâmetros Fósforo Total, Nitrogênio Total e Carbono Orgânico Dissolvido durante eventos de precipitação e características dos eventos, da bacia e componentes do ciclo hidrológico.

A partir das simulações, a Tabela 4 destaca o melhor coeficiente de determinação R^2 (coeficiente de determinação) de cada teste e seu respectivo RSME (Raiz do Erro Quadrático Médio), para cada um dos parâmetros, assim como a quantidade de colunas (variáveis) utilizada no teste com melhor R^2 (mais próximo a 1) e o regressor que gerou tal resultado. As células em verde destacam os melhores resultados para R^2 (maior) e RSME (menor).

Na Tabela 7 a Tabela 9 são apresentadas as variáveis (colunas) usadas por esses resultados para os parâmetros Fósforo Total, Nitrogênio Total e Carbono Total, respectivamente. As siglas utilizadas nas tabelas estão descritas no APÊNDICE B – *Variáveis e Siglas*. Nas tabelas são indicadas quais colunas geraram os resultados de melhor R^2 para cada simulação, sendo as colunas em azuis aquelas utilizadas, as em preto que não entraram no modelo e as em branco que eram possíveis de serem escolhidas aleatoriamente, mas não formaram o melhor resultado.

A análise foi realizada sob três critérios:

- Maior coeficiente de determinação (R^2);
- Menor raiz do erro quadrático médio (RSME);
- Variáveis selecionadas pelos regressores com maior R^2 e menor RMSE, em termos de quantidade, frequência de ocorrência pelos regressores e facilidade de obtenção.

TABELA 4 - MELHORES RESULTADOS PELO INDICADOR R² DAS SIMULAÇÕES PARA OS PARÂMETROS FÓSFORO TOTAL, NITROGÊNIO TOTAL E CARBONO ORGÂNICO DISSOLVIDO

Nome Teste/ Quantidade de rodadas	PT				NT				COD			
	R ²	RSME	Quant. de variáveis usadas	Regressor	R ²	RSME	Quant. de variáveis usadas	Regressor	R ²	RSME	Quant. de variáveis usadas	Regressor
20	0,851	0,333	24	ExtraTrees	0,731	2,111	59	LGBM	0,410	2,322	22	ExtraTree
1000	0,935	0,314	31	Bagging	0,903	2,102	63	ExtraTree	0,620	2,3488	8	DecisionTree
4500	0,935	0,309	14	Bagging	0,967	2,160	23	ExtraTree	0,689	2,344	22	ExtraTree
20sc	0,849	0,383	17	HistGradientBoosting	0,802	1,801	19	KNeighbors	0,283	2,358	18	Bagging
1000sc	0,898	0,331	8	RandomForest	0,855	2,207	8	KNeighbors	0,510	2,144	9	DecisionTree
4500sc	0,900	0,368	12	ExtraTrees	0,875	2,134	3	KNeighbors	0,632	2,399	10	DecisionTree
1000sceu	0,890	0,409	9	RandomForest	-	-	-	-	-	-	-	-
1000scet	0,883	0,372	12	RandomForest	-	-	-	-	-	-	-	-
1000sceu	0,898	0,355	6	GradientBoosting	-	-	-	-	-	-	-	-
1000scetu	0,910	0,379	11	Bagging	-	-	-	-	-	-	-	-
1000scuteth	0,911	0,375	4	Bagging	-	-	-	-	-	-	-	-
Utodas	0,848	0,290	68	ExtraTrees	0,715	2,159	68	ExtraTree	0,206	2,326	68	Bagging
Utodas-sc	0,751	0,370	66	ExtraTrees	0,743	2,052	66	ExtraTree	0,210	2,319	66	Bagging
Utodas-sceu	0,738	0,380	42	ExtraTrees	0,700	2,215	42	ExtraTree	0,167	2,382	42	RandomForest
Utodas-sceu	0,744	0,376	36	DecisionTree	0,707	2,191	36	DecisionTree	0,163	2,388	36	RandomForest
Utodas-scet	0,778	0,35	19	ExtraTrees	0,671	2,322	19	ExtraTree	0,190	2,349	19	SGDR
Utodas- scetu	0,711	0,399	25	ExtraTrees	0,658	2,365	25	ExtraTree	0,201	2,333	25	AdaBoost

Utodas-simples1	0,871	0,267	10	XGB	0,637	2,439	10	XGBR	0,185	2,356	10	AdaBoost
Utodas-simples2	0,875	0,262	5	LGBM	0,571	2,651	5	LGBM	0,184	2,357	5	SVR
t1 / 20	0,942	0,220	56	XGB	-	-	-	-	-	-	-	-
t2 / 20	0,942	0,549	56	XGB	-	-	-	-	-	-	-	-
t3 / 20	0,862	1,101	44	MLP	-	-	-	-	-	-	-	-
t4 / 20	0,874	0,486	21	GradientBoosting	-	-	-	-	-	-	-	-

Fonte: A Autora (2021)

Para o parâmetro Fósforo Total, em que foram realizadas todas as simulações descritas na Tabela 3, diversas são as considerações que podem ser apontadas. Ao se analisar qual o regressor que gerou o melhor resultado, foi considerado o R^2 como indicador, o regressor Bagging foi o que apresentou os melhores resultados, atingindo um R^2 de 0,935. Estes resultados foram obtidos pelas simulações aleatórias entre todas as variáveis possíveis em 1000 e 4500 simulações. O RMSE para ambos ficou próximo, sendo 0,314 e 0,309 mg/L, respectivamente. Este resultado mostrou-se relevante pelo fato de que o indicador R^2 determinou resultados que, apesar de não serem os melhores RSME dentre os testes, não ficaram próximos aos piores.

Entretanto, a quantidade de variáveis utilizada para obter tais resultados foi muito diferente, sendo 31 e 14, respectivamente, com 10 variáveis comuns entre elas, como pode ser observado na Tabela 7 (células azuis). Nestes testes foram utilizadas as concentrações de nitrogênio total, vazão do rio, alguns tipos de uso da terra e tipo de solo, além de precipitações antecedentes referentes à bacia e a estação próxima ao ponto, que indicam principalmente a condição de saturação do solo. Interessantemente, para estes casos, não foram necessários para o melhor resultado as indicações diretas de localização do ponto amostrado, tal como o ponto, o tempo de concentração, área da bacia e declividade da bacia. Isto pode indicar que as características da bacia e condições do evento, por si só, já são o suficiente para obter bons resultados. Em especial, as características da bacia e condições do evento em que se mostram relevantes são uso da terra, tipo do solo e precipitação antecedente, além da condição do próprio rio no momento da amostra coletada, representado pela vazão e concentração de nitrogênio total.

Por apresentar, portanto, o R^2 mais alto, o RSME mais baixo e a menor quantidade de variáveis utilizadas, o melhor resultado da simulação 4500 foi considerado como o melhor resultado geral, correspondendo ao teste “Melhor R”, que sofreu outras análises a seguir.

Outro resultado expressivo foi a simulação intitulada “Utodas-simples2”, na qual foram utilizadas apenas 5 variáveis pré-determinadas, que resultaram no melhor RSME de 0,262 mg/L e em um R^2 de 0,875, dentre as simulações com melhores R^2 . O regressor foi o LGBM e as variáveis utilizadas contemplam, de forma simplificada, o ciclo hidrológico, sendo o escoamento superficial representado pela vazão do rio, a evapotranspiração real, as condições da bacia que influenciam nas interceptações e

no escoamento superficial, tais quais a área da bacia e a declividade da bacia, além da potencial infiltração na bacia e posterior escoamento subterrâneo, representado pela precipitação antecedente de 1 dia calculada por Thiessen.

Em análise inicial, a quantidade de variáveis utilizada nas melhores predições não está diretamente relacionada com os melhores resultados, mas a utilização de poucas variáveis que geram bons resultados pode gerar um modelo mais simples e rápido, a depender também de quais são estas informações. Em adição, deve-se atentar também a quais são as variáveis utilizadas, pois estas podem gerar modelos sem significado físico ou teórico.

Para o nitrogênio total, a melhor solução também foi encontrada na rodada 4500, com R^2 de 0,903. Isso indica a tendência de encontrar melhores resultados quanto maior foi o número de conjuntos aleatórios de variáveis testadas, significando que pode haver resultados ainda melhores. O regressor Extra Tree Regressor foi aquele que gerou o melhor modelo, a partir de 23 variáveis combinadas, incluindo a concentração de COD. Sem utilizar os dados de concentração, o melhor resultado obteve um R^2 de 0,875, também por rodagem de 4500 sem concentração, utilizando apenas três dados de entrada: ponto, precipitação antecedente de 5 dias (2549090) e precipitação antecedente de 7 dias (2549006). Fisicamente, entretanto, este resultado pode não representar um resultado viável, uma vez que apenas considera o espaço e as condições do solo no momento da precipitação. A determinação de um modelo representativo precisa levar em consideração o significado de suas variáveis e sua aderência em situações diversas de inputs, sendo de grande importância a avaliação do usuário do modelo entender se ele é adequado.

Para o carbono orgânico dissolvido, o melhor regressor performou um R^2 de 0,689 que, se comparado com os outros parâmetros, é um valor baixo. O melhor resultado é composto por um conjunto de 22 dados de entrada, que não incluem concentração de outro parâmetro e foi igualmente obtido pela rodada de 4500. Assim como para o nitrogênio total, o melhor regressor foi o Extra Tree Regressor.

A Tabela 5 resume os 10 melhores resultados entre todos os resultados obtidos em relação ao indicador R^2 , além do melhor resultado sem que se utilizem dados de concentração, para cada um dos parâmetros, com exceção dos testes t1, t2, t3 e t4, por serem exclusivamente utilizados como comparação de aplicação.

TABELA 5 - MELHORES 10 RESULTADOS ENTRE TODOS OS TESTES (EXCETO T1, T2, T3 E T4) PARA TODOS OS PARÂMETROS

Melhores Resultados	PT			NT			COD		
	R ²	RSME	Regressor (quant. Variáveis)	R ²	RSME	Regressor (quant. Variáveis)	R ²	RSME	Regressor (quant. Variáveis)
1	0,935	0,314	Bagging (31)	0,967	2,160	ExtraTree (23)	0,689	2,344	ExtraTree (22)
2	0,935	0,309	Bagging (14)	0,919	2,450	ExtraTree (32)	0,635	2,307	DecisionTree (13)
3	0,930	0,271	Bagging (64)	0,916	2,222	ExtraTree (57)	0,632	2,399	DecisionTree (10)
4	0,929	0,388	Bagging (37)	0,912	2,161	ExtraTree (41)	0,620	2,348	DecisionTree (8)
5	0,925	0,307	Bagging (15)	0,903	2,102	ExtraTree (63)	0,609	2,370	DecisionTree (4)
6	0,924	0,305	Bagging (32)	0,902	1,996	Kneighbors (9)	0,581	2,363	ExtraTree (8)
7	0,919	0,426	XGB (18)	0,901	2,920	ExtraTree (52)	0,561	2,330	ExtraTree (16)
8	0,918	0,306	Bagging (39)	0,900	1,938	ExtraTree (32)	0,560	2,458	DecisionTree (13)
9	0,917	0,464	XGB (41)	0,896	1,481	ExtraTree (52)	0,544	2,358	DecisionTree (7)
10	0,915	0,252	Bagging (63)	0,893	2,273	ExtraTree (63)	0,536	2,167	ExtraTree (60)
1 sc	0,911	0,375	Bagging (4)	0,875	2,134	Kneighbors (3)	0,689	2,344	ExtraTree (22)

Fonte: A Autora (2021)

Dentre os regressores, o Bagging ganhou destaque para fósforo total por gerar 8 dos 10 melhores resultados de R². Ainda, os melhores regressores são caracterizados como conjunto de modelos, desenvolvidos a partir de árvores de decisão, o que pode indicar que este tipo de construção se adequa bem para explicar e prever as concentrações de fósforo total. Para nitrogênio total, o regressor Extra Tree foi maioria, enquanto para carbono orgânico dissolvido, o Extra Tree e o Decision Tree foram os mais relevantes, sendo ambos categorizados como modelos de árvore de decisão. Sendo assim, é possível afirmar que, para este conjunto de dados de concentração na bacia do rio Barigui, os modelos do tipo árvore de decisão, sendo simples ou conjuntos, são os melhores previsores de concentração de PT, NT e COD durante eventos de precipitação.

Dentre os testes de diferentes considerações, os testes t1 e t2, que considera uma separação entre as amostras de treino e verificação de proporção (90%-10% para o primeiro e 70%-30% para o segundo), indicaram o melhor indicador R², de

0,942 para ambos, enquanto o primeiro obteve também o melhor RSME, de 0,220 mg/L. As variáveis utilizadas por ambos os testes foram as mesmas, o que indica uma convergência para o resultado. Em questão de comparação, pode-se dizer que o teste t1 resulta em um modelo de regressão mais acurado, uma vez que têm o menor RSME.

Por conta deste resultado, foi realizada uma simulação com as variáveis utilizadas na melhor simulação (teste 4500), mas alterando-se a proporção treino-verificação para 90%-10%, o que resultou em um R^2 de 0,88 pelo regressor DecisionTree, com RMSE de 0,22 mg/L, e para proporção treino-verificação para 70%-30%, o que resultou em um R^2 de 0,76 e RMSE de 0,39 mg/L pelo regressor Bagging. A proporção inicial adotada demonstrou-se mais adequada para obter o melhor resultado neste caso, além da indicação de que os resultados podem se comportar diferentes em cada conjunto de dados. Os resultados estão indicados na Tabela 6. Ainda, foi realizado um teste sem utilizar a variável de concentração, uma vez que este dado é de difícil obtenção e disponibilidade, resultado em um modelo de pior performance em relação a R^2 e melhor em relação a RSME, além do regressor mudar para o Random Forest.

O melhor resultado obtido continuou sendo, portanto, o encontrado pela simulação 4500, utilizando uma proporção treino-verificação de 80%-20% pelo regressor Bagging. A Figura 14 mostra, para o considerado melhor resultado, as concentrações reais medidas em campo comparadas às concentrações calculadas pelo regressor, com boas previsões levemente superestimadas quando próximos a 1mg/L, chegando a 0,6mg/L no erro absoluto, com média de 0,1mg/L de erro.

As variáveis utilizadas para a geração deste modelo mostraram, ainda, significado físico quando comparada com o ciclo hidrológico temporal, diagramado na Figura 5. As condições do corpo hídrico estão representadas pela vazão no rio e concentração de NT, enquanto os escoamentos subterrâneo e superficial são diretamente ligados com as precipitações antecedentes, o tipo de solo e uso da terra. As precipitações antecedentes utilizadas são referentes às estações próximas ao exutório das bacias, às dentro da bacia (2549100 e 2549107) e a uma estação a jusante de BA2. Os usos do solo utilizados pelo modelo refletem aqueles já conhecidos por gerarem as maiores cargas difusas poluentes, área urbana baixa e cultura temporária. Os tipos de solo utilizados estão entre os menos expressivos na

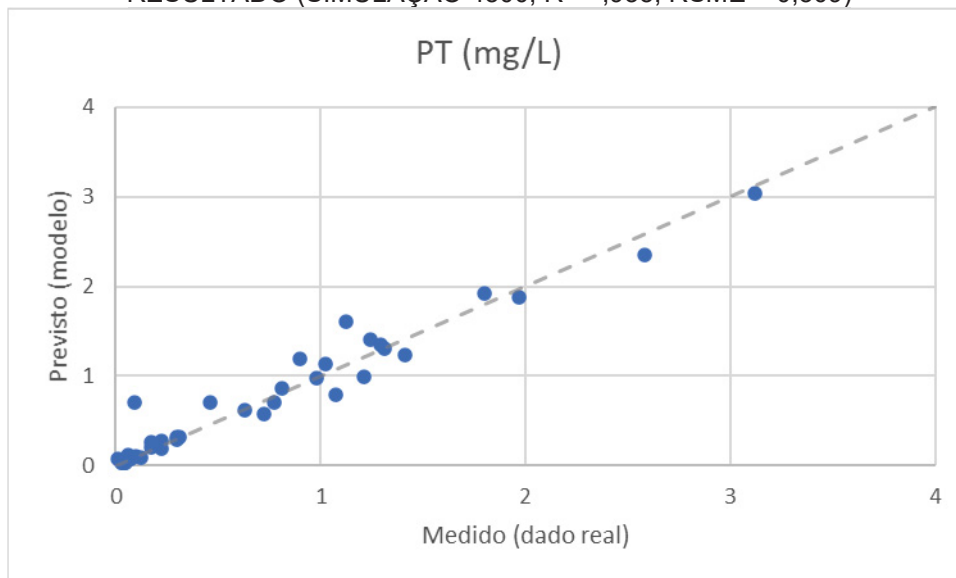
bacia, sendo cerca de 8% podzólico vermelho-amarelo eutrófico e 1% latossolo vermelho-escuro distrófico.

TABELA 6 - TESTES SIMULADOS A PARTIR DA MELHOR RESPOSTA (SIMULAÇÃO 4500, $R^2 = 0,935$, RSME = 0,309) DOS TESTES INICIAIS DA TABELA 3

Variáveis usadas (melhor resultado da Tabela 3)	Treino- Verificação	Número pseudo aleatório	R^2	RSME	Regressor	
Concentração NT Pacum ant. 5 dias Pacum ant. 7 dias Vazão no rio Tipo de Solo – PVAE Tipo de Solo – LVED Uso da terra – AUB Uso da terra – CT P ant 3 dias – Thiessen P ant 5 dias – Thiessen P ant 3 dias – 2549100 P ant 5 dias – 2549100 P ant 1 dia – 2549107 P ant 5 dias – 2549006	90/10	42	0,88	0,22	DecisionTree	
	70/30		0,76	0,39	Bagging	
			0,94	0,31	Bagging	
		80/20	3	0,85	0,45	LinearSVR
				0,72	0,63	Bagging
			94	0,85	0,36	GradientBoosting
Pacum ant. 5 dias Pacum ant. 7 dias Vazão no rio Tipo de Solo – PVAE Tipo de Solo – LVED Uso da terra – AUB Uso da terra – CT P ant 3 dias – Thiessen P ant 5 dias – Thiessen P ant 3 dias – 2549100 P ant 5 dias – 2549100 P ant 1 dia – 2549107 P ant 5 dias – 2549006	80/20	42	0,86	0,28	RandomForest	

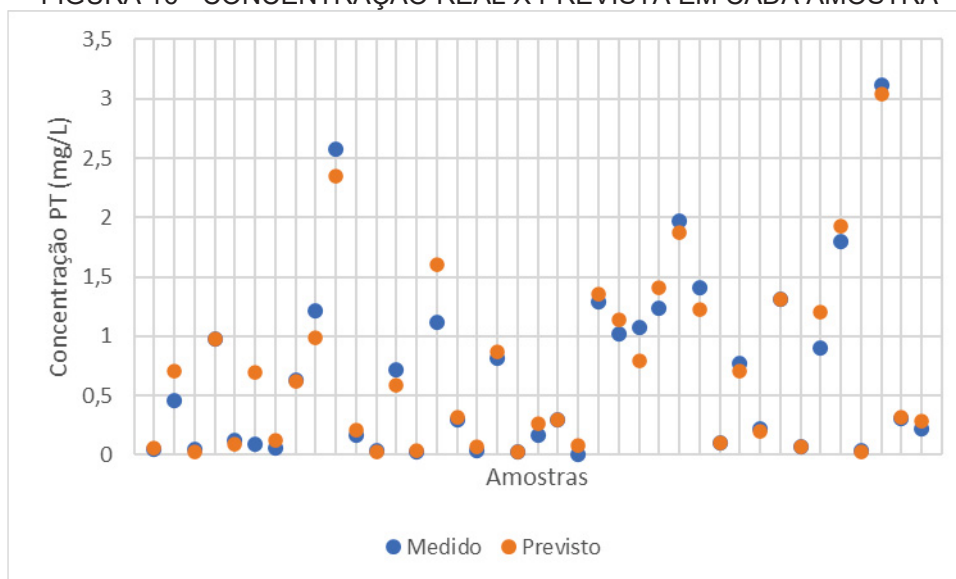
Fonte: A Autora (2021)

FIGURA 15 - CONCENTRAÇÕES REAL E MODELADA DE FÓSFORO TOTAL PELO MELHOR RESULTADO (SIMULAÇÃO 4500, $R^2 = ,935$, RSME = 0,309)



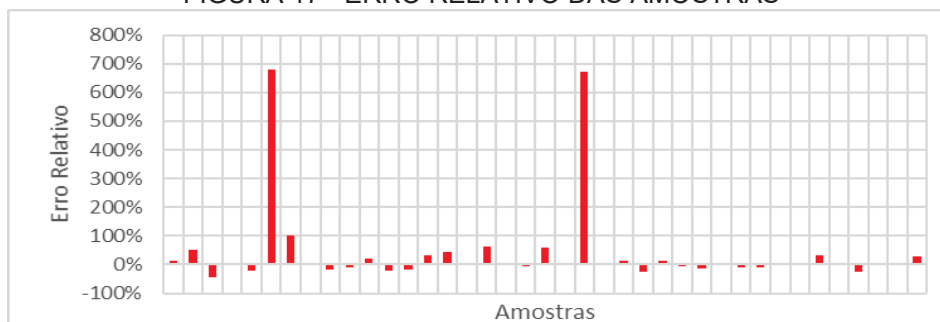
Fonte: A Autora (2021)

FIGURA 16 - CONCENTRAÇÃO REAL X PREVISTA EM CADA AMOSTRA



Fonte: A Autora (2021)

FIGURA 17 - ERRO RELATIVO DAS AMOSTRAS



Fonte: A Autora (2021)

FIGURA 18 - ERRO ABSOLUTO DAS AMOSTRAS



Fonte: A Autora (2021)

A fim de se avaliar o comportamento da regressão com um conjunto diferente de treino-verificação, mantendo-se a proporção de 80%-20%, foram realizados testes similares a t3 e t4, mudando-se o número pseudoaleatório, de 42 para 3 e 94, respectivamente com o mesmo conjunto de 14 variáveis da simulação 4500. Os resultados obtidos foram de $R^2=0,85$, $RSME=0,45$ mg/L e regressor LinearSVR para o primeiro caso, enquanto o regressor Bagging, o mesmo que gerou o melhor resultado, produziu $R^2=0,72$ e $RSME=0,63$ mg/L. Para o segundo caso, os resultados foram $R^2=0,85$ e $RSME=0,36$ mg/L para o regressor GradientBoosting, enquanto o regressor Bagging produziu $R^2=0,82$ e $RSME=0,40$ mg/L.

Tais resultados evidenciam que os dados de entrada utilizados para treino e verificação têm influência significativa no resultado do modelo e, portanto, em suas previsões. Em qualquer modelo é importante que sejam utilizados dados para treino que contemplem diferentes especificidades, para que o modelo fique o mais global possível, podendo prever as concentrações em diferentes condições. A coluna, por exemplo, de tempo desde a chuva inicial é relevante neste aspecto pois separa as amostras coletadas durante eventos de chuva e durante o “baseline”.

A simulação utilizando todos os dados de entrada (Utodos) em que inicialmente era esperado o melhor resultado por considerar todos os elementos que foram considerados importantes para a formulação das variáveis, não foi o caso para R^2 , mas obteve um bom RSME. Isso indica que existem variáveis que podem piorar a performance de um modelo se forem consideradas, por terem dados não relevantes à predição de concentração. Vale evidenciar que, em comparação com a simulação que usa todos os dados de entrada exceto dados de concentração (Utodos-sc), a utilização da concentração melhorou a performance do regressor (que é o mesmo para ambos, ExtraTrees), mostrando que a concentração de outros poluentes ajuda o modelo a estimar a concentração de certo poluente. Todavia, a disponibilidade de dados de concentração de outros parâmetros de qualidade da água em situações de chuva, como é analisado neste trabalho, não é de fácil acesso e nem tão frequentes de serem medidos e encontrados. A consideração de concentrações de outros poluentes como dados de entrada, portanto, pode não gerar um modelo utilizável em termos práticos.

FIGURA 19 - MATRIZ DE CORRELAÇÃO ENTRE VARIÁVEIS DE ENTRADA

	Amostra	Ponto	Evento	Tc (min)	Tcl (min)	Conc. NT (mg/L)	Conc. PT (mg/L)	Conc. DOC (mg/L)	P. Acum. - Tc (mm)
Amostra	1,00	0,59	0,99	-0,28	-0,21	0,19	0,00	-0,05	-0,27
Ponto	0,59	1,00	0,63	0,13	-0,09	0,49	0,45	0,29	0,00
Evento	0,99	0,63	1,00	-0,24	-0,20	0,26	0,08	-0,03	-0,22
Tc (min)	-0,28	0,13	-0,24	1,00	0,65	0,05	0,10	0,23	0,32
Tcl (min)	-0,21	-0,09	-0,20	0,65	1,00	-0,16	-0,17	0,18	-0,14
Conc. NT (mg/L)	0,19	0,49	0,26	0,05	-0,16	1,00	0,66	0,25	0,14
Conc. PT (mg/L)	0,00	0,45	0,08	0,10	-0,17	0,66	1,00	0,23	0,33
Conc. DOC (mg/L)	-0,05	0,29	-0,03	0,23	0,18	0,25	0,23	1,00	0,05
P. Acum. - Tc (mm)	-0,27	0,00	-0,22	0,32	-0,14	0,14	0,33	0,05	1,00
P. Acum. - Tcl (mm)	-0,40	-0,22	-0,38	0,69	0,65	-0,14	-0,03	0,14	0,32
P. Antec. 1 d (mm)	-0,44	-0,14	-0,40	0,67	0,35	-0,08	0,05	0,14	0,45
P. Antec. 3 d (mm)	-0,42	-0,06	-0,37	0,72	0,55	0,00	0,08	0,19	0,31
P. Antec. 5 d (mm)	-0,26	-0,08	-0,24	0,27	0,26	-0,03	0,05	0,14	0,04
P. Antec. 7 d (mm)	-0,54	-0,06	-0,50	0,41	0,34	0,00	0,06	0,15	0,14
Qrio (m³/s)	-0,47	0,08	-0,45	0,14	-0,05	-0,06	0,29	0,20	0,25
Hidrog. sobre (2)/desce (1)	0,21	0,13	0,23	-0,36	-0,48	0,14	0,17	-0,11	0,17
ET real	-0,45	0,06	-0,39	0,18	0,05	0,15	0,28	0,25	0,03
A bacia (km²)	-0,18	0,29	-0,11	0,68	0,47	-0,06	0,23	0,19	0,12
l (m/km)	-0,57	-0,03	-0,49	0,63	0,44	0,05	0,29	0,27	0,36
TS - LVA (%)	-0,61	-0,30	-0,56	0,45	0,39	-0,27	0,08	0,10	0,23
TS - CA (%)	-0,58	-0,20	-0,53	0,57	0,42	-0,20	0,08	0,21	0,38
TS - PVAA (%)	-0,07	0,42	0,03	0,27	0,07	0,67	0,58	0,27	0,15
TS - LVED (%)	-0,47	-0,30	-0,46	0,48	0,58	-0,31	-0,16	0,11	-0,13
TS - Ctba (%)	0,12	0,53	0,21	0,11	-0,09	0,70	0,60	0,24	0,13
TS - PVAE (%)	0,22	0,56	0,28	0,39	0,21	0,00	0,25	0,12	-0,02
US - AUB (%)	-0,14	0,14	-0,07	0,03	-0,08	0,22	0,41	0,08	0,16
US - VAN (%)	-0,50	0,08	-0,41	0,62	0,39	0,19	0,38	0,30	0,37
US - Campo (%)	-0,53	0,04	-0,44	0,62	0,40	0,11	0,33	0,29	0,40
US - Água (%)	-0,42	0,17	-0,32	0,55	0,29	0,22	0,44	0,33	0,44
US - VAP (%)	-0,59	-0,24	-0,55	0,61	0,54	-0,20	0,02	0,19	0,18
US - VAN2 (%)	-0,60	-0,23	-0,56	0,61	0,52	-0,18	0,05	0,17	0,23
US - M/O (%)	-0,59	-0,40	-0,57	0,29	0,31	-0,31	-0,11	0,06	0,23
US - AI (%)	-0,43	-0,05	-0,37	0,34	0,18	0,00	0,21	0,18	0,39
US - Vila (%)	-0,59	-0,25	-0,54	0,43	0,33	-0,13	0,11	0,21	0,39
US - CT (%)	-0,34	0,21	-0,25	0,69	0,42	0,08	0,34	0,26	0,31
US - AA (%)	-0,43	-0,32	-0,43	0,16	0,13	-0,19	-0,03	0,23	0,34
US - Granja (%)	-0,02	0,43	0,07	0,32	0,05	0,42	0,48	0,21	0,29
US - CP (%)	-0,50	-0,28	-0,47	0,48	0,41	-0,28	-0,04	0,08	0,25
US - SE (%)	-0,23	0,31	-0,14	0,54	0,37	0,31	0,39	0,30	0,10
US - AUM (%)	0,11	0,47	0,19	0,07	-0,09	0,62	0,52	0,22	0,12
US - Lot. (%)	0,13	0,54	0,22	0,12	-0,08	0,64	0,56	0,24	0,15
US - Lixão (%)	0,16	0,56	0,23	0,29	0,01	0,32	0,40	0,18	0,20
P. Antec. 1 d (mm) - Thiessen	-0,29	0,05	-0,23	0,39	0,26	-0,11	0,17	0,16	0,10
P. Antec. 3 d (mm) - Thiessen	-0,30	0,12	-0,22	0,42	0,43	0,05	0,25	0,20	0,06
P. Antec. 5 d (mm) - Thiessen	-0,41	0,09	-0,35	0,34	0,36	0,05	0,25	0,18	0,02
P. Antec. 7 d (mm) - Thiessen	-0,41	0,06	-0,35	0,22	0,25	0,08	0,24	0,18	0,00
P. Antec. 1 d (mm) - 2549100	-0,06	0,16	0,01	0,19	0,14	-0,01	0,27	0,16	0,05
P. Antec. 3 d (mm) - 2549100	-0,07	0,19	0,00	0,27	0,29	0,01	0,24	0,19	-0,02
P. Antec. 5 d (mm) - 2549100	-0,21	0,17	-0,14	0,21	0,25	0,01	0,29	0,18	-0,03
P. Antec. 7 d (mm) - 2549100	-0,15	0,22	-0,08	0,12	0,20	0,08	0,33	0,18	-0,07
P. Antec. 1 d (mm) - 2549107	-0,34	-0,03	-0,31	0,36	0,23	-0,15	0,07	0,14	0,08
P. Antec. 3 d (mm) - 2549107	-0,37	0,09	-0,30	0,39	0,36	0,24	0,30	0,20	0,15
P. Antec. 5 d (mm) - 2549107	-0,41	0,10	-0,35	0,34	0,31	0,20	0,28	0,19	0,10
P. Antec. 7 d (mm) - 2549107	-0,45	0,04	-0,40	0,29	0,25	0,16	0,22	0,18	0,10
P. Antec. 1 d (mm) - 2549128	-0,30	-0,01	-0,24	0,36	0,26	-0,14	0,14	0,05	0,13
P. Antec. 3 d (mm) - 2549128	-0,28	0,11	-0,21	0,47	0,48	-0,01	0,20	0,18	0,02
P. Antec. 5 d (mm) - 2549128	-0,44	0,02	-0,41	0,27	0,34	-0,02	0,12	0,11	-0,03
P. Antec. 7 d (mm) - 2549128	-0,47	0,02	-0,43	0,19	0,22	0,04	0,14	0,13	-0,05
P. Antec. 1 d (mm) - 2549077	0,20	0,60	0,26	0,34	0,01	0,24	0,38	0,17	0,10
P. Antec. 3 d (mm) - 2549077	0,23	0,66	0,30	0,31	0,10	0,22	0,40	0,18	0,02
P. Antec. 5 d (mm) - 2549077	0,23	0,69	0,31	0,35	0,12	0,33	0,45	0,19	0,06
P. Antec. 7 d (mm) - 2549077	0,32	0,78	0,40	0,19	0,01	0,39	0,47	0,20	0,00
P. Antec. 1 d (mm) - 2549090	-0,31	-0,04	-0,27	0,42	0,32	-0,23	0,05	0,07	0,12
P. Antec. 3 d (mm) - 2549090	-0,22	0,14	-0,15	0,52	0,51	0,01	0,15	0,15	0,07
P. Antec. 5 d (mm) - 2549090	-0,30	0,13	-0,25	0,40	0,44	0,05	0,15	0,12	0,00
P. Antec. 7 d (mm) - 2549090	-0,29	0,15	-0,23	0,29	0,33	0,11	0,16	0,10	-0,01
P. Antec. 1 d (mm) - 2549006	0,17	0,58	0,25	0,48	0,10	0,22	0,37	0,16	0,13
P. Antec. 3 d (mm) - 2549006	0,24	0,64	0,31	0,34	0,11	0,11	0,37	0,15	-0,02
P. Antec. 5 d (mm) - 2549006	0,24	0,64	0,31	0,37	0,21	0,11	0,36	0,15	-0,01
P. Antec. 7 d (mm) - 2549006	0,25	0,70	0,33	0,28	0,10	0,30	0,48	0,16	0,01

	P. Acum. - Tci (mm)	P. Antec. 1 d (mm)	P. Antec. 3 d (mm)	P. Antec. 5 d (mm)	P. Antec. 7 d (mm)	Qrio (m³/s)	Hidrog. sobe (2)/desce (1)	ET real
Amostra	-0,40	-0,44	-0,42	-0,26	-0,54	-0,47	0,21	-0,45
Ponto	-0,22	-0,14	-0,06	-0,08	-0,06	0,08	0,13	0,06
Evento	-0,38	-0,40	-0,37	-0,24	-0,50	-0,45	0,23	-0,39
Tc (min)	0,69	0,67	0,72	0,27	0,41	0,14	-0,36	0,18
Tci (min)	0,65	0,35	0,55	0,26	0,34	-0,05	-0,48	0,05
Conc. NT (mg/L)	-0,14	-0,08	0,00	-0,03	0,00	-0,06	0,14	0,15
Conc. PT (mg/L)	-0,03	0,05	0,08	0,05	0,06	0,29	0,17	0,28
Conc. DOC (mg/L)	0,14	0,14	0,19	0,14	0,15	0,20	-0,11	0,25
P. Acum. - Tc (mm)	0,32	0,45	0,31	0,04	0,14	0,25	0,17	0,03
P. Acum. - Tci (mm)	1,00	0,82	0,77	0,32	0,36	0,15	-0,39	-0,06
P. Antec. 1 d (mm)	0,82	1,00	0,79	0,36	0,43	0,26	-0,29	0,03
P. Antec. 3 d (mm)	0,77	0,79	1,00	0,47	0,71	0,22	-0,42	0,33
P. Antec. 5 d (mm)	0,32	0,36	0,47	1,00	0,51	0,17	-0,27	0,32
P. Antec. 7 d (mm)	0,36	0,43	0,71	0,51	1,00	0,38	-0,41	0,59
Qrio (m³/s)	0,15	0,26	0,22	0,17	0,38	1,00	0,00	0,41
Hidrog. sobe (2)/desce (1)	-0,39	-0,29	-0,42	-0,27	-0,41	0,00	1,00	-0,19
ET real	-0,06	0,03	0,33	0,32	0,59	0,41	-0,19	1,00
A bacia (km²)	0,45	0,41	0,50	0,22	0,36	0,43	-0,25	0,35
I (m/m)	0,59	0,58	0,63	0,25	0,40	0,49	-0,18	0,45
TS - LVA (%)	0,52	0,45	0,44	0,17	0,21	0,35	-0,16	0,31
TS - CA (%)	0,59	0,58	0,59	0,24	0,37	0,49	-0,24	0,37
TS - PVAA (%)	0,09	0,16	0,26	0,11	0,22	0,17	0,11	0,33
TS - LVED (%)	0,59	0,43	0,44	0,23	0,25	0,32	-0,35	0,18
TS - Ctba (%)	-0,10	0,00	0,10	0,02	0,11	0,07	0,19	0,26
TS - PVAE (%)	0,07	0,06	0,20	0,08	0,19	0,27	-0,16	0,26
US - AUB (%)	-0,05	0,01	0,03	-0,02	-0,02	0,04	0,18	0,27
US - VAN (%)	0,53	0,55	0,61	0,25	0,41	0,48	-0,13	0,47
US - Campo (%)	0,55	0,57	0,63	0,25	0,41	0,51	-0,16	0,48
US - Água (%)	0,43	0,50	0,57	0,21	0,37	0,48	-0,09	0,51
US - VAP (%)	0,66	0,57	0,58	0,26	0,35	0,44	-0,28	0,30
US - VAN2 (%)	0,65	0,57	0,58	0,25	0,35	0,44	-0,26	0,31
US - M/O (%)	0,46	0,39	0,35	0,15	0,17	0,36	-0,14	0,22
US - AI (%)	0,34	0,39	0,42	0,15	0,27	0,41	-0,08	0,39
US - Vila (%)	0,52	0,52	0,53	0,20	0,31	0,46	-0,18	0,36
US - CT (%)	0,50	0,52	0,59	0,24	0,42	0,48	-0,20	0,44
US - AA (%)	0,32	0,32	0,30	0,11	0,16	0,31	-0,10	0,23
US - Granja (%)	0,10	0,21	0,32	0,12	0,30	0,26	0,02	0,40
US - CP (%)	0,54	0,48	0,48	0,20	0,29	0,36	-0,25	0,21
US - SE (%)	0,38	0,38	0,47	0,21	0,32	0,35	-0,10	0,39
US - AUM (%)	-0,09	0,00	0,08	0,01	0,07	0,04	0,18	0,22
US - Lot. (%)	-0,09	0,01	0,11	0,03	0,12	0,08	0,18	0,27
US - Lixão (%)	-0,01	0,13	0,21	0,08	0,22	0,18	0,02	0,31
P. Antec. 1 d (mm) - Thiessen	0,50	0,56	0,49	0,37	0,32	0,39	-0,27	0,24
P. Antec. 3 d (mm) - Thiessen	0,49	0,41	0,67	0,43	0,56	0,34	-0,36	0,51
P. Antec. 5 d (mm) - Thiessen	0,35	0,33	0,61	0,51	0,77	0,45	-0,40	0,68
P. Antec. 7 d (mm) - Thiessen	0,24	0,23	0,50	0,48	0,78	0,48	-0,34	0,68
P. Antec. 1 d (mm) - 2549100	0,28	0,29	0,37	0,31	0,21	0,15	-0,22	0,32
P. Antec. 3 d (mm) - 2549100	0,30	0,23	0,51	0,34	0,36	0,15	-0,28	0,48
P. Antec. 5 d (mm) - 2549100	0,23	0,19	0,49	0,44	0,57	0,30	-0,32	0,64
P. Antec. 7 d (mm) - 2549100	0,15	0,09	0,37	0,42	0,55	0,34	-0,27	0,64
P. Antec. 1 d (mm) - 2549107	0,47	0,52	0,35	0,31	0,28	0,45	-0,20	0,14
P. Antec. 3 d (mm) - 2549107	0,47	0,42	0,58	0,39	0,58	0,39	-0,31	0,46
P. Antec. 5 d (mm) - 2549107	0,34	0,35	0,53	0,44	0,71	0,46	-0,32	0,61
P. Antec. 7 d (mm) - 2549107	0,27	0,30	0,52	0,43	0,80	0,43	-0,30	0,61
P. Antec. 1 d (mm) - 2549128	0,52	0,61	0,63	0,32	0,35	0,36	-0,29	0,22
P. Antec. 3 d (mm) - 2549128	0,48	0,41	0,70	0,41	0,55	0,35	-0,36	0,52
P. Antec. 5 d (mm) - 2549128	0,25	0,26	0,58	0,50	0,85	0,43	-0,42	0,66
P. Antec. 7 d (mm) - 2549128	0,10	0,13	0,44	0,44	0,87	0,49	-0,33	0,74
P. Antec. 1 d (mm) - 2549077	0,07	0,18	0,22	0,07	0,17	0,23	-0,11	0,28
P. Antec. 3 d (mm) - 2549077	-0,03	0,04	0,17	0,07	0,20	0,27	-0,09	0,34
P. Antec. 5 d (mm) - 2549077	-0,04	0,06	0,18	0,09	0,24	0,23	-0,08	0,36
P. Antec. 7 d (mm) - 2549077	-0,15	-0,08	0,04	0,03	0,16	0,25	-0,03	0,33
P. Antec. 1 d (mm) - 2549090	0,62	0,66	0,60	0,31	0,28	0,41	-0,31	0,11
P. Antec. 3 d (mm) - 2549090	0,59	0,52	0,77	0,39	0,51	0,32	-0,37	0,32
P. Antec. 5 d (mm) - 2549090	0,37	0,37	0,67	0,48	0,78	0,38	-0,44	0,50
P. Antec. 7 d (mm) - 2549090	0,24	0,26	0,58	0,46	0,86	0,37	-0,37	0,50
P. Antec. 1 d (mm) - 2549006	0,21	0,32	0,40	0,11	0,23	0,20	0,00	0,23
P. Antec. 3 d (mm) - 2549006	0,03	0,08	0,26	0,08	0,22	0,31	-0,04	0,30
P. Antec. 5 d (mm) - 2549006	0,03	0,07	0,23	0,09	0,23	0,28	-0,07	0,29
P. Antec. 7 d (mm) - 2549006	-0,04	0,01	0,17	0,08	0,24	0,33	0,01	0,36

	A bacia (km ²)	I (m/m)	TS - LVA (%)	TS - CA (%)	TS - PVAA (%)	TS - LVED (%)	TS - Ctba (%)	TS - PVAE (%)	US - AUB (%)
Amostra	-0,18	-0,57	-0,61	-0,58	-0,07	-0,47	0,12	0,22	-0,14
Ponto	0,29	-0,03	-0,30	-0,20	0,42	-0,30	0,53	0,56	0,14
Evento	-0,11	-0,49	-0,56	-0,53	0,03	-0,46	0,21	0,28	-0,07
Tc (min)	0,68	0,63	0,45	0,57	0,27	0,48	0,11	0,39	0,03
Tci (min)	0,47	0,44	0,39	0,42	0,07	0,58	-0,09	0,21	-0,08
Conc. NT (mg/L)	-0,06	0,05	-0,27	-0,20	0,67	-0,31	0,70	0,00	0,22
Conc. PT (mg/L)	0,23	0,29	0,08	0,08	0,58	-0,16	0,60	0,25	0,41
Conc. DOC (mg/L)	0,19	0,27	0,10	0,21	0,27	0,11	0,24	0,12	0,08
P. Acum. - Tc (mm)	0,12	0,36	0,23	0,38	0,15	-0,13	0,13	-0,02	0,16
P. Acum. - Tci (mm)	0,45	0,59	0,52	0,59	0,09	0,59	-0,10	0,07	-0,05
P. Antec. 1 d (mm)	0,41	0,58	0,45	0,58	0,16	0,43	0,00	0,06	0,01
P. Antec. 3 d (mm)	0,50	0,63	0,44	0,59	0,26	0,44	0,10	0,20	0,03
P. Antec. 5 d (mm)	0,22	0,25	0,17	0,24	0,11	0,23	0,02	0,08	-0,02
P. Antec. 7 d (mm)	0,36	0,40	0,21	0,37	0,22	0,25	0,11	0,19	-0,02
Qrio (m ³ /s)	0,43	0,49	0,35	0,49	0,17	0,32	0,07	0,27	0,04
Hidrog. sobe (2)/desce (1)	-0,25	-0,18	-0,16	-0,24	0,11	-0,35	0,19	-0,16	0,18
ET real	0,35	0,45	0,31	0,37	0,33	0,18	0,26	0,26	0,27
A bacia (km ²)	1,00	0,74	0,56	0,69	0,22	0,63	0,07	0,79	0,02
I (m/m)	0,74	1,00	0,72	0,91	0,44	0,64	0,22	0,31	0,15
TS - LVA (%)	0,56	0,72	1,00	0,74	-0,03	0,62	-0,22	0,21	0,47
TS - CA (%)	0,69	0,91	0,74	1,00	0,08	0,63	-0,14	0,28	0,00
TS - PVAA (%)	0,22	0,44	-0,03	0,08	1,00	0,02	0,90	0,07	0,29
TS - LVED (%)	0,63	0,64	0,62	0,63	0,02	1,00	-0,25	0,18	-0,19
TS - Ctba (%)	0,07	0,22	-0,22	-0,14	0,90	-0,25	1,00	0,11	0,36
TS - PVAE (%)	0,79	0,31	0,21	0,28	0,07	0,18	0,11	1,00	0,09
US - AUB (%)	0,02	0,15	0,47	0,00	0,29	-0,19	0,36	0,09	1,00
US - VAN (%)	0,71	0,98	0,61	0,83	0,59	0,54	0,39	0,31	0,16
US - Campo (%)	0,72	0,98	0,65	0,90	0,48	0,55	0,27	0,33	0,14
US - Água (%)	0,65	0,89	0,54	0,81	0,52	0,34	0,40	0,38	0,26
US - VAP (%)	0,74	0,89	0,75	0,87	0,18	0,88	-0,09	0,24	-0,09
US - VANZ (%)	0,74	0,91	0,78	0,88	0,18	0,83	-0,08	0,24	-0,05
US - M/O (%)	0,33	0,62	0,62	0,76	-0,13	0,53	-0,28	0,03	0,03
US - AI (%)	0,38	0,69	0,45	0,78	0,19	0,23	0,06	0,18	0,17
US - Vila (%)	0,49	0,83	0,63	0,93	0,09	0,50	-0,07	0,15	0,01
US - CT (%)	0,89	0,90	0,60	0,81	0,42	0,56	0,21	0,56	0,05
US - AA (%)	0,15	0,46	0,39	0,64	-0,15	0,22	-0,21	-0,03	-0,07
US - Granja (%)	0,34	0,45	0,07	0,27	0,72	-0,06	0,58	0,24	0,24
US - CP (%)	0,61	0,74	0,67	0,79	-0,07	0,64	-0,25	0,22	-0,18
US - SE (%)	0,66	0,73	0,33	0,44	0,76	0,53	0,65	0,36	0,19
US - AUM (%)	0,03	0,19	-0,21	-0,16	0,81	-0,23	0,97	0,07	0,34
US - Lot. (%)	0,09	0,23	-0,21	-0,12	0,87	-0,24	0,98	0,13	0,36
US - Lixão (%)	0,34	0,27	-0,03	0,13	0,57	-0,14	0,50	0,36	0,28
P. Antec. 1 d (mm) - Thiessen	0,68	0,63	0,52	0,58	0,20	0,56	0,05	0,41	0,07
P. Antec. 3 d (mm) - Thiessen	0,68	0,70	0,49	0,60	0,37	0,52	0,20	0,45	0,13
P. Antec. 5 d (mm) - Thiessen	0,59	0,58	0,39	0,49	0,34	0,41	0,20	0,42	0,12
P. Antec. 7 d (mm) - Thiessen	0,52	0,52	0,32	0,44	0,33	0,34	0,19	0,36	0,08
P. Antec. 1 d (mm) - 2549100	0,54	0,47	0,39	0,43	0,17	0,33	0,12	0,42	0,15
P. Antec. 3 d (mm) - 2549100	0,60	0,53	0,40	0,47	0,24	0,38	0,16	0,47	0,14
P. Antec. 5 d (mm) - 2549100	0,58	0,48	0,37	0,42	0,24	0,33	0,18	0,50	0,17
P. Antec. 7 d (mm) - 2549100	0,56	0,45	0,29	0,36	0,30	0,29	0,22	0,51	0,15
P. Antec. 1 d (mm) - 2549107	0,59	0,55	0,47	0,50	0,15	0,56	-0,03	0,30	0,00
P. Antec. 3 d (mm) - 2549107	0,52	0,64	0,38	0,50	0,52	0,41	0,31	0,25	0,13
P. Antec. 5 d (mm) - 2549107	0,50	0,58	0,33	0,46	0,46	0,35	0,28	0,28	0,13
P. Antec. 7 d (mm) - 2549107	0,42	0,54	0,29	0,44	0,42	0,28	0,25	0,20	0,08
P. Antec. 1 d (mm) - 2549128	0,63	0,64	0,51	0,61	0,16	0,55	0,02	0,38	0,03
P. Antec. 3 d (mm) - 2549128	0,71	0,70	0,51	0,61	0,34	0,56	0,17	0,47	0,10
P. Antec. 5 d (mm) - 2549128	0,46	0,40	0,26	0,34	0,23	0,32	0,13	0,35	0,05
P. Antec. 7 d (mm) - 2549128	0,40	0,39	0,21	0,32	0,29	0,27	0,17	0,29	0,02
P. Antec. 1 d (mm) - 2549077	0,60	0,31	0,09	0,19	0,37	0,01	0,38	0,71	0,20
P. Antec. 3 d (mm) - 2549077	0,68	0,34	0,11	0,21	0,39	0,04	0,40	0,82	0,21
P. Antec. 5 d (mm) - 2549077	0,63	0,35	0,07	0,19	0,50	-0,01	0,50	0,76	0,25
P. Antec. 7 d (mm) - 2549077	0,55	0,25	-0,03	0,09	0,50	-0,08	0,50	0,72	0,22
P. Antec. 1 d (mm) - 2549090	0,63	0,61	0,53	0,59	0,11	0,59	-0,03	0,38	0,02
P. Antec. 3 d (mm) - 2549090	0,67	0,64	0,45	0,55	0,34	0,49	0,19	0,47	0,11
P. Antec. 5 d (mm) - 2549090	0,55	0,46	0,27	0,37	0,31	0,34	0,20	0,44	0,07
P. Antec. 7 d (mm) - 2549090	0,45	0,38	0,15	0,28	0,35	0,22	0,24	0,38	0,04
P. Antec. 1 d (mm) - 2549006	0,59	0,30	0,09	0,18	0,34	0,02	0,37	0,72	0,19
P. Antec. 3 d (mm) - 2549006	0,73	0,33	0,14	0,24	0,28	0,09	0,32	0,90	0,17
P. Antec. 5 d (mm) - 2549006	0,74	0,34	0,14	0,24	0,29	0,09	0,32	0,91	0,18
P. Antec. 7 d (mm) - 2549006	0,68	0,35	0,08	0,20	0,46	0,03	0,47	0,84	0,22

	US - VAN (%)	US - Campo (%)	US - Água (%)	US - VAP (%)	US - VAN2 (%)	US - M/O (%)	US - AI (%)	US - Vila (%)	US - CT (%)
Amostra	-0,50	-0,53	-0,42	-0,59	-0,60	-0,59	-0,43	-0,59	-0,34
Ponto	0,08	0,04	0,17	-0,24	-0,23	-0,40	-0,05	-0,25	0,21
Evento	-0,41	-0,44	-0,32	-0,55	-0,56	-0,57	-0,37	-0,54	-0,25
Tc (min)	0,62	0,62	0,55	0,61	0,61	0,29	0,34	0,43	0,69
Tci (min)	0,39	0,40	0,29	0,54	0,52	0,31	0,18	0,33	0,42
Conc. NT (mg/L)	0,19	0,11	0,22	-0,20	-0,18	-0,31	0,00	-0,13	0,08
Conc. PT (mg/L)	0,38	0,33	0,44	0,02	0,05	-0,11	0,21	0,11	0,34
Conc. DOC (mg/L)	0,30	0,29	0,33	0,19	0,17	0,06	0,18	0,21	0,26
P. Acum. - Tc (mm)	0,37	0,40	0,44	0,18	0,23	0,23	0,39	0,39	0,31
P. Acum. - Tci (mm)	0,53	0,55	0,43	0,66	0,65	0,46	0,34	0,52	0,50
P. Antec. 1 d (mm)	0,55	0,57	0,50	0,57	0,57	0,39	0,39	0,52	0,52
P. Antec. 3 d (mm)	0,61	0,63	0,57	0,58	0,58	0,35	0,42	0,53	0,59
P. Antec. 5 d (mm)	0,25	0,25	0,21	0,26	0,25	0,15	0,15	0,20	0,24
P. Antec. 7 d (mm)	0,41	0,41	0,37	0,35	0,35	0,17	0,27	0,31	0,42
Qrio (m³/s)	0,48	0,51	0,48	0,44	0,44	0,36	0,41	0,46	0,48
Hidrog. sobe (2)/desce (1)	-0,13	-0,16	-0,09	-0,28	-0,26	-0,14	-0,08	-0,18	-0,20
ET real	0,47	0,48	0,51	0,30	0,31	0,22	0,39	0,36	0,44
A bacia (km²)	0,71	0,72	0,65	0,74	0,74	0,33	0,38	0,49	0,89
I (m/m)	0,98	0,98	0,89	0,89	0,91	0,62	0,69	0,83	0,90
TS - LVA (%)	0,61	0,65	0,54	0,75	0,78	0,62	0,45	0,63	0,60
TS - CA (%)	0,83	0,90	0,81	0,87	0,88	0,76	0,78	0,93	0,81
TS - PVAA (%)	0,59	0,48	0,52	0,18	0,18	-0,13	0,19	0,09	0,42
TS - LVED (%)	0,54	0,55	0,34	0,88	0,83	0,53	0,23	0,50	0,56
TS - Ctba (%)	0,39	0,27	0,40	-0,09	-0,08	-0,28	0,06	-0,07	0,21
TS - PVAE (%)	0,31	0,33	0,38	0,24	0,24	0,03	0,18	0,15	0,56
US - AUB (%)	0,16	0,14	0,26	-0,09	-0,05	0,03	0,17	0,01	0,05
US - VAN (%)	1,00	0,98	0,92	0,82	0,84	0,53	0,66	0,77	0,90
US - Campo (%)	0,98	1,00	0,94	0,83	0,85	0,62	0,76	0,86	0,89
US - Água (%)	0,92	0,94	1,00	0,66	0,68	0,50	0,77	0,81	0,82
US - VAP (%)	0,82	0,83	0,66	1,00	0,99	0,62	0,47	0,75	0,81
US - VAN2 (%)	0,84	0,85	0,68	0,99	1,00	0,62	0,48	0,75	0,83
US - M/O (%)	0,53	0,62	0,50	0,62	0,62	1,00	0,77	0,80	0,35
US - AI (%)	0,66	0,76	0,77	0,47	0,48	0,77	1,00	0,87	0,52
US - Vila (%)	0,77	0,86	0,81	0,75	0,75	0,80	0,87	1,00	0,66
US - CT (%)	0,90	0,89	0,82	0,81	0,83	0,35	0,52	0,66	1,00
US - AA (%)	0,39	0,50	0,52	0,44	0,42	0,49	0,52	0,75	0,35
US - Granja (%)	0,56	0,53	0,55	0,18	0,20	-0,09	0,39	0,22	0,56
US - CP (%)	0,64	0,67	0,51	0,84	0,87	0,47	0,35	0,67	0,74
US - SE (%)	0,79	0,71	0,68	0,61	0,59	0,16	0,29	0,35	0,71
US - AUM (%)	0,34	0,23	0,36	-0,10	-0,09	-0,26	0,02	-0,07	0,15
US - Lot. (%)	0,39	0,28	0,41	-0,09	-0,08	-0,28	0,08	-0,07	0,22
US - Lixão (%)	0,36	0,34	0,41	0,00	0,01	-0,19	0,24	0,00	0,40
P. Antec. 1 d (mm) - Thiessen	0,60	0,60	0,53	0,64	0,64	0,34	0,33	0,46	0,65
P. Antec. 3 d (mm) - Thiessen	0,69	0,69	0,63	0,63	0,63	0,32	0,42	0,50	0,72
P. Antec. 5 d (mm) - Thiessen	0,58	0,58	0,54	0,50	0,50	0,26	0,37	0,41	0,62
P. Antec. 7 d (mm) - Thiessen	0,54	0,53	0,49	0,45	0,45	0,21	0,33	0,36	0,56
P. Antec. 1 d (mm) - 2549100	0,45	0,47	0,47	0,42	0,42	0,19	0,28	0,35	0,51
P. Antec. 3 d (mm) - 2549100	0,53	0,53	0,52	0,47	0,47	0,21	0,32	0,39	0,59
P. Antec. 5 d (mm) - 2549100	0,48	0,49	0,49	0,41	0,41	0,18	0,30	0,34	0,56
P. Antec. 7 d (mm) - 2549100	0,46	0,46	0,45	0,36	0,36	0,12	0,27	0,28	0,53
P. Antec. 1 d (mm) - 2549107	0,51	0,50	0,40	0,60	0,59	0,32	0,26	0,38	0,55
P. Antec. 3 d (mm) - 2549107	0,67	0,65	0,58	0,54	0,54	0,26	0,41	0,43	0,63
P. Antec. 5 d (mm) - 2549107	0,61	0,59	0,55	0,47	0,47	0,24	0,39	0,39	0,59
P. Antec. 7 d (mm) - 2549107	0,57	0,55	0,50	0,42	0,44	0,21	0,36	0,37	0,54
P. Antec. 1 d (mm) - 2549128	0,60	0,61	0,54	0,65	0,64	0,37	0,37	0,52	0,64
P. Antec. 3 d (mm) - 2549128	0,69	0,69	0,62	0,66	0,65	0,33	0,41	0,51	0,73
P. Antec. 5 d (mm) - 2549128	0,40	0,41	0,37	0,37	0,36	0,17	0,25	0,28	0,45
P. Antec. 7 d (mm) - 2549128	0,41	0,40	0,36	0,34	0,34	0,15	0,25	0,26	0,44
P. Antec. 1 d (mm) - 2549077	0,37	0,35	0,44	0,12	0,14	-0,09	0,20	0,09	0,52
P. Antec. 3 d (mm) - 2549077	0,40	0,39	0,48	0,15	0,16	-0,09	0,22	0,10	0,57
P. Antec. 5 d (mm) - 2549077	0,43	0,40	0,49	0,12	0,13	-0,13	0,23	0,08	0,57
P. Antec. 7 d (mm) - 2549077	0,34	0,31	0,41	0,03	0,04	-0,20	0,17	-0,01	0,48
P. Antec. 1 d (mm) - 2549090	0,56	0,58	0,49	0,65	0,64	0,42	0,35	0,50	0,60
P. Antec. 3 d (mm) - 2549090	0,64	0,64	0,58	0,59	0,58	0,29	0,39	0,46	0,68
P. Antec. 5 d (mm) - 2549090	0,47	0,47	0,43	0,40	0,39	0,15	0,27	0,28	0,54
P. Antec. 7 d (mm) - 2549090	0,42	0,39	0,37	0,30	0,30	0,05	0,22	0,20	0,47
P. Antec. 1 d (mm) - 2549006	0,36	0,34	0,42	0,13	0,14	-0,08	0,18	0,09	0,50
P. Antec. 3 d (mm) - 2549006	0,38	0,37	0,45	0,18	0,19	-0,05	0,20	0,12	0,57
P. Antec. 5 d (mm) - 2549006	0,38	0,37	0,45	0,18	0,19	-0,05	0,20	0,12	0,58
P. Antec. 7 d (mm) - 2549006	0,43	0,40	0,48	0,14	0,16	-0,11	0,22	0,09	0,59

	US - AA (%)	US - Granja (%)	US - CP (%)	US - SE (%)	US - AUM (%)	US - Lot. (%)	US - Lixão (%)	P. Antec. 1 d (mm) - Thiessen	P. Antec. 3 d (mm) - Thiessen
Amostra	-0,43	-0,02	-0,50	-0,23	0,11	0,13	0,16	-0,29	-0,30
Ponto	-0,32	0,43	-0,28	0,31	0,47	0,54	0,56	0,05	0,12
Evento	-0,43	0,07	-0,47	-0,14	0,19	0,22	0,23	-0,23	-0,22
Tc (min)	0,16	0,32	0,48	0,54	0,07	0,12	0,29	0,39	0,42
Tcl (min)	0,13	0,05	0,41	0,37	-0,09	-0,08	0,01	0,26	0,43
Conc. NT (mg/L)	-0,19	0,42	-0,28	0,31	0,62	0,64	0,32	-0,11	0,05
Conc. PT (mg/L)	-0,03	0,48	-0,04	0,39	0,52	0,56	0,40	0,17	0,25
Conc. DOC (mg/L)	0,23	0,21	0,08	0,30	0,22	0,24	0,18	0,16	0,20
P. Acum. - Tc (mm)	0,34	0,29	0,25	0,10	0,12	0,15	0,20	0,10	0,06
P. Acum. - Tcl (mm)	0,32	0,10	0,54	0,38	-0,09	-0,09	-0,01	0,50	0,49
P. Antec. 1 d (mm)	0,32	0,21	0,48	0,38	0,00	0,01	0,13	0,56	0,41
P. Antec. 3 d (mm)	0,30	0,32	0,48	0,47	0,08	0,11	0,21	0,49	0,67
P. Antec. 5 d (mm)	0,11	0,12	0,20	0,21	0,01	0,03	0,08	0,37	0,43
P. Antec. 7 d (mm)	0,16	0,30	0,29	0,32	0,07	0,12	0,22	0,32	0,56
Qrio (m³/s)	0,31	0,26	0,36	0,35	0,04	0,08	0,18	0,39	0,34
Hidrog. sobe (2)/desce (1)	-0,10	0,02	-0,25	-0,10	0,18	0,18	0,02	-0,27	-0,36
ET real	0,23	0,40	0,21	0,39	0,22	0,27	0,31	0,24	0,51
A bacía (km²)	0,15	0,34	0,61	0,66	0,03	0,09	0,34	0,68	0,68
I (m/m)	0,46	0,45	0,74	0,73	0,19	0,23	0,27	0,63	0,70
TS - LVA (%)	0,39	0,07	0,67	0,33	-0,21	-0,21	-0,03	0,52	0,49
TS - CA (%)	0,64	0,27	0,79	0,44	-0,16	-0,12	0,13	0,58	0,60
TS - PVAA (%)	-0,15	0,72	-0,07	0,76	0,81	0,87	0,57	0,20	0,37
TS - LVED (%)	0,22	-0,06	0,64	0,53	-0,23	-0,24	-0,14	0,56	0,52
TS - Ctba (%)	-0,21	0,58	-0,25	0,65	0,97	0,98	0,50	0,05	0,20
TS - PVAE (%)	-0,03	0,24	0,22	0,36	0,07	0,13	0,36	0,41	0,45
US - AUB (%)	-0,07	0,24	-0,18	0,19	0,34	0,36	0,28	0,07	0,13
US - VAN (%)	0,39	0,56	0,64	0,79	0,34	0,39	0,36	0,60	0,69
US - Campo (%)	0,50	0,53	0,67	0,71	0,23	0,28	0,34	0,60	0,69
US - Água (%)	0,52	0,55	0,51	0,68	0,36	0,41	0,41	0,53	0,63
US - VAP (%)	0,44	0,18	0,84	0,61	-0,10	-0,09	0,00	0,64	0,63
US - VAN2 (%)	0,42	0,20	0,87	0,59	-0,09	-0,08	0,01	0,64	0,63
US - M/O (%)	0,49	-0,09	0,47	0,16	-0,26	-0,28	-0,19	0,34	0,32
US - AI (%)	0,52	0,39	0,35	0,29	0,02	0,08	0,24	0,33	0,42
US - Vila (%)	0,75	0,22	0,67	0,35	-0,07	-0,07	0,00	0,46	0,50
US - CT (%)	0,35	0,56	0,74	0,71	0,15	0,22	0,40	0,65	0,72
US - AA (%)	1,00	0,08	0,45	-0,01	-0,19	-0,21	-0,15	0,17	0,21
US - Granja (%)	0,08	1,00	0,07	0,54	0,45	0,62	0,72	0,22	0,43
US - CP (%)	0,45	0,07	1,00	0,31	-0,24	-0,25	-0,11	0,50	0,47
US - SE (%)	-0,01	0,54	0,31	1,00	0,64	0,67	0,47	0,53	0,62
US - AUM (%)	-0,19	0,45	-0,24	0,64	1,00	0,97	0,39	0,03	0,15
US - Lot. (%)	-0,21	0,62	-0,25	0,67	0,97	1,00	0,54	0,06	0,21
US - Lixão (%)	-0,15	0,72	-0,11	0,47	0,39	0,54	1,00	0,19	0,31
P. Antec. 1 d (mm) - Thiessen	0,17	0,22	0,50	0,53	0,03	0,06	0,19	1,00	0,81
P. Antec. 3 d (mm) - Thiessen	0,21	0,43	0,47	0,62	0,15	0,21	0,31	0,81	1,00
P. Antec. 5 d (mm) - Thiessen	0,17	0,40	0,37	0,52	0,14	0,20	0,31	0,67	0,89
P. Antec. 7 d (mm) - Thiessen	0,15	0,41	0,34	0,46	0,14	0,20	0,28	0,59	0,83
P. Antec. 1 d (mm) - 2549100	0,15	0,21	0,35	0,41	0,11	0,12	0,23	0,81	0,77
P. Antec. 3 d (mm) - 2549100	0,16	0,31	0,38	0,48	0,14	0,17	0,29	0,72	0,88
P. Antec. 5 d (mm) - 2549100	0,14	0,31	0,32	0,45	0,14	0,18	0,29	0,68	0,87
P. Antec. 7 d (mm) - 2549100	0,07	0,37	0,26	0,46	0,18	0,23	0,32	0,65	0,85
P. Antec. 1 d (mm) - 2549107	0,12	0,14	0,46	0,45	-0,04	-0,02	0,11	0,89	0,62
P. Antec. 3 d (mm) - 2549107	0,15	0,55	0,36	0,60	0,23	0,32	0,33	0,64	0,85
P. Antec. 5 d (mm) - 2549107	0,15	0,52	0,31	0,54	0,21	0,29	0,34	0,53	0,76
P. Antec. 7 d (mm) - 2549107	0,15	0,49	0,30	0,46	0,18	0,25	0,34	0,47	0,71
P. Antec. 1 d (mm) - 2549128	0,25	0,21	0,53	0,49	0,01	0,03	0,14	0,86	0,80
P. Antec. 3 d (mm) - 2549128	0,21	0,40	0,49	0,62	0,13	0,18	0,29	0,77	0,97
P. Antec. 5 d (mm) - 2549128	0,10	0,29	0,26	0,38	0,08	0,13	0,22	0,51	0,76
P. Antec. 7 d (mm) - 2549128	0,10	0,36	0,24	0,37	0,11	0,17	0,25	0,37	0,64
P. Antec. 1 d (mm) - 2549077	-0,10	0,47	0,06	0,44	0,31	0,39	0,55	0,55	0,55
P. Antec. 3 d (mm) - 2549077	-0,10	0,49	0,08	0,49	0,32	0,41	0,59	0,50	0,57
P. Antec. 5 d (mm) - 2549077	-0,13	0,60	0,03	0,53	0,40	0,51	0,66	0,43	0,55
P. Antec. 7 d (mm) - 2549077	-0,18	0,59	-0,06	0,47	0,40	0,51	0,64	0,35	0,48
P. Antec. 1 d (mm) - 2549090	0,25	0,14	0,53	0,47	-0,03	-0,02	0,09	0,84	0,69
P. Antec. 3 d (mm) - 2549090	0,17	0,40	0,44	0,58	0,14	0,20	0,29	0,70	0,88
P. Antec. 5 d (mm) - 2549090	0,06	0,38	0,27	0,47	0,15	0,21	0,30	0,55	0,80
P. Antec. 7 d (mm) - 2549090	0,01	0,43	0,20	0,41	0,17	0,25	0,34	0,43	0,71
P. Antec. 1 d (mm) - 2549006	-0,09	0,41	0,07	0,43	0,30	0,37	0,50	0,29	0,33
P. Antec. 3 d (mm) - 2549006	-0,08	0,39	0,13	0,46	0,26	0,33	0,51	0,43	0,49
P. Antec. 5 d (mm) - 2549006	-0,08	0,40	0,13	0,46	0,26	0,33	0,52	0,41	0,48
P. Antec. 7 d (mm) - 2549006	-0,12	0,56	0,06	0,53	0,38	0,48	0,60	0,34	0,48

	P. Antec. 5 d (mm) - Thiessen	P. Antec. 7 d (mm) - Thiessen	P. Antec. 1 d (mm) - 2549100	P. Antec. 3 d (mm) - 2549100	P. Antec. 5 d (mm) - 2549100	P. Antec. 7 d (mm) - 2549100	P. Antec. 1 d (mm) - 2549107
Amostra	-0,41	-0,41	-0,06	-0,07	-0,21	-0,15	-0,34
Ponto	0,09	0,06	0,16	0,19	0,17	0,22	-0,03
Evento	-0,35	-0,35	0,01	0,00	-0,14	-0,08	-0,31
Tc (min)	0,34	0,22	0,19	0,27	0,21	0,12	0,36
Tcl (min)	0,36	0,25	0,14	0,29	0,25	0,20	0,23
Conc. NT (mg/L)	0,05	0,08	-0,01	0,01	0,01	0,08	-0,15
Conc. PT (mg/L)	0,25	0,24	0,27	0,24	0,29	0,33	0,07
Conc. DOC (mg/L)	0,18	0,18	0,16	0,19	0,18	0,18	0,14
P. Acum. - Tc (mm)	0,02	0,00	0,05	-0,02	-0,03	-0,07	0,08
P. Acum. - Tcl (mm)	0,35	0,24	0,28	0,30	0,23	0,15	0,47
P. Antec. 1 d (mm)	0,33	0,23	0,29	0,23	0,19	0,09	0,52
P. Antec. 3 d (mm)	0,61	0,50	0,37	0,51	0,49	0,37	0,35
P. Antec. 5 d (mm)	0,51	0,48	0,31	0,34	0,44	0,42	0,31
P. Antec. 7 d (mm)	0,77	0,78	0,21	0,36	0,57	0,55	0,28
Qrlo (m³/s)	0,45	0,48	0,15	0,15	0,30	0,34	0,45
Hidrog. sobe (2)/desce (1)	-0,40	-0,34	-0,22	-0,28	-0,32	-0,27	-0,20
ET real	0,68	0,68	0,32	0,48	0,64	0,64	0,14
A bacia (km²)	0,59	0,52	0,54	0,60	0,58	0,56	0,59
I (m/m)	0,58	0,52	0,47	0,53	0,48	0,45	0,55
TS - LVA (%)	0,39	0,32	0,39	0,40	0,37	0,29	0,47
TS - CA (%)	0,49	0,44	0,43	0,47	0,42	0,36	0,50
TS - PVAA (%)	0,34	0,33	0,17	0,24	0,24	0,30	0,15
TS - LVED (%)	0,41	0,34	0,33	0,38	0,33	0,29	0,56
TS - Ctba (%)	0,20	0,19	0,12	0,16	0,18	0,22	-0,03
TS - PVAE (%)	0,42	0,36	0,42	0,47	0,50	0,51	0,30
US - AUB (%)	0,12	0,08	0,15	0,14	0,17	0,15	0,00
US - VAN (%)	0,58	0,54	0,45	0,53	0,48	0,46	0,51
US - Campo (%)	0,58	0,53	0,47	0,53	0,49	0,46	0,50
US - Água (%)	0,54	0,49	0,47	0,52	0,49	0,45	0,40
US - VAP (%)	0,50	0,45	0,42	0,47	0,41	0,36	0,60
US - VANZ (%)	0,50	0,45	0,42	0,47	0,41	0,36	0,59
US - M/O (%)	0,26	0,21	0,19	0,21	0,18	0,12	0,32
US - AJ (%)	0,37	0,33	0,28	0,32	0,30	0,27	0,26
US - Vila (%)	0,41	0,36	0,35	0,39	0,34	0,28	0,38
US - CT (%)	0,62	0,56	0,51	0,59	0,56	0,53	0,55
US - AA (%)	0,17	0,15	0,15	0,16	0,14	0,07	0,12
US - Granja (%)	0,40	0,41	0,21	0,31	0,31	0,37	0,14
US - CP (%)	0,37	0,34	0,35	0,38	0,32	0,26	0,46
US - SE (%)	0,52	0,46	0,41	0,48	0,45	0,46	0,45
US - AUM (%)	0,14	0,14	0,11	0,14	0,14	0,18	-0,04
US - Lot. (%)	0,20	0,20	0,12	0,17	0,18	0,23	-0,02
US - Lixão (%)	0,31	0,28	0,23	0,29	0,29	0,32	0,11
P. Antec. 1 d (mm) - Thiessen	0,67	0,59	0,81	0,72	0,68	0,65	0,89
P. Antec. 3 d (mm) - Thiessen	0,89	0,83	0,77	0,88	0,87	0,85	0,62
P. Antec. 5 d (mm) - Thiessen	1,00	0,95	0,61	0,73	0,89	0,90	0,54
P. Antec. 7 d (mm) - Thiessen	0,95	1,00	0,54	0,66	0,84	0,89	0,51
P. Antec. 1 d (mm) - 2549100	0,61	0,54	1,00	0,90	0,81	0,75	0,52
P. Antec. 3 d (mm) - 2549100	0,73	0,66	0,90	1,00	0,91	0,84	0,41
P. Antec. 5 d (mm) - 2549100	0,89	0,84	0,81	0,91	1,00	0,96	0,43
P. Antec. 7 d (mm) - 2549100	0,90	0,89	0,75	0,84	0,96	1,00	0,47
P. Antec. 1 d (mm) - 2549107	0,54	0,51	0,52	0,41	0,43	0,47	1,00
P. Antec. 3 d (mm) - 2549107	0,83	0,81	0,46	0,54	0,61	0,68	0,61
P. Antec. 5 d (mm) - 2549107	0,90	0,88	0,38	0,47	0,65	0,72	0,53
P. Antec. 7 d (mm) - 2549107	0,86	0,90	0,38	0,46	0,63	0,70	0,46
P. Antec. 1 d (mm) - 2549128	0,63	0,53	0,72	0,72	0,65	0,57	0,65
P. Antec. 3 d (mm) - 2549128	0,86	0,77	0,68	0,84	0,83	0,79	0,59
P. Antec. 5 d (mm) - 2549128	0,94	0,90	0,41	0,56	0,79	0,78	0,44
P. Antec. 7 d (mm) - 2549128	0,87	0,92	0,26	0,43	0,69	0,74	0,36
P. Antec. 1 d (mm) - 2549077	0,48	0,42	0,56	0,55	0,54	0,53	0,43
P. Antec. 3 d (mm) - 2549077	0,54	0,47	0,55	0,57	0,59	0,61	0,37
P. Antec. 5 d (mm) - 2549077	0,55	0,49	0,46	0,51	0,55	0,58	0,32
P. Antec. 7 d (mm) - 2549077	0,48	0,46	0,41	0,45	0,49	0,57	0,25
P. Antec. 1 d (mm) - 2549090	0,56	0,42	0,62	0,59	0,54	0,47	0,70
P. Antec. 3 d (mm) - 2549090	0,77	0,65	0,58	0,71	0,70	0,65	0,52
P. Antec. 5 d (mm) - 2549090	0,91	0,84	0,44	0,58	0,75	0,74	0,46
P. Antec. 7 d (mm) - 2549090	0,87	0,88	0,33	0,48	0,68	0,72	0,37
P. Antec. 1 d (mm) - 2549006	0,27	0,22	0,24	0,35	0,33	0,30	0,18
P. Antec. 3 d (mm) - 2549006	0,45	0,38	0,44	0,52	0,54	0,54	0,31
P. Antec. 5 d (mm) - 2549006	0,46	0,38	0,43	0,51	0,54	0,54	0,30
P. Antec. 7 d (mm) - 2549006	0,49	0,46	0,34	0,44	0,50	0,57	0,26

	P. Antec. 3 d (mm) - 2549107	P. Antec. 5 d (mm) - 2549107	P. Antec. 7 d (mm) - 2549107	P. Antec. 1 d (mm) - 2549128	P. Antec. 3 d (mm) - 2549128	P. Antec. 5 d (mm) - 2549128	P. Antec. 7 d (mm) - 2549128	P. Antec. 1 d (mm) - 2549077
Amostra	-0,37	-0,41	-0,45	-0,30	-0,28	-0,44	-0,47	0,20
Ponto	0,09	0,10	0,04	-0,01	0,11	0,02	0,02	0,60
Evento	-0,30	-0,35	-0,40	-0,24	-0,21	-0,41	-0,43	0,26
Tc (min)	0,39	0,34	0,29	0,36	0,47	0,27	0,19	0,34
Tci (min)	0,36	0,31	0,25	0,26	0,48	0,34	0,22	0,01
Conc. NT (mg/L)	0,24	0,20	0,16	-0,14	-0,01	-0,02	0,04	0,24
Conc. PT (mg/L)	0,30	0,28	0,22	0,14	0,20	0,12	0,14	0,38
Conc. DOC (mg/L)	0,20	0,19	0,18	0,05	0,18	0,11	0,13	0,17
P. Acum. - Tc (mm)	0,15	0,10	0,10	0,13	0,02	-0,03	-0,05	0,10
P. Acum. - Tci (mm)	0,47	0,34	0,27	0,52	0,48	0,25	0,10	0,07
P. Antec. 1 d (mm)	0,42	0,35	0,30	0,61	0,41	0,26	0,13	0,18
P. Antec. 3 d (mm)	0,58	0,53	0,52	0,63	0,70	0,58	0,44	0,22
P. Antec. 5 d (mm)	0,39	0,44	0,43	0,32	0,41	0,50	0,44	0,07
P. Antec. 7 d (mm)	0,58	0,71	0,80	0,35	0,55	0,85	0,87	0,17
Qrio (m³/s)	0,39	0,46	0,43	0,36	0,35	0,43	0,49	0,23
Hidrog. sobe (2)/desce (1)	-0,31	-0,32	-0,30	-0,29	-0,36	-0,42	-0,33	-0,11
ET real	0,46	0,61	0,61	0,22	0,52	0,66	0,74	0,28
A bacia (km²)	0,52	0,50	0,42	0,63	0,71	0,46	0,40	0,60
I (m/m)	0,64	0,58	0,54	0,64	0,70	0,40	0,39	0,31
TS - LVA (%)	0,38	0,33	0,29	0,51	0,51	0,26	0,21	0,09
TS - CA (%)	0,50	0,46	0,44	0,61	0,61	0,34	0,32	0,19
TS - PVAA (%)	0,52	0,46	0,42	0,16	0,34	0,23	0,29	0,37
TS - LVED (%)	0,41	0,35	0,28	0,55	0,56	0,32	0,27	0,01
TS - Ctba (%)	0,31	0,28	0,25	0,02	0,17	0,13	0,17	0,38
TS - PVAE (%)	0,25	0,28	0,20	0,38	0,47	0,35	0,29	0,71
US - AUB (%)	0,13	0,13	0,08	0,03	0,10	0,05	0,02	0,20
US - VAN (%)	0,67	0,61	0,57	0,60	0,69	0,40	0,41	0,37
US - Campo (%)	0,65	0,59	0,55	0,61	0,69	0,41	0,40	0,35
US - Água (%)	0,58	0,55	0,50	0,54	0,62	0,37	0,36	0,44
US - VAP (%)	0,54	0,47	0,42	0,65	0,66	0,37	0,34	0,12
US - VAN2 (%)	0,54	0,47	0,44	0,64	0,65	0,36	0,34	0,14
US - M/O (%)	0,26	0,24	0,21	0,37	0,33	0,17	0,15	-0,09
US - Al (%)	0,41	0,39	0,36	0,37	0,41	0,25	0,25	0,20
US - Vila (%)	0,43	0,39	0,37	0,52	0,51	0,28	0,26	0,09
US - CT (%)	0,63	0,59	0,54	0,64	0,73	0,45	0,44	0,52
US - AA (%)	0,15	0,15	0,15	0,25	0,21	0,10	0,10	-0,10
US - Granja (%)	0,55	0,52	0,49	0,21	0,40	0,29	0,36	0,47
US - CP (%)	0,36	0,31	0,30	0,53	0,49	0,26	0,24	0,06
US - SE (%)	0,60	0,54	0,46	0,49	0,62	0,38	0,37	0,44
US - AUM (%)	0,23	0,21	0,18	0,01	0,13	0,08	0,11	0,31
US - Lot. (%)	0,32	0,29	0,25	0,03	0,18	0,13	0,17	0,39
US - Lixão (%)	0,33	0,34	0,34	0,14	0,29	0,22	0,25	0,55
P. Antec. 1 d (mm) - Thiessen	0,64	0,53	0,47	0,86	0,77	0,51	0,37	0,55
P. Antec. 3 d (mm) - Thiessen	0,85	0,76	0,71	0,80	0,97	0,76	0,64	0,55
P. Antec. 5 d (mm) - Thiessen	0,83	0,90	0,86	0,63	0,86	0,94	0,87	0,48
P. Antec. 7 d (mm) - Thiessen	0,81	0,88	0,90	0,53	0,77	0,90	0,92	0,42
P. Antec. 1 d (mm) - 2549100	0,46	0,38	0,38	0,72	0,68	0,41	0,26	0,56
P. Antec. 3 d (mm) - 2549100	0,54	0,47	0,46	0,72	0,84	0,56	0,43	0,55
P. Antec. 5 d (mm) - 2549100	0,61	0,65	0,63	0,65	0,83	0,79	0,69	0,54
P. Antec. 7 d (mm) - 2549100	0,68	0,72	0,70	0,57	0,79	0,78	0,74	0,53
P. Antec. 1 d (mm) - 2549107	0,61	0,53	0,46	0,65	0,59	0,44	0,36	0,43
P. Antec. 3 d (mm) - 2549107	1,00	0,91	0,84	0,58	0,79	0,71	0,68	0,44
P. Antec. 5 d (mm) - 2549107	0,91	1,00	0,94	0,46	0,71	0,81	0,83	0,41
P. Antec. 7 d (mm) - 2549107	0,84	0,94	1,00	0,39	0,64	0,79	0,87	0,33
P. Antec. 1 d (mm) - 2549128	0,58	0,46	0,39	1,00	0,82	0,53	0,35	0,42
P. Antec. 3 d (mm) - 2549128	0,79	0,71	0,64	0,82	1,00	0,76	0,64	0,50
P. Antec. 5 d (mm) - 2549128	0,71	0,81	0,79	0,53	0,76	1,00	0,93	0,36
P. Antec. 7 d (mm) - 2549128	0,68	0,83	0,87	0,35	0,64	0,93	1,00	0,27
P. Antec. 1 d (mm) - 2549077	0,44	0,41	0,33	0,42	0,50	0,36	0,27	1,00
P. Antec. 3 d (mm) - 2549077	0,43	0,44	0,38	0,39	0,54	0,41	0,35	0,92
P. Antec. 5 d (mm) - 2549077	0,49	0,52	0,46	0,32	0,52	0,43	0,40	0,88
P. Antec. 7 d (mm) - 2549077	0,46	0,49	0,43	0,25	0,44	0,35	0,38	0,82
P. Antec. 1 d (mm) - 2549090	0,50	0,40	0,28	0,91	0,72	0,46	0,25	0,43
P. Antec. 3 d (mm) - 2549090	0,75	0,65	0,54	0,78	0,91	0,68	0,50	0,52
P. Antec. 5 d (mm) - 2549090	0,77	0,81	0,74	0,60	0,81	0,93	0,80	0,46
P. Antec. 7 d (mm) - 2549090	0,75	0,83	0,85	0,44	0,69	0,90	0,89	0,39
P. Antec. 1 d (mm) - 2549006	0,20	0,19	0,11	0,35	0,38	0,22	0,18	0,65
P. Antec. 3 d (mm) - 2549006	0,29	0,32	0,25	0,41	0,54	0,37	0,33	0,78
P. Antec. 5 d (mm) - 2549006	0,27	0,32	0,26	0,38	0,52	0,38	0,33	0,76
P. Antec. 7 d (mm) - 2549006	0,41	0,45	0,37	0,32	0,50	0,39	0,41	0,72

	P. Antec. 3 d (mm) - 2549077	P. Antec. 5 d (mm) - 2549077	P. Antec. 7 d (mm) - 2549077	P. Antec. 1 d (mm) - 2549090	P. Antec. 3 d (mm) - 2549090	P. Antec. 5 d (mm) - 2549090	P. Antec. 7 d (mm) - 2549090	P. Antec. 1 d (mm) - 2549006
Amostra	0,23	0,23	0,32	-0,31	-0,22	-0,30	-0,29	0,17
Ponto	0,66	0,69	0,78	-0,04	0,14	0,13	0,15	0,58
Evento	0,30	0,31	0,40	-0,27	-0,15	-0,25	-0,23	0,25
Tc (min)	0,31	0,35	0,19	0,42	0,52	0,40	0,29	0,48
Tci (min)	0,10	0,12	0,01	0,32	0,51	0,44	0,33	0,10
Conc. NT (mg/L)	0,22	0,33	0,39	-0,23	0,01	0,05	0,11	0,22
Conc. PT (mg/L)	0,40	0,45	0,47	0,05	0,15	0,15	0,16	0,37
Conc. DOC (mg/L)	0,18	0,19	0,20	0,07	0,15	0,12	0,10	0,16
P. Acum. - Tc (mm)	0,02	0,06	0,00	0,12	0,07	0,00	-0,01	0,13
P. Acum. - Tci (mm)	-0,03	-0,04	-0,15	0,62	0,59	0,37	0,24	0,21
P. Antec. 1 d (mm)	0,04	0,06	-0,08	0,66	0,52	0,37	0,26	0,32
P. Antec. 3 d (mm)	0,17	0,18	0,04	0,60	0,77	0,67	0,58	0,40
P. Antec. 5 d (mm)	0,07	0,09	0,03	0,31	0,39	0,48	0,46	0,11
P. Antec. 7 d (mm)	0,20	0,24	0,16	0,28	0,51	0,78	0,86	0,23
Qrio (m³/s)	0,27	0,23	0,25	0,41	0,32	0,38	0,37	0,20
Hidrog. sobe (2)/desce (1)	-0,09	-0,08	-0,03	-0,31	-0,37	-0,44	-0,37	0,00
ET real	0,34	0,36	0,33	0,11	0,32	0,50	0,50	0,23
A bacía (km²)	0,68	0,63	0,55	0,63	0,67	0,55	0,45	0,59
I (m/m)	0,34	0,35	0,25	0,61	0,64	0,46	0,38	0,30
TS - LVA (%)	0,11	0,07	-0,03	0,53	0,45	0,27	0,15	0,09
TS - CA (%)	0,21	0,19	0,09	0,59	0,55	0,37	0,28	0,18
TS - PVAA (%)	0,39	0,50	0,50	0,11	0,34	0,31	0,35	0,34
TS - LVED (%)	0,04	-0,01	-0,08	0,59	0,49	0,34	0,22	0,02
TS - Ctba (%)	0,40	0,50	0,50	-0,03	0,19	0,20	0,24	0,37
TS - PVAE (%)	0,82	0,76	0,72	0,38	0,47	0,44	0,38	0,72
US - AUB (%)	0,21	0,25	0,22	0,02	0,11	0,07	0,04	0,19
US - VAN (%)	0,40	0,43	0,34	0,56	0,64	0,47	0,42	0,36
US - Campo (%)	0,39	0,40	0,31	0,58	0,64	0,47	0,39	0,34
US - Água (%)	0,48	0,49	0,41	0,49	0,58	0,43	0,37	0,42
US - VAP (%)	0,15	0,12	0,03	0,65	0,59	0,40	0,30	0,13
US - VANZ (%)	0,16	0,13	0,04	0,64	0,58	0,39	0,30	0,14
US - M/O (%)	-0,09	-0,13	-0,20	0,42	0,29	0,15	0,05	-0,08
US - Al (%)	0,22	0,23	0,17	0,35	0,39	0,27	0,22	0,18
US - Vila (%)	0,10	0,08	-0,01	0,50	0,46	0,28	0,20	0,09
US - CT (%)	0,57	0,57	0,48	0,60	0,68	0,54	0,47	0,50
US - AA (%)	-0,10	-0,13	-0,18	0,25	0,17	0,06	0,01	-0,09
US - Granja (%)	0,49	0,60	0,59	0,14	0,40	0,38	0,43	0,41
US - CP (%)	0,08	0,03	-0,06	0,53	0,44	0,27	0,20	0,07
US - SE (%)	0,49	0,53	0,47	0,47	0,58	0,47	0,41	0,43
US - AUM (%)	0,32	0,40	0,40	-0,03	0,14	0,15	0,17	0,30
US - Lot. (%)	0,41	0,51	0,51	-0,02	0,20	0,21	0,25	0,37
US - Lixão (%)	0,59	0,66	0,64	0,09	0,29	0,30	0,34	0,50
P. Antec. 1 d (mm) - Thiessen	0,50	0,43	0,35	0,84	0,70	0,55	0,43	0,29
P. Antec. 3 d (mm) - Thiessen	0,57	0,55	0,48	0,69	0,88	0,80	0,71	0,33
P. Antec. 5 d (mm) - Thiessen	0,54	0,55	0,48	0,56	0,77	0,91	0,87	0,27
P. Antec. 7 d (mm) - Thiessen	0,47	0,49	0,46	0,42	0,65	0,84	0,88	0,22
P. Antec. 1 d (mm) - 2549100	0,55	0,46	0,41	0,62	0,58	0,44	0,33	0,24
P. Antec. 3 d (mm) - 2549100	0,57	0,51	0,45	0,59	0,71	0,58	0,48	0,35
P. Antec. 5 d (mm) - 2549100	0,59	0,55	0,49	0,54	0,70	0,75	0,68	0,33
P. Antec. 7 d (mm) - 2549100	0,61	0,58	0,57	0,47	0,65	0,74	0,72	0,30
P. Antec. 1 d (mm) - 2549107	0,37	0,32	0,25	0,70	0,52	0,46	0,37	0,18
P. Antec. 3 d (mm) - 2549107	0,43	0,49	0,46	0,50	0,75	0,77	0,75	0,20
P. Antec. 5 d (mm) - 2549107	0,44	0,52	0,49	0,40	0,65	0,81	0,83	0,19
P. Antec. 7 d (mm) - 2549107	0,38	0,46	0,43	0,28	0,54	0,74	0,85	0,11
P. Antec. 1 d (mm) - 2549128	0,39	0,32	0,25	0,91	0,78	0,60	0,44	0,35
P. Antec. 3 d (mm) - 2549128	0,54	0,52	0,44	0,72	0,91	0,81	0,69	0,38
P. Antec. 5 d (mm) - 2549128	0,41	0,43	0,35	0,46	0,68	0,93	0,90	0,22
P. Antec. 7 d (mm) - 2549128	0,35	0,40	0,38	0,25	0,50	0,80	0,89	0,18
P. Antec. 1 d (mm) - 2549077	0,92	0,88	0,82	0,43	0,52	0,46	0,39	0,65
P. Antec. 3 d (mm) - 2549077	1,00	0,97	0,92	0,37	0,54	0,52	0,46	0,63
P. Antec. 5 d (mm) - 2549077	0,97	1,00	0,95	0,27	0,50	0,53	0,52	0,62
P. Antec. 7 d (mm) - 2549077	0,92	0,95	1,00	0,18	0,41	0,45	0,47	0,56
P. Antec. 1 d (mm) - 2549090	0,37	0,27	0,18	1,00	0,80	0,58	0,38	0,35
P. Antec. 3 d (mm) - 2549090	0,54	0,50	0,41	0,80	1,00	0,85	0,70	0,43
P. Antec. 5 d (mm) - 2549090	0,52	0,53	0,45	0,58	0,85	1,00	0,93	0,34
P. Antec. 7 d (mm) - 2549090	0,46	0,52	0,47	0,38	0,70	0,93	1,00	0,29
P. Antec. 1 d (mm) - 2549006	0,63	0,62	0,56	0,35	0,43	0,34	0,29	1,00
P. Antec. 3 d (mm) - 2549006	0,89	0,83	0,79	0,41	0,54	0,47	0,42	0,81
P. Antec. 5 d (mm) - 2549006	0,90	0,86	0,80	0,38	0,52	0,49	0,44	0,77
P. Antec. 7 d (mm) - 2549006	0,87	0,88	0,88	0,28	0,48	0,50	0,51	0,77

	P. Antec. 3 d (mm) - 2549006	P. Antec. 5 d (mm) - 2549006	P. Antec. 7 d (mm) - 2549006
Amostra	0,24	0,24	0,25
Ponto	0,64	0,64	0,70
Evento	0,31	0,31	0,33
Tc (min)	0,34	0,37	0,28
Tci (min)	0,11	0,21	0,10
Conc. NT (mg/L)	0,11	0,11	0,30
Conc. PT (mg/L)	0,37	0,36	0,48
Conc. DOC (mg/L)	0,15	0,15	0,16
P. Acum. - Tc (mm)	-0,02	-0,01	0,01
P. Acum. - Tci (mm)	0,03	0,03	-0,04
P. Antec. 1 d (mm)	0,08	0,07	0,01
P. Antec. 3 d (mm)	0,26	0,23	0,17
P. Antec. 5 d (mm)	0,08	0,09	0,08
P. Antec. 7 d (mm)	0,22	0,23	0,24
Qrio (m³/s)	0,31	0,28	0,33
Hidrog. sobe (Z)/desce (1)	-0,04	-0,07	0,01
ET real	0,30	0,29	0,36
A bacia (km²)	0,73	0,74	0,68
I (m/m)	0,33	0,34	0,35
TS - LVA (%)	0,14	0,14	0,08
TS - CA (%)	0,24	0,24	0,20
TS - PVAA (%)	0,28	0,29	0,46
TS - LVED (%)	0,09	0,09	0,03
TS - Ctba (%)	0,32	0,32	0,47
TS- PVAE (%)	0,90	0,91	0,84
US - AUB (%)	0,17	0,18	0,22
US - VAN (%)	0,38	0,38	0,43
US - Campo (%)	0,37	0,37	0,40
US - Água (%)	0,45	0,45	0,48
US - VAP (%)	0,18	0,18	0,14
US - VAN2 (%)	0,19	0,19	0,16
US - M/O (%)	-0,05	-0,05	-0,11
US - AI (%)	0,20	0,20	0,22
US - Vila (%)	0,12	0,12	0,09
US - CT (%)	0,57	0,58	0,59
US - AA (%)	-0,08	-0,08	-0,12
US - Granja (%)	0,39	0,40	0,56
US - CP (%)	0,13	0,13	0,06
US - SE (%)	0,46	0,46	0,53
US - AUM (%)	0,26	0,26	0,38
US - Lot. (%)	0,33	0,33	0,48
US - Lixão (%)	0,51	0,52	0,60
P. Antec. 1 d (mm) - Thiessen	0,43	0,41	0,34
P. Antec. 3 d (mm) - Thiessen	0,49	0,48	0,48
P. Antec. 5 d (mm) - Thiessen	0,45	0,46	0,49
P. Antec. 7 d (mm) - Thiessen	0,38	0,38	0,46
P. Antec. 1 d (mm) - 2549100	0,44	0,43	0,34
P. Antec. 3 d (mm) - 2549100	0,52	0,51	0,44
P. Antec. 5 d (mm) - 2549100	0,54	0,54	0,50
P. Antec. 7 d (mm) - 2549100	0,54	0,54	0,57
P. Antec. 1 d (mm) - 2549107	0,31	0,30	0,26
P. Antec. 3 d (mm) - 2549107	0,29	0,27	0,41
P. Antec. 5 d (mm) - 2549107	0,32	0,32	0,45
P. Antec. 7 d (mm) - 2549107	0,25	0,26	0,37
P. Antec. 1 d (mm) - 2549128	0,41	0,38	0,32
P. Antec. 3 d (mm) - 2549128	0,54	0,52	0,50
P. Antec. 5 d (mm) - 2549128	0,37	0,38	0,39
P. Antec. 7 d (mm) - 2549128	0,33	0,33	0,41
P. Antec. 1 d (mm) - 2549077	0,78	0,76	0,72
P. Antec. 3 d (mm) - 2549077	0,89	0,90	0,87
P. Antec. 5 d (mm) - 2549077	0,83	0,86	0,88
P. Antec. 7 d (mm) - 2549077	0,79	0,80	0,88
P. Antec. 1 d (mm) - 2549090	0,41	0,38	0,28
P. Antec. 3 d (mm) - 2549090	0,54	0,52	0,48
P. Antec. 5 d (mm) - 2549090	0,47	0,49	0,50
P. Antec. 7 d (mm) - 2549090	0,42	0,44	0,51
P. Antec. 1 d (mm) - 2549006	0,81	0,77	0,77
P. Antec. 3 d (mm) - 2549006	1,00	0,97	0,92
P. Antec. 5 d (mm) - 2549006	0,97	1,00	0,93
P. Antec. 7 d (mm) - 2549006	0,92	0,93	1,00

Legenda:

Tc (Tempo de concentração);
Tci (tempo desde chuva inicial);
Conc. (concentração);
P. Acum. (Precipitação Acumulada);
P. Antec. (Precipitação Antecedente);
d (dias);
ET (Evapotranspiração);
A (área);
I (declividade média);
TS (Tipo de Solo);
US (Uso do Solo);
LVA (Latossolo vermelho-amarelo);
CA (Cambissolo álico);
PVAA (Podzólico Vermelho-amarelo Álico);
LVED (Latossolo Vermelho-escuro Distrófico);
Ctba (Curitiba);
PVAE (Podzólico Vermelho-amarelo Eutrófico);
AUB (Área Urbana Baixa);
VAN (Vegetação Arbórea Natural);
VAP (Vegetação Arbórea Plantada);
VAN2 (Vegetação Arbustiva Natural);
M/O (Mineração/Outros);
AI (Área Industrial);
CT (Cultura Temporária);
AA (Área Alagada);
CP (Cultura Permanente);
SE (Solo Exposto);
AUM (Área Urbana Média);
Lot. (Loteamentos).

Fonte: A Autora (2021)

A matriz de correlação entre variáveis (em que as correlações mais fortes estão em cores mais claras) é apresentada na Figura 19. Ela permite identificar para o Fósforo Total, por exemplo, que as variáveis que mais se relacionam diretamente com ele são, nesta ordem, a concentração de Nitrogênio Total, Tipo do Solo – C, Tipo do Solo – PVAA, Uso da terra – L e Uso da terra – AUM. A melhor regressão entre tais variáveis, entretanto, é de $R^2 = 0,53$ e $RSME = 0,51$ mg/L, o que apresenta que apenas boas correlações entre variáveis não são o suficiente para prever uma variável, necessitando-se de diferentes tipos de informações.

Outra característica apresentada na matriz de correlação é que são altas as correlações entre, principalmente, as estações pluviométricas no entorno da bacia, o que pode significar que apenas uma ou então apenas a informação de Thiessen já seja o suficiente para um modelo de previsão, evitando variáveis que serão redundantes para o resultado.

As variáveis amostra, evento e condição do hidrograma foram as que apresentaram as menores correlações com as outras variáveis. O número do evento e a condição do hidrograma realmente não foram expressivos para a obtenção dos melhores resultados, sendo que a matriz de correlação pode ser um possível indicativo de quais dados de entrada eliminar para gerar as relações.

Sendo a matriz uma avaliação linear da correlação entre variáveis, suas conclusões são somente aplicadas para modelagem deste tipo. A matriz apresenta que a correlação linear entre suas variáveis não é suficiente para representar as condições poluentes e, portanto, a aplicação de modelos multicolineares e não-lineares é relevante.

TABELA 10 - QUANTIDADE DE VEZES QUE CADA COLUNA FOI UTILIZADA PARA GERAR AS 11 MELHORES SIMULAÇÕES PARA FÓSFORO TOTAL PELO INDICADOR R^2

PT – Variáveis		20	1000	4500	20 sc	1000 sc	4500 Sc	1000 Sce	1000 scet	1000 sceu	1000 scetu	1000 scuteth	10 Melhores Resultados
1	Ponto	6	6	7	7	2	3	1	6	2	3	3	5
2	Evento	9	8	8	7	3	6	5	7	9	9	8	9
3	Tc	6	5	7	7	4	3	5	4	4	4	6	3
4	Tempo desde chuva inicial	8	6	5	4	1	7	4	4	4	5	6	3
5	Concentração NT	7	11	10									10
6	Concentração PT												
7	Concentração COD	5	7	5									3
8	Pacum - dentro do tc	6	8	6	6	4	2	2	3	2	3	1	7

9	Pacum - desde chuva inicial	6	7	5	4	2	2	3	3	7	7	7	6
10	Pacum antecedente 1 dia	7	8	7	5	3	1	4	3	7	3	7	4
11	Pacum antecedente 3 dia	7	11	7	5	6	4	2	2	6	3	8	5
12	Pacum antecedente 5 dia	8	9	8	6	6	6	4	3	5	8	8	6
13	Pacum antecedente 7 dia	7	9	8	4	2	3	6	5	7	7	5	8
14	Qrio	8	11	11	9	4	8	8	10	8	11	11	10
15	Condição hidrograma	6	8	4	7	3	2	3	3	6	7	6	7
16	Evapotranspiração real	6	7	7	6	4	4	5	10	7	7	9	7
17	Área da bacia	8	8	9	3	2	2	3	1	5	5	3	5
18	Declividade média	5	6	7	3	1	2	1	3	4	2	1	6
19	Tipo de Solo - LVA	6	7	4	4	2	4	3		4			3
20	Tipo de Solo - CA	7	8	7	3	2	6	1		2			6
21	Tipo de Solo - PVAE	6	6	5	5	2	1	4		3			4
22	Tipo de Solo - LVED	8	10	9	4	2	2	4		5			10
23	Tipo de Solo - C	7	8	6	7	2	1	2		3			2
24	Tipo de Solo - PVAA	8	9	6	6	3	4	3		2			3
25	Uso da terra - AUB	5	11	9	6	3	3	1	1				7
26	Uso da terra - VAN	6	8	6	4	1	2	2	3				6
27	Uso da terra - C	8	7	5	6	0	2	3	3				7
28	Uso da terra - A	7	9	8	5	2	4	5	5				8
29	Uso da terra - VAP	5	8	4	6	3	3	2	3				1
30	Uso da terra - VAN	8	10	5	4	4	7	1	3				7
31	Uso da terra - M/O	7	9	8	4	2	3	2	4				4
32	Uso da terra - AI	7	6	7	4	2	2	4	4				5
33	Uso da terra - V	8	6	6	5	2	1	1	4				4
34	Uso da terra - CT	7	8	8	5	0	4	6	3				6
35	Uso da terra - AA	7	9	6	5	3	3	3	3				5
36	Uso da terra - G	8	7	7	6	4	2	5	3				6
37	Uso da terra - CP	7	9	6	4	4	3	2	3				5
38	Uso da terra - SE	8	6	4	6	4	4	2	2				4
39	Uso da terra - AUM	7	9	4	6	2	1	4	4				2
40	Uso da terra - Lo	6	8	7	5	2	3	4	6				9
41	Uso da terra - Li	8	10	5	5	2	1	5	5				6
42	P antecedente 1 dia- Thiessen	6	7	5	5	2	3	4	6	8	8		3
43	P antecedente 3 dia- Thiessen	8	9	8	5	2	3	6	3	5	6		7
44	P antecedente 5 dia- Thiessen	6	6	8	4	5	5	1	1	2	3		6
45	P antecedente 7 dia- Thiessen	8	8	7	3	0	1	1	3	0	2		6
46	P antecedente 1 dia- 2549100	7	9	7	2	2	1						8
47	P antecedente 3 dia- 2549100	9	9	7	5	4	6						7
48	P antecedente 5 dia- 2549100	9	5	4	7	6	5						3
49	P antecedente 7 dia- 2549100	6	8	7	4	6	5						5
50	P antecedente 1 dia- 2549107	8	9	6	6	3	5						6
51	P antecedente 3 dia- 2549107	6	9	6	6	4	6						6
52	P antecedente 5 dia- 2549107	8	7	7	6	1	3						8
53	P antecedente 7 dia- 2549107	8	8	6	6	1	3						7
54	P antecedente 1 dia- 2549128	5	8	5	4	2	3						6

55	P antecedente 3 dia- 2549128	8	2	3	3	3	2	5
56	P antecedente 5 dia- 2549128	8	8	7	3	3	3	6
57	P antecedente 7 dia- 2549128	9	7	7	6	3	4	5
58	P antecedente 1 dia- 2549077	8	7	5	6	3	4	5
59	P antecedente 3 dia- 2549077	8	7	6	5	4	2	4
60	P antecedente 5 dia- 2549077	7	11	6	4	2	3	7
61	P antecedente 7 dia- 2549077	6	7	5	4	6	7	7
62	P antecedente 1 dia- 2549090	7	6	4	3	2	3	5
63	P antecedente 3 dia- 2549090	7	3	4	5	2	5	5
64	P antecedente 5 dia- 2549090	8	8	7	5	6	6	7
65	P antecedente 7 dia- 2549090	7	8	4	5	2	2	3
66	P antecedente 1 dia- 2549006	5	7	6	3	0	2	7
67	P antecedente 3 dia- 2549006	8	8	5	6	2	2	2
68	P antecedente 5 dia- 2549006	7	8	7	6	2	2	8
69	P antecedente 7 dia- 2549006	5	8	6	7	1	2	5

Fonte: A Autora (2021)

TABELA 11 - QUANTIDADE DE VEZES QUE CADA COLUNA FOI UTILIZADA PARA GERAR AS 11 MELHORES SIMULAÇÕES PARA NITROGÊNIO TOTAL PELO INDICADOR R²

NT - Variáveis		20	1000	4500	20sc	1000sc	4500sc	10 Melhores Resultados
1	Ponto	8	4	5	4	4	2	5
2	Evento	8	8	7	7	3	1	8
3	tc	6	8	5	6	2	1	5
4	Tempo desde chuva inicial	7	4	5	7	2	2	7
5	Concentração NT							
6	Concentração PT	10	9	8				8
7	Concentração COD	10	6	8				9
8	Pacum - dentro do tc	10	8	6	8	5	0	7
9	Pacum - desde chuva inicial	8	4	7	5	4	0	7
10	Pacum antecedente 1 dia	11	6	6	8	3	1	7
11	Pacum antecedente 3 dia	6	7	4	5	2	3	5
12	Pacum antecedente 5 dia	7	4	6	4	4	1	7
13	Pacum antecedente 7 dia	8	4	7	5	5	0	7
14	Qrio	8	6	3	4	2	2	4
15	Condição hidrograma	8	6	5	5	1	2	6
16	Evapotranspiração real	8	6	6	6	0	3	7
17	Área da bacia	7	7	3	5	4	3	4
18	Declividade média	9	6	5	5	2	4	6
19	Tipo de Solo - LVA	6	4	5	3	7	2	5
20	Tipo de Solo - CA	7	9	4	6	3	1	5
21	Tipo de Solo - PVAE	8	5	6	4	0	1	7
22	Tipo de Solo - LVED	7	8	6	5	1	2	7
23	Tipo de Solo - C	7	5	5	6	1	0	6
24	Tipo de Solo - PVAA	8	5	3	7	4	1	3
25	Uso da terra - AUB	10	5	4	5	2	1	5
26	Uso da terra - VAN	8	4	6	5	2	2	5

27	Uso da terra - C	10	5	7	8	4	5	8
28	Uso da terra - A	7	5	7	7	1	3	8
29	Uso da terra - VAP	8	5	7	6	2	2	8
30	Uso da terra - VAN	9	5	6	5	4	1	5
31	Uso da terra - M/O	8	7	9	6	4	2	10
32	Uso da terra - AI	8	4	6	6	4	1	8
33	Uso da terra - V	8	8	7	6	2	1	6
34	Uso da terra - CT	6	6	6	5	4	0	7
35	Uso da terra - AA	7	6	8	6	4	2	8
36	Uso da terra - G	6	6	4	5	0	1	4
37	Uso da terra - CP	8	5	6	7	6	3	7
38	Uso da terra - SE	7	5	5	6	1	4	7
39	Uso da terra - AUM	10	5	6	7	1	0	7
40	Uso da terra - Lo	8	6	5	8	1	2	6
41	Uso da terra - Li	8	2	5	6	1	1	4
42	P antecedente 1 dia- Thiessen	8	5	7	7	4	0	7
43	P antecedente 3 dia- Thiessen	8	5	6	6	5	1	5
44	P antecedente 5 dia- Thiessen	9	5	6	4	3	3	6
45	P antecedente 7 dia- Thiessen	8	6	6	8	2	1	7
46	P antecedente 1 dia- 2549100	6	6	6	5	4	1	6
47	P antecedente 3 dia- 2549100	9	6	6	6	3	4	7
48	P antecedente 5 dia- 2549100	9	7	7	7	4	1	8
49	P antecedente 7 dia- 2549100	7	5	6	6	1	2	6
50	P antecedente 1 dia- 2549107	10	6	4	6	5	1	6
51	P antecedente 3 dia- 2549107	9	6	3	5	3	2	5
52	P antecedente 5 dia- 2549107	8	4	5	9	2	2	5
53	P antecedente 7 dia- 2549107	9	8	9	5	2	3	9
54	P antecedente 1 dia- 2549128	8	7	5	6	6	6	4
55	P antecedente 3 dia- 2549128	8	6	4	6	4	5	5
56	P antecedente 5 dia- 2549128	7	5	5	5	4	2	5
57	P antecedente 7 dia- 2549128	7	7	7	5	5	4	7
58	P antecedente 1 dia- 2549077	7	6	4	5	5	0	5
59	P antecedente 3 dia- 2549077	9	5	7	7	4	3	8
60	P antecedente 5 dia- 2549077	7	6	5	6	5	6	5
61	P antecedente 7 dia- 2549077	8	6	6	5	3	4	7
62	P antecedente 1 dia- 2549090	7	6	7	7	4	3	7
63	P antecedente 3 dia- 2549090	8	6	5	5	4	3	5
64	P antecedente 5 dia- 2549090	8	5	5	7	2	1	5
65	P antecedente 7 dia- 2549090	8	6	4	7	5	3	5
66	P antecedente 1 dia- 2549006	8	6	5	3	5	3	6
67	P antecedente 3 dia- 2549006	6	6	4	6	3	5	6
68	P antecedente 5 dia- 2549006	9	7	5	7	4	3	7
69	P antecedente 7 dia- 2549006	11	7	4	8	3	3	5

Fonte: A Autora (2021)

TABELA 12 - QUANTIDADE DE VEZES QUE CADA COLUNA FOI UTILIZADA PARA GERAR AS 11 MELHORES SIMULAÇÕES PARA CARBONO ORGÂNICO DISSOLVIDO PELO INDICADOR R²

COD - Variáveis		20	1000	4500	20sc	1000sc	4500sc	10 Melhores Resultados
1	Ponto	6	2	1	6	4	2	0
2	Evento	7	2	1	5	6	3	2
3	tc	4	3	5	4	5	3	2
4	Tempo desde chuva inicial	7	3	2	5	3	3	2
5	Concentração NT	7	4	3				1
6	Concentração PT	3	3	1				1
7	Concentração COD							
8	Pacum - dentro do tc	7	4	6	7	5	2	3
9	Pacum - desde chuva inicial	6	4	3	4	2	2	3
10	Pacum antecedente 1 dia	5	5	4	5	4	1	2
11	Pacum antecedente 3 dia	4	5	4	6	4	3	1
12	Pacum antecedente 5 dia	5	2	4	5	0	2	1
13	Pacum antecedente 7 dia	6	5	1	4	4	4	2
14	Qrio	7	4	3	4	2	1	1
15	Condição hidrograma	6	4	2	5	4	4	1
16	Evapotranspiração real	6	3	7	6	3	3	4
17	Área da bacia	8	3	4	8	5	3	2
18	Declividade média	5	4	3	5	5	3	2
19	Tipo de Solo - LVA	6	3	4	6	2	2	3
20	Tipo de Solo - CA	7	6	3	7	5	4	3
21	Tipo de Solo - PVAE	6	4	3	6	3	4	4
22	Tipo de Solo - LVED	7	3	2	7	0	3	2
23	Tipo de Solo - C	5	4	2	6	4	1	1
24	Tipo de Solo - PVAA	5	4	2	6	3	2	1
25	Uso da terra - AUB	6	4	5	4	5	2	3
26	Uso da terra - VAN	7	5	3	7	4	5	3
27	Uso da terra - C	7	4	3	6	4	4	2
28	Uso da terra - A	6	5	5	6	5	6	7
29	Uso da terra - VAP	3	5	6	6	5	4	5
30	Uso da terra - VAN	7	5	7	6	4	5	5
31	Uso da terra - M/O	4	4	5	6	3	3	3
32	Uso da terra - AI	7	4	5	5	5	5	4
33	Uso da terra - V	8	4	3	5	4	1	1
34	Uso da terra - CT	5	4	8	8	5	5	6
35	Uso da terra - AA	6	4	5	6	4	3	3
36	Uso da terra - G	4	3	4	6	2	4	2
37	Uso da terra - CP	3	2	3	6	4	2	2
38	Uso da terra - SE	7	3	3	5	3	1	0
39	Uso da terra - AUM	6	5	3	7	3	3	4
40	Uso da terra - Lo	6	3	4	9	6	3	1
41	Uso da terra - Li	5	1	4	6	6	0	2
42	P antecedente 1 dia- Thiessen	6	4	3	5	3	3	1
43	P antecedente 3 dia- Thiessen	5	2	3	3	6	3	1

44	P antecedente 5 dia- Thiessen	6	4	3	8	5	4	4
45	P antecedente 7 dia- Thiessen	5	5	4	6	5	3	4
46	P antecedente 1 dia- 2549100	4	4	2	4	3	1	1
47	P antecedente 3 dia- 2549100	4	3	3	7	5	1	1
48	P antecedente 5 dia- 2549100	4	4	1	7	5	2	1
49	P antecedente 7 dia- 2549100	5	3	4	5	5	4	3
50	P antecedente 1 dia- 2549107	6	4	1	4	1	2	1
51	P antecedente 3 dia- 2549107	6	4	4	6	4	4	3
52	P antecedente 5 dia- 2549107	7	3	6	8	3	1	3
53	P antecedente 7 dia- 2549107	8	3	4	4	4	3	2
54	P antecedente 1 dia- 2549128	6	4	5	6	4	6	8
55	P antecedente 3 dia- 2549128	8	6	4	7	5	2	2
56	P antecedente 5 dia- 2549128	8	3	4	8	5	3	3
57	P antecedente 7 dia- 2549128	6	4	4	6	4	4	3
58	P antecedente 1 dia- 2549077	7	2	2	4	3	3	2
59	P antecedente 3 dia- 2549077	7	5	0	6	3	2	2
60	P antecedente 5 dia- 2549077	6	2	3	5	2	5	2
61	P antecedente 7 dia- 2549077	5	3	4	6	5	3	2
62	P antecedente 1 dia- 2549090	7	4	2	6	4	3	1
63	P antecedente 3 dia- 2549090	5	3	4	4	1	1	1
64	P antecedente 5 dia- 2549090	2	5	6	6	5	2	3
65	P antecedente 7 dia- 2549090	5	4	3	9	5	5	2
66	P antecedente 1 dia- 2549006	6	4	4	6	4	2	1
67	P antecedente 3 dia- 2549006	5	4	6	6	2	4	3
68	P antecedente 5 dia- 2549006	5	3	5	6	5	3	2
69	P antecedente 7 dia- 2549006	5	3	3	8	5	2	2

Fonte: A Autora (2021)

A partir das Tabela 10, Tabela 11 e Tabela 12 outros testes foram realizados, baseado no número de vezes que cada coluna contribui para os melhores 11 resultados de cada teste e para os melhores 10 resultados gerais e, portanto, que são provavelmente importantes para a predição dos valores de concentração. Para o Fósforo Total, foram analisadas quatro situações diferentes com duas diferentes abordagens: a primeira consistia em uma avaliação com todos os parâmetros e a segunda sem dados de concentração (por serem de difícil acesso: existem porém são de propriedade particular/ não são liberados ao público; ou são escassos) e identificação do evento (para tentar generalizar o modelo). Posteriormente, essas situações foram simuladas a fim de observar se bons resultados podem ser obtidos ao se fazer tal tipo de análise e estão indicados na Tabela 13. O melhor resultado deu-se por um R^2 de 0,86 e $RSME = 0,28$ mg/L que utilizou as variáveis mais usadas considerando todos os testes.

TABELA 13 - SIMULAÇÕES REALIZADAS A PARTIR DAS VARIÁVEIS MAIS UTILIZADAS PARA GERAR OS MELHORES 11 RESULTADOS DE FÓSFORO TOTAL PELO INDICADOR R²

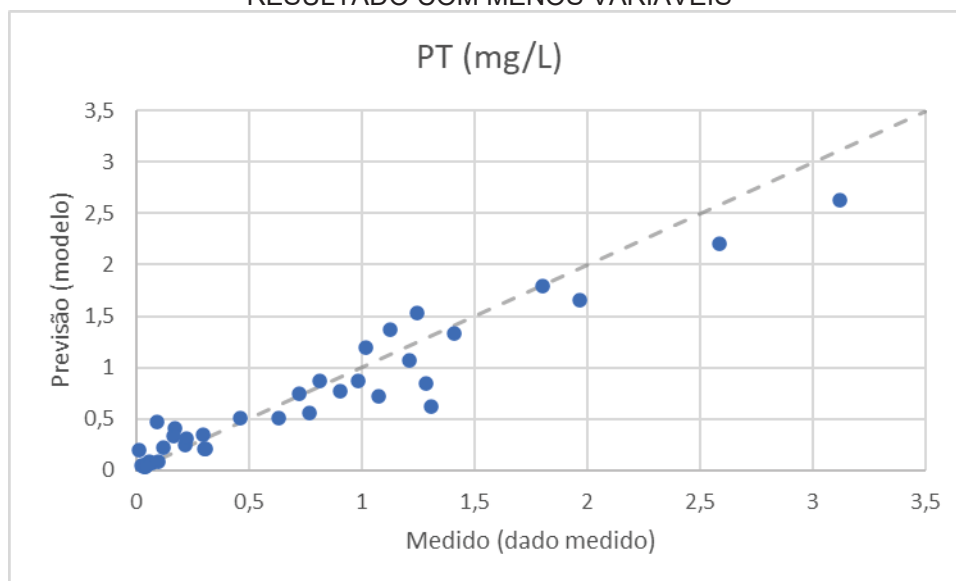
Teste	Variáveis usadas	R²	RSME	Regressor
Variáveis mais usadas nas 11 melhores respostas de cada teste	Evento, Concentração NT, Vazão no rio, Precipitação acumulada 5 dias, Evapotranspiração real, Precipitação antecedente 3 dias - 2549100, Precipitação antecedente 5 dias - 2549090	0,86	0,28	Bagging
Variáveis mais usadas nas 11 melhores respostas de cada teste (sem concentração e identificação do evento)	Vazão no rio, Precipitação acumulada 5 dias, Evapotranspiração real, Precipitação antecedente 3 dias - 2549100, Precipitação antecedente 5 dias - 2549090	0,86	0,28	Bagging
Variáveis mais usadas nas 11 melhores respostas de cada teste nas simulações que consideraram concentração	Evento, Concentração NT, Vazão no Rio, Precipitação acumulada 3 dias, Precipitação acumulada 5 dias, Área da Bacia, Tipo de Solo – LVED, Uso da terra – AUB, Precipitação antecedente 3 dias – Thiessen, Precipitação antecedente 3 dias 2549100	0,82	0,31	XGB
Variáveis mais usadas nas 11 melhores respostas de cada teste nas simulações que consideraram concentração(sem concentração e identificação do evento)	Vazão no Rio, Precipitação acumulada 3 dias, Precipitação acumulada 5 dias, Área da Bacia, Tipo de Solo – LVED, Uso da terra – AUB, Precipitação antecedente 3 dias – Thiessen, Precipitação antecedente 3 dias 2549100	0,78	0,35	Random Forest
Variáveis mais usadas nas 11 melhores respostas de cada teste nas simulações que	Evento, Evapotranspiração real, Vazão no rio, Precipitação Antecedente 5 dias - 2549100	0,83 0,81	0,31 0,32	KNeighbors

não consideraram concentração				
Variáveis mais usadas nas 11 melhores respostas de cada teste nas simulações que não consideraram concentração (sem concentração e identificação do evento)	Evapotranspiração real, Vazão no rio, Precipitação Antecedente 5 dias - 2549100	0,81	0,32	Extra Trees
Variáveis mais usadas nos 10 melhores resultados gerais	Evento, Concentração NT, Vazão no rio, Tipo de Solo – LVED, Uso da terra - Lo	0,85	0,29	Bagging
Variáveis mais usadas nos 10 melhores resultados gerais (sem concentração e identificação do evento)	Vazão no rio, Tipo de Solo – LVED, Uso da terra – Lo	0,52	0,51	NuSVR

Fonte: A Autora (2021)

Outro resultado relevante foi de que, ao se analisarem todas as melhores 11 combinações de cada teste e, como critério, observar-se aquelas com poucas variáveis e, portanto, com menor quantidade de dados a serem obtidos, tornando a predição mais simples, houve uma combinação (simulação 1000scuteth) de 4 variáveis (vazão no rio, evapotranspiração real, tempo desde a chuva inicial e precipitação antecedente acumulada de 3 dias) com $R^2=0,911$ e $RSME=0,375$ mg/L, um resultado excelente, com predições conforme a Figura 20, que indica um bom ajuste com tendência de subestimação das concentrações para os maiores valores.

FIGURA 20 - CONCENTRAÇÕES REAL E MODELADA DE FÓSFORO TOTAL PELO MELHOR RESULTADO COM MENOS VARIÁVEIS



Fonte: A Autora (2021)

Com isso, percebe-se que é possível fazer a simulação da concentração a partir de inteligência artificial, chegando-se para as condições de separação de amostras de treino e verificação de 80%-20% a um ajuste de R^2 de 0,935 e $RSME = 0,309$ mg/L, com 14 variáveis para o fósforo total. Para o nitrogênio total, chegou-se a um modelo com R^2 de 0,967 e $RSME = 2,160$ mg/L. Para carbono orgânico dissolvido os modelos não obtiveram um coeficiente de determinação tão altos quanto para os outros parâmetros, indicando dificuldade maior para ajustar a linha de regressão aos dados, uma vez que o maior R^2 foi de 0,689 e $RSME = 2,344$ mg/L.

A utilização de dados de fácil acesso, sem utilizar concentrações de outros poluentes durante os eventos de precipitação, e com menor dificuldade de obtenção, verificou modelos de predição ligeiramente piores que aqueles que as utilizam.

As análises de modelagem indicaram uma importante informação quanto ao regressor que melhor se ajusta a cada tipo de parâmetro. Para o fósforo total, o regressor Bagging se mostrou o mais indicado. Para o nitrogênio total, o regressor Extra Tree. Para o carbono orgânico dissolvido, os regressores Extra Tree e Decision Tree.

5 CONCLUSÕES

As poluições difusa e pontual que chegam no corpo hídrico são variáveis no tempo e no espaço. Esta variabilidade depende de condições de lançamento de efluentes pontuais, dos eventos de precipitação, dos fatores territoriais da bacia hidrográfica como tipo do solo, uso da terra e condições do terreno. Entender a variabilidade é uma ferramenta necessária para o desenvolvimento dos instrumentos de gestão hídrica e, conseqüentemente, a alocação de cargas poluentes. Ainda, é uma forma de otimizar as ações e alocar os recursos financeiros e humanos para ações que efetivamente causem efeito.

Modelos que utilizem dados disponíveis e de fácil acesso tornam o processo de monitoramento e acompanhamento de índices mais simples e rápidos. A aplicação da inteligência artificial à estimativa e predição de concentrações de poluentes em corpos hídricos pode vir a ser uma ótima solução para aprender o comportamento do rio com o tempo, ao serem adicionados novos dados ao sistema.

Além disso, o modelo pode ser utilizado para gerar uma ferramenta de simulação de cargas poluidoras discretizadas temporalmente. Uma vez obtidos dados de precipitação de certo ano, do comportamento hidrológico e as características do território, poderá haver uma previsão de concentrações ao longo do ano que, multiplicadas pela vazão medida, acarretará um modelo de cargas poluentes no tempo.

A discretização temporal das cargas estimadas pode ser uma ferramenta de compreensão da importância da variação, principalmente das cargas provenientes por vias difusas, de acordo com o regime de precipitação na bacia e seus impactos nos corpos hídricos. Os planos de ação poderão ser elaborados a partir destes modelos.

Esta pesquisa mostrou resultados promissores para a criação de um modelo de estimativa de concentração, principalmente em eventos de precipitação que indicarão a contribuição da poluição difusa. Para a consolidação de um modelo, a quantidade de amostras coletadas e a variabilidade espacial das coletas é um passo importante para a elaboração de um modelo genérico. Sugere-se, para futuros estudos, que não sejam considerados como variáveis elegíveis ao modelo aqueles referentes ao ponto de estudo ou evento, uma vez que estas variáveis podem induzir o modelo a respostas localizadas e regionais.

Podem ser objeto de estudo, ainda, a proporção de separação das amostras de treinamento e verificação que gerem os melhores modelos, como verificado nas respostas dos testes t1 e t2, e a forma de separação, conforme verificado nas respostas dos testes t3 e t4.

Uma melhor assertividade na geração de modelos ou estimativas para cada local é dependente da quantidade e qualidade dos dados de entrada. Para este estudo, houve uma limitação na representatividade da precipitação na bacia, uma vez que os dados de precipitação com escala temporal menores que horárias não representavam a bacia como um todo. Para BA1, a estação localizava-se fora da bacia, enquanto para BA2 a estação estava próxima ao exutório, podendo não representar a bacia. Em uma tentativa de minimizar os impactos desta limitação, foram utilizados os dados de estações que formaram os polígonos do Thiessen. Para analisar as precipitações antecedentes, as informações eram satisfatórias, mas não para representar cada evento, pois as leituras eram feitas somente duas vezes no dia.

Ainda, a análise dos modelos por outros indicadores em conjunto e não somente o R^2 pode gerar resultados ainda mais satisfatórios em questão de previsão de concentrações. É preciso avaliar as vantagens e desvantagens de cada indicador, para ser cada vez mais assertivo. A partir dos modelos apresentados neste trabalho é possível fazer uma análise de sensibilidade das variáveis de entrada, o que pode ser útil para reduzir o tamanho do modelo.

As elaborações de modelos de estimativa de cargas poderão utilizar-se desta pesquisa para a utilização dos regressores. A aplicação da metodologia em três diferentes parâmetros de qualidade da água permitiu visualizar que cada parâmetro é mais bem entendido por um regressor diferente. De forma geral, as árvores de decisão, sejam aplicadas diretamente ou como conjunto, foram explicativas na interrelação da concentração dos parâmetros com as variáveis dos eventos pluviométricos, características da bacia e condições hidrológicas do rio.

É possível desenvolver modelos de previsão de concentrações a partir do modelo de estimativa. Para isso, outros modelos de previsão dos dados de entrada devem trabalhar em conjunto, em especial aqueles que simulam os dados de precipitação, vazão e evapotranspiração.

Dá-se destaque aos dados de Kozak (2016,2020) que permitem uma perspectiva sobre a poluição difusa relacionando-as com seus efeitos durante os

eventos de precipitação e auxiliando inúmeras outras pesquisas e lançando novos desafios e questionamentos quanto a qualidade das águas.

Esta pesquisa obteve resultados de modelos de estimativa de carga poluente com indicadores de resultados (R^2 - coeficiente de determinação) em valores superiores a 0,90 para nitrogênio total e fósforo total, indicando forte adequação do modelo aos dados reais.

Por fim, uma das conclusões mais relevantes deste trabalho é que os poluentes em um corpo hídrico se comportam de maneira não-linear. Especialmente aqueles na bacia do Rio Barigui são melhor representados por modelos de árvores de decisão e precisam de relativamente poucas variáveis para serem estimados. Ainda, a estimativa pode utilizar diferentes modelos de acordo com as informações disponíveis, sem afetar significativamente o resultado.

6 REFERÊNCIAS BIBLIOGRÁFICAS

ACOSTA, E. A. et al. Estimativa de cargas de fósforo provenientes da erosão que atingem as águas superficiais da Bacia do Alto Iguaçu/PR. In: SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, 20, 2013, Bento Gonçalves. **Anais...** São Paulo: ABRH, 2013.

ACOSTA, M. J. P. **Simulação do escoamento superficial e erosão do solo utilizando o Método do Reticulado de Boltzmann.** Tese (Doutorado em Engenharia de Recursos Hídricos e Ambiental) – Setor de Tecnologia, Universidade Federal do Paraná, Curitiba, 2019.

AGÊNCIA NACIONAL DE ÁGUAS (ANA). **Implementação do Enquadramento em Bacias Hidrográficas no Brasil:** Cadernos de Recursos Hídricos 6. Brasília: ANA, 2009.

AGÊNCIA NACIONAL DE ÁGUAS (ANA). **Plano Integrado de Recursos Hídricos da Unidade de Gestão de Recursos Hídricos Paranapanema.** Comitê da Bacia Hidrográfica do Rio Paranapanema, 2016.

AGÊNCIA NACIONAL DE ÁGUAS (ANA). **Conjuntura de Recursos Hídricos no Brasil.** 177p. 2017.

AGÊNCIA NACIONAL DE ÁGUAS (ANA). **Encarte: Cobrança pelo uso dos recursos hídricos.** Brasília, 80p. 2019a.

AGÊNCIA NACIONAL DE ÁGUAS (ANA). **Encarte: Outorga dos direitos de uso dos recursos hídricos.** Brasília, 76p. 2019b.

AGÊNCIA NACIONAL DE ÁGUAS (ANA). **Encarte: Enquadramento dos corpos d'água em classes.** Brasília, 57p. 2020a.

AGÊNCIA NACIONAL DE ÁGUAS (ANA). **Encarte: Sistema de informações sobre recursos hídricos.** Brasília, 33p. 2020b.

AGÊNCIA NACIONAL DE ÁGUAS (ANA). **Estimativas de Evapotranspiração Real por Sensoriamento Remoto no Brasil.** Brasília, 2020c.

ALMEIDA, L.; SERRA, J. C. V. **Modelos hidrológicos, tipos e aplicações mais utilizadas.** FAE, v. 20, n. 1, p 129 – 137. 2017.

BARROS, A. M. L. **Modelagem da poluição pontual e difusa: aplicação do modelo MONERIS à Bacia Hidrográfica do Rio Ipojuca, Pernambuco.** Dissertação (Mestrado em Engenharia Civil) – Programa de Pós-Graduação em Engenharia Civil, Universidade Federal de Pernambuco, Recife, 2008.

BATISTEL, L. **Estudo do coeficiente de recessão em três bacias hidrográficas embutidas no município de Rio Negrinho – SC.** Monografia (Bacharelado em Engenharia Ambiental) – Universidade Federal da Fronteira Sul, Chapecó, 2015.

BI, J. et al. Large-scale water quality prediction with integrated deep neural network. **Information Sciences**, v. 571, p. 191-205, 2021.

BITENCOURT, C. C. A. **Integração de análise estatística multivariada e geoprocessamento: aplicação para estudos de enquadramento de corpos d'água**. Dissertação (Mestrado em Engenharia de Recursos Hídricos e Ambiental) – Setor de Tecnologia, Universidade Federal do Paraná, Curitiba, 2018.

BRAGA, S. M. **Uma nova abordagem para integração entre quantidade e qualidade da água para a avaliação da poluição difusa**. Tese (Doutorado em Engenharia de Recursos Hídricos e Ambiental) – Setor de Tecnologia, Universidade Federal do Paraná, Curitiba, 2013.

BRASIL. **Constituição (1988)**. Constituição da República Federativa do Brasil. Brasília, DF, 1988.

BRASIL. **Casa Civil**. Lei Federal n. 6.938, de 31 ago. de 1981. Dispõe sobre a Política Nacional do Meio Ambiente, seus fins e mecanismos de formulação e aplicação, e dá outras providências. Brasília, DF, ago, 1981.

BRASIL. **Casa Civil**. Lei nº 9.433, de 8 de janeiro de 1997. Institui a Política Nacional de Recursos Hídricos, cria o Sistema Nacional de Gerenciamento de Recursos Hídricos, regulamenta o inciso XIX do art. 21 da Constituição Federal, e altera o art. 1º da Lei no 8.001, de 13 de março de 1990, que modificou a Lei no 7.990, de 28 de dezembro de 1989.

BRASIL. **Resolução CONAMA n. 274**, de 29 de novembro de 2000. Define os critérios de balneabilidade em águas brasileiras. Publicado no DOU n. 018, de 25 de janeiro de 2001.

BRASIL. **Resolução CONAMA n. 357**, de 17 de março de 2005. Classificação de águas, doces, salobras e salinas do Território Nacional. Publicado no DOU n. 053, de 18 de março de 2005.

BRASIL. Ministério do meio ambiente (MMA). **Plano Nacional de Recursos Hídricos**. Panorama e estado dos recursos hídricos do Brasil: Volume 1. Brasília: MMA, 2006.

BRASIL. **Resolução CONAMA n. 430**, de 13 de maio de 2011a. Dispõe sobre as condições e padrões de lançamento de efluentes, complementa e altera a Resolução n. 357, de 17 de março de 2005, do Conselho Nacional do Meio Ambiente-CONAMA. Publicado no DOU n. 092, de 16 de maio de 2011.

BRASIL. Portaria n. 2.914, de 12 de dez. de 2011. **Dispões sobre os procedimentos de controle e de vigilância da qualidade da água para o consumo humano e seu padrão de potabilidade**. Brasília, DF, dez 2011b.

BRITES, A. P. Z. **Enquadramento dos corpos de água através de metas progressivas: probabilidade de ocorrência e custos de despoluição hídrica**.

Tese (Doutorado em Engenharia) – Escola Politécnica, Universidade de São Paulo, São Paulo, 2010.

CALIFORNIA DEPARTMENT OF TRANSPORTATION (CALTRANS). **First Flush Phenomenon Characterization**. Sacramento, 2005.

CENTRO DE ESTUDOS E DEBATES ESTRATÉGICOS. **Instrumentos de Gestão das Águas**. Brasília: Centro de Documentação e Informação, 2015.

CHANG, N. B. et al. Evaluation of Non-point Source Loads in the Reservoir Watershed using the GIS/GPS/RS Information Technologies and Numerical Models. **Water International**, v. 26, n. 2, p. 239-251, 2001.

CHEN, L; CHEN, S; LI, S; SHEN, Z. **Temporal and spatial scaling effects of parameter sensitivity in relation to non-point source pollution simulation**. *Journal of Hydrology* 571. 36-49, 2019.

CORRÊA, G. P.; CUNHA, C. L. N.; BITENCOURT, C. C. A. Determinação do coeficiente de exportação de fósforo em áreas agrícolas. In: SIMPÓSIO DE HIDRÁULICA E RECURSOS HÍDRICOS DOS PAÍSES DE LÍNGUA PORTUGUESA, 12, 2015, Brasília. **Anais...** Brasília: ABRH, 2015.

COSTA, D. A. et al. Dos instrumentos de gestão de recursos hídricos – o Enquadramento – como ferramenta para reabilitação de rios. **Saúde Debate**, v. 43, n. 3, p. 35-50. Rio de Janeiro, 2019.

CUNHA, D. G. F.; CALIJURI, M. C. Análise probabilística de ocorrência de incompatibilidade da qualidade da água com o enquadramento legal de sistemas aquáticos – estudo de caso do rio Pariquera-Açu (SP). **Engenharia Sanitária Ambiental**, v. 15, n. 4, p. 337-346. São Paulo, 2010.

CUNHA, D. G. F.; CALIJURI, M. C.; MENDIONDO, E. M. Integração entre curvas de permanência de quantidade e qualidade da água como uma ferramenta para a gestão eficiente dos recursos hídricos. **Engenharia Sanitária Ambiental**, v. 17, n. 4, p. 369-376. São Paulo, 2012.

DANIEL, E. B. et al. Watershed Modeling and its Applications: A State-of-the-Art Review. **The Open Hydrology Journal**, v. 5, p. 26-50, 2011.

DAVIE, Tim. **Fundamentals of Hydrology**. 2. ed. New York: Taylor & Francis Group, 2008.

DAWOOD, D. et al. Toward urban sustainability and clean potable water: Prediction of water quality via artificial neural networks. **Journal of Cleaner Production**, v. 291, 2021.

DRUMMOND, S. B. M. **Análise temporal de poluentes difusos no rio passaúna por meio de amostrador automático e dados de uso e cobertura do solo**. Dissertação (Mestrado em Engenharia de Recursos Hídricos e Ambiental) – Setor de Tecnologia, Universidade Federal do Paraná, Curitiba, 2020.

DULAC, V. F. **Análise dos instrumentos de gestão dos recursos hídricos com ênfase nas ações do comitê de Bacia do Rio Santa Maria, RS.** Dissertação (Mestrado em Engenharia Civil e Ambiental) – Centro de Tecnologia, Universidade Federal de Santa Maria, Santa Maria, 2013.

EIGER, S. et al. Modelagem matemática de cargas difusas na Bacia do Guarapiranga, São Paulo. In: SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, 13, 1999, Belo Horizonte. **Anais...** Brasília: ABRH, 1999.

FAGUNDES, H. O. **Modelagem hidrossedimentológica de grandes bacias com apoio de dados in situ e sensoriamento remoto.** Dissertação (Mestrado em Recursos Hídricos e Saneamento Ambiental) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.

FAYAL, M. A. A. **Previsão de vazão por redes neurais artificiais e transformada wavelet.** Dissertação. PUC-Rio. Rio de Janeiro, f. 158. 2008.

FERREIRA, D. B. **Nonpoint pollution and Best Management Practices assessment using Geographic Information Systems data – application in the Barigui Basin, State of Parana.** Dissertação (Mestrado em Engenharia de Recursos Hídricos e Ambiental) – Setor de Tecnologia, Universidade Federal do Paraná, Curitiba, 2016.

FERREIRA, D. B.; MÜHLENHOFF, A. P.; FERNANDES, C. V. S. Modelos de poluição difusa: desafios, estratégias e impacto para a gestão de recursos hídricos. **REGA**, v. 15, n. 10, 2018.

FLECK, L.; TAVARES, M. H. F.; EYNG, E. Principais modelos matemáticos de qualidade da água e suas aplicações: Uma revisão. **Revista Eletrônica Científica Inovação e Tecnologia**, v. 1, n. 7, 2013.

FUCHS, S.; TOSHOVSKI, S. Modelling of micro pollutant emissions to water and evaluation of reduction measures. In: NOVATECH, 9, 2016, Lyon. **Anais...** London: IWA, 2016.

FUCHS, S. et al. Modeling of Regionalized Emissions (MoRE) into Water Bodies: An Open-Source River Basin Management System. **Water**, v. 9, n. 239, 2017.

FUJIWARA, T; YOSHIDA, A; OHTOSHI, K. **Temporal Variations od Pollutant Loads During Storm Events in a Small River Basin.** Journal of Water and Environment Technology, vol.3, no 2, 183-190, 2005.

GOMES, F. C.; SIMÕES, S. J. C. Simulação de Modelagem Qualitativa para Avaliação Preliminar da Qualidade da Água na Bacia do Ribeirão das Perdizes em Campos do Jordão/SP, como Subsídio ao Enquadramento. **Revista Brasileira de Recursos Hídricos**, v. 19, n. 3, p. 309-326, 2014.

GRUDZIEN, J. P. **Utilização de amostrador automático experimental para a identificação do aporte de poluentes no Rio Passaúna.** Dissertação (Mestrado

em Engenharia de Recursos Hídricos e Ambiental) – Setor de Tecnologia, Universidade Federal do Paraná, Curitiba, 2019.

HA, H; STENSTROM, M. **Identification of land use with water quality data in stormwater using a neural network**. Water Research 37. 4222-4230, 2003.

HAYKIN, S. **Redes Neurais: Princípios e Práticas**. 2ª Ed. Porto Alegre: Bookman, 2001.

INSTITUTO ÁGUA E TERRA DO PARANÁ (IAT). **Relatório de conjuntura dos recursos hídricos do Estado do Paraná**. Curitiba, 2020.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Manual técnico de uso da terra**. 3. ed. Rio de Janeiro, 2013.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Censo Demográfico de 2010**, Rio de Janeiro, 2010. Disponível em: <<http://www.ibge.gov.br>>. Acesso em: 14 abr. 2018.

KNAPIK, H. G. et al. Estratégia para consolidação de metas progressivas de qualidade de água: estudo de caso do enquadramento do Rio Iguaçu. In: SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, 18, 2009, Campo Grande. **Anais...** Brasília: ABRH, 2009.

KNAPIK, H. G.; FERNANDES, C. V. S.; AZEVEDO, J. C. R. Aplicabilidade de técnicas de espectroscopia e da concentração de carbono orgânico dissolvido na caracterização da matéria orgânica em rios urbanos. **Revista Brasileira de Recursos Hídricos**, v. 19, n. 4, p. 214-225, 2014.

KOZAK, C. **Water quality assessment and its effects on diffuse pollution considering a new water quality and quantity approach**. 2016. 265 f. Dissertação (Mestrado em Engenharia de Recursos Hídricos e Ambiental) – Setor de Tecnologia, Universidade Federal do Paraná, Curitiba, 2016.

KOZAK, C. Water quality dynamic during rainfall episodes: integrated approach to assess diffuse pollution using automatic sampling. **Environ Monit Assess**, v. 191, n. 402, 2019.

KOZAK, C. **Non point source pollution assessment based on event driven influence**. Tese (Doutorado em Engenharia de Recursos Hídricos e Ambiental) – Setor de Tecnologia, Universidade Federal do Paraná, Curitiba, 2020.

KRTOLICA, I. et al. Water quality and macrophytes in the Danube River: Artificial neural network modelling. **Ecological Indicators**, v. 121, 2021.

LI, H. Z.; ZHANG, M. X. A review on the calculation of non-point source pollution loads. In: INTERNATIONAL CONFERENCE ON WATER RESOURCE AND ENVIRONMENT, 5, 2019, Macao. **Anais...** WRE, 2019.

LIMA, R. N. S. et al. Estudo da poluição pontual e difusa na bacia de contribuição do reservatório da usina hidrelétrica de Funil utilizando modelagem espacialmente distribuída em Sistema de Informação Geográfica. **Engenharia Sanitária Ambiental**, v. 21, n. 1, p. 139-150, 2016.

LIU, M.; LU, J. Support vector machine—an alternative to artificial neuron network for water quality forecasting in an agricultural nonpoint source polluted river? **Environmental Science and Pollution Research**, v. 21, n. 18, p. 11036-11053, 2014.

LOU, R. F. **Modelagem Hidrológica Chuva-Vazão e Hidrodinâmica aplicada na bacia experimental do Rio Piabanha/RJ**. 2010

LUZ, E. P. et al. Ferramenta computacional “ISOCHRONES”: análise na Bacia Hidrográfica do Arroio Pelotas visando modelagem de cheias. In: CONGRESSO DE INICIAÇÃO CIENTÍFICA DA UNIVERSIDADE FEDERAL DE PELOTAS, 13, 2014, Pelotas. **Anais...** Pelotas: UFPel, 2014.

MACHADO, E. S.; KNAPIK, H. G.; BITTENCOURT, C. C. A. **Considerações sobre o processo de enquadramento de corpos de água**. Engenharia Sanitária e Ambiental. Volume 24, n. 2, p. 261-269, mar. 2019.

MAGALHÃES, I. A. L. et al. Análise de métodos de interpolação para espacialização da precipitação fluvial na região Norte do estado do Espírito Santo, Brasil. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 16, 2013, Foz do Iguaçu. **Anais...** São José dos Campos: INPE, 2013.

MAY, D. B.; SIVAKUMAR, M. Prediction of urban stormwater quality using artificial neural networks. **Environmental Modelling & Software**, v. 24, p. 296-302, 2009.

MOMEM, A.; UDDIN, E.; HASSAN, E. Discharge, rainfall and pollution load on Dhaleshwari River: a study on trend analysis. In: INTERNATIONAL CONFERENCE ON CIVIL ENGINEERING FOR SUSTAINABLE DEVELOPMENT, 5, 2020, Khulna. **Anais...** Khulna: KUET, 2020.

MONJO, R. Measure of rainfall time structure using the dimensionless n-index. **Climate Research**, v. 67, p. 71-86, 2016.

MORUZZI, R. B. et al. Avaliação de cargas difusas e simulação de autodepuração no Córrego da Água Branca, Itirapina (SP). **Geociências**, v. 31, n. 3, p. 447-458, 2012.

MUHLENHOFF, A; KOZAK, C; IVANCHENECN, F; TORRES, M; KYOCHI, V. **Avaliação da qualidade da água a partir de espectros de fluorescência e estimativa de carga associados ao uso e ocupação do solo: Estudo de Caso na Bacia do Rio Barigui**.

NUNES, F. G.; FIORI, A. P.; FIRPO, G. Estimativa de coeficientes de escoamento superficial na bacia hidrográfica do rio Atuba: Curitiba e região metropolitana – Paraná/ Brasil. **Boletim Paranaense de Geociências**, v. 64-65, p. 27-39, 2011.

OLIVEIRA, C. P. M. **Plataforma de análise e simulação hidrológica PLASH**. 183 f. Dissertação (Mestrado em Engenharia) – Escola Politécnica, Universidade de São Paulo, São Paulo, 2012.

OLIVEIRA, L. M. **Análise da Relação entre Uso do Solo e Liberação dos Nutrientes Fósforo e Nitrogênio nas Águas Superficiais do Rio das Velhas utilizando Árvores de Decisão e Regressão Múltipla**. Tese (Doutorado em Geografia) – Departamento de Geografia, Universidade Federal de Minas Gerais, Belo Horizonte, 2016.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS (ONU). **Transformando Nosso Mundo: A Agenda 2030 para o Desenvolvimento Sustentável**, 2015. Disponível em: <<https://nacoesunidas.org/pos2015/agenda2030/>>. Acesso em: 10 mai. 2020.

PALMERI, L.; BENDORICCHIO, G.; ARTIOLI, Y. Modelling nutrient emissions from river systems and loads to the coastal zone: Po River case study, Italy. **Ecological Modelling**, v. 184, p. 37-53, 2005.

PEREIRA, G. H. A.; **Avaliação do uso combinado de métodos de modelagem da direção do escoamento superficial e modelos digitais de superfície derivados de laser scanner em regiões urbanas**. 2013. 182 p. Dissertação (Mestrado em Engenharia de Recursos Hídricos e Ambiental) – Setor de Tecnologia, Universidade Federal do Paraná, Curitiba, 2013.

PORRAS, E. A. A. **Estudo da estimativa de cargas de fósforo aplicando MoRE (modeling of regionalized emissions) na bacia hidrográfica do Alto Iguaçu – Paraná**. 127 f. Dissertação (Mestrado em Engenharia de Recursos Hídricos e Ambiental) – Setor de Tecnologia, Universidade Federal do Paraná, Curitiba, 2014.

PORRAS, E. A. A. et al. Estimation of phosphorus emissions in the Upper Iguazu Basin (Brazil) using GIS and the MoRE model. In: INTERNATIONAL SOCIETY FOR PHOTOGRAMMETRY AND REMOTE SENSING CONGRESS, 13, 2016, Praga. **Anais...** Nice: ISPRS, 2016.

RUSSEL, S. J.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. New Jersey, Prentice-Hall, 1995.

SARKAR, A.; PANDEY, P. River Water Quality Modelling using Artificial Neural Network Technique. **Aquatic Procedia**, v. 4, p. 1070-1077, 2015.

SECRETARIA DE ESTADO DO MEIO AMBIENTE DE SÃO PAULO (SEMA). **Elaboração do Plano de Desenvolvimento e Proteção Ambiental da Bacia Hidrográfica do Reservatório Billings: Relatório Final**. São Paulo: SEMA, 2010.

SERRANO, L. O. et al. A New Approach to Use Load Duration Curves to Evaluate Water Quality: A Study in the Doce River Basin, Brazil. **Water**, v. 12, n. 811, 2020.

SHEN, Z; QIU, J; HONG, Q; CHEN, L. **Simulation of spatial and temporal distributions of non-point source pollution load in the Three Gorges Reservoir Region.** Science of the Total Environment 493. 138-146, 2014.

SHRESTHA, S; KAZAMA, F; NEWHAM, L. **A framework for estimating pollutant export coefficients from long-term in-stream water quality monitoring data.** Environmental Modelling & Software 23. 182-194, 2008.

SILVA, I. V. M. **Modelagem da qualidade da água e capacidade de autodepuração do Rio das Antas-RS.** Monografia (Bacharelado em Engenharia Sanitária e Ambiental) – Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2009.

SODRÉ, F. F. Fontes Difusas de Poluição da Água: Características e métodos de controle. **AQUA**, v. 1, p. 9-16, 2012.

SPERLING, M. V. **Introdução à qualidade das águas e ao tratamento de esgotos.** Editora UFMG, volume 1, 4ª edição. 2017.

TEODORO, A. et al. Implementação do conceito Capacidade de Diluição de Efluentes no modelo de qualidade da água QUAL-UFMG: estudo de caso no Rio Taquarizinho (MS). **Engenharia Sanitária Ambiental**, v. 18, n. 3, p. 275-288, 2013.

TERCINI, J. F. B. **Modelagem da qualidade da água integrando rio e reservatório.** Dissertação (Mestrado em Ciências) – Escola Politécnica, Universidade de São Paulo, São Paulo, 2014.

TSCHIEDEL, A. F. **Modelagem da qualidade de água para cenários futuros através do modelo QUAL- 2k – um estudo de caso para o Rio Macaé, RJ.** Monografia (Bacharelado em Engenharia Ambiental) – Instituto de Pesquisas Hidráulicas, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2013.

TUCCI, C. E. M. **Modelos Hidrológicos.** 2. ed. Porto Alegre: Editora UFRGS, 2005.

TUCCI, C. E. M. **Hidrologia: ciência e aplicação.** 4. ed. Porto Alegre: Editora UFRGS, 2015.

UNITED STATES GEOLOGICAL SURVEY (USGS). **O Ciclo da Água.** 2019. Disponível em: <https://www.usgs.gov/special-topics/water-science-school/science/o-ciclo-dagua-water-cycle-portuguese?qt-science_center_objects=0#qt-science_center_objects>. Acesso em: 10 mai. 2020.

WANG, K; WANG, P; ZHANG, R; LIN, Z. **Determination of spatiotemporal characteristics of agricultural non-point source pollution of river basins using the dynamic time warping distance.** Journal of Hydrology. 2019.

ZHAO, J; LIN, L; YANG, K; LIU, Q; QIAN, G. **Influences of land use on water quality in a reticular river network area: A case study in Shanghai, China.** Landscape and Urban Planning 137. 20-29, 2015.

6. APÊNDICE A – Programa em Python de Regressão Linear para Dados de Concentração Durante Eventos de Precipitação

Planilha de dados de entrada

Colunas =

```
[Amostra', 'Ponto', 'Evento', 'Tempo de Concentração (min)', 'Tempo
desde chuva inicial (min)', 'Concentração NT (mg/L)', 'Concentração PT
(mg/L)', 'Concentração DOC (mg/L)', 'Precipitação Acumulada - dentro do
tempo de concentração (mm)', 'Precipitação Acumulada - desde chuva
inicial (mm)', 'Precipitação Antecedente 1 dia (mm)', 'Precipitação
Antecedente 3 dias (mm)', 'Precipitação Antecedente 5 dias (mm)',
'Precipitação Antecedente 7 dias (mm)', 'Vazão no rio (m³/s)',
'Hydrograma sobe (2) ou desce (1)', 'Evapotranspiração real', 'Área da
bacia (km²)', 'Declividade média (m/m)', 'Tipo de Solo - Latossolo
vermelho-amarelo (%)', 'Tipo de Solo - Cambissolo Álico (%)', 'Tipo de
Solo - Podzólico Vermelho-amarelo Álico (%)', 'Tipo de Solo - Latossolo
Vermelho-escuro Distrófico (%)', 'Tipo de Solo - Curitiba (%)', 'Tipo
de Solo - Podzólico Vermelho-amarelo Eutrófico (%)', 'Uso do Solo -
Área Urbana Baixa (%)', 'Uso do Solo - Vegetação Arbórea Natural (%)',
'Uso do Solo - Campo (%)', 'Uso do Solo - Água (%)', 'Uso do Solo -
Vegetação Arbórea Plantada (%)', 'Uso do Solo - Vegetação Arbustiva
Natural (%)', 'Uso do Solo - Mineração/Outros (%)', 'Uso do Solo - Área
Industrial (%)', 'Uso do Solo - Vila (%)', 'Uso do Solo - Cultura
Temporária (%)', 'Uso do Solo - Área Alagada (%)', 'Uso do Solo -
Granja (%)', 'Uso do Solo - Cultura Permanente (%)', 'Uso do Solo -
Solo exposto (%)', 'Uso do Solo - Área Urbana Média (%)', 'Uso do Solo
- Loteamentos (%)', 'Uso do Solo - Lixão (%)', 'Precipitação
Antecedente 1 dia (mm) - Thiessen', 'Precipitação Antecedente 3 dias
(mm) - Thiessen', 'Precipitação Antecedente 5 dias (mm) - Thiessen',
'Precipitação Antecedente 7 dias (mm) - Thiessen', 'Precipitação
Antecedente 1 dia (mm) - 2549100', 'Precipitação Antecedente 3 dias
(mm) - 2549100', 'Precipitação Antecedente 5 dias (mm) - 2549100',
'Precipitação Antecedente 7 dias (mm) - 2549100', 'Precipitação
Antecedente 1 dia (mm) - 2549107', 'Precipitação Antecedente 3 dias
(mm) - 2549107', 'Precipitação Antecedente 5 dias (mm) - 2549107',
'Precipitação Antecedente 7 dias (mm) - 2549107', 'Precipitação
Antecedente 1 dia (mm) - 2549128', 'Precipitação Antecedente 3 dias
(mm) - 2549128', 'Precipitação Antecedente 5 dias (mm) - 2549128',
'Precipitação Antecedente 7 dias (mm) - 2549128', 'Precipitação
Antecedente 1 dia (mm) - 2549077', 'Precipitação Antecedente 3 dias
(mm) - 2549077', 'Precipitação Antecedente 5 dias (mm) - 2549077',
'Precipitação Antecedente 7 dias (mm) - 2549077', 'Precipitação
Antecedente 1 dia (mm) - 2549090', 'Precipitação Antecedente 3 dias
(mm) - 2549090', 'Precipitação Antecedente 5 dias (mm) - 2549090',
'Precipitação Antecedente 7 dias (mm) - 2549090', 'Precipitação
Antecedente 1 dia (mm) - 2549006', 'Precipitação Antecedente 3 dias (mm)
- 2549006', 'Precipitação Antecedente 5 dias (mm) - 2549006',
'Precipitação Antecedente 7 dias (mm) - 2549006']
```

Programa:

```

# Instalar lazypredict
!pip install lazypredict

# Imports
import pandas as pd
from lazypredict.Supervised import LazyRegressor
from sklearn.model_selection import train_test_split
import numpy as np
import itertools
from sklearn.preprocessing import StandardScaler
import random
from itertools import zip_longest

# Ler tabela de dados
data = pd.read_excel('Dados_de_entrada.xlsx' ,0)
data.head(20)

# Eliminar células em branco das variáveis a serem usadas
data.dropna(inplace=True)
data

#Iniciando vetores
Resposta_R2 = []
Resposta_Regressor = []
Resposta_Colunas = []
Resposta_R2ajustado = []
Resposta_RMSE = []
Numero_combinacoes = 4500

#Programa
for L in range(Numero_combinacoes):
    Colunas_selecionadas=[]
    N_colunas = random.randint(1,len(Colunas)) #Escolhe aleatoriamente
o número de colunas a serem utilizadas
    Coluna_aleatoria = random.sample(range(0,len(Colunas)),N_colunas)
#Aleatoriamente escolhe as colunas a serem utilizadas, sem repetição
    for k in range(N_colunas):
        Nome_Coluna = Colunas[Coluna_aleatoria[k]]
        Colunas_selecionadas.append(Nome_Coluna)

```

```

colunas_usadas = list(Colunas_selecionadas)
X, y = data[colunas_usadas], data['Concentração NT (mg/L)'].values
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=.2, random_state=42) #separa variáveis de treino e de teste
ss = StandardScaler() #normalizar variáveis
X_train_scaled = ss.fit_transform(X_train)
X_test_scaled = ss.transform(X_test)
reg = LazyRegressor(predictions=True)
models, predictions = reg.fit(X_train_scaled, X_test_scaled,
y_train, y_test)

#correção de erros
if models.empty == False:
    for i in range (len(models.index)):
        if models.loc[models.index[i], 'R-Squared'] > 1:
            models.loc[models.index[i], 'Rquadrado arrumado'] = 0
        else:
            models.loc[models.index[i], 'Rquadrado arrumado'] =
models.loc[models.index[i], 'R-Squared']

# 11 melhores combinações
if len(Resposta_R2)<11:
    Resposta_R2.append(models['Rquadrado arrumado'].max())
    Resposta_Regressor.append(models.index[models['Rquadrado
arrumado'].argmax()])
    Resposta_Colunas.append(colunas_usadas)
    Resposta_R2ajustado.append(models.iat[models['Rquadrado
arrumado'].argmax(),0])
    Resposta_RMSE.append(models.iat[models['Rquadrado
arrumado'].argmax(),2])
else:
    Menor_da_lista = Resposta_R2.index(min(Resposta_R2))
    Resposta_R2.pop(Menor_da_lista)
    Resposta_Regressor.pop(Menor_da_lista)
    Resposta_Colunas.pop(Menor_da_lista)
    Resposta_R2.append(models['Rquadrado arrumado'].max())
    Resposta_Regressor.append(models.index[models['Rquadrado
arrumado'].argmax()])
    Resposta_Colunas.append(colunas_usadas)

```

7. APÊNDICE B – Variáveis e Siglas

Número de identificação da variável	Variável - sigla	Variável
1	Ponto	Ponto
2	Evento	Evento
3	tc	Tempo de Concentração (min)
4	Tempo desde chuva inicial	Tempo desde chuva inicial (min)
5	Concentração NT	Concentração NT (mg/L)
6	Concentração PT	Concentração PT (mg/L)
7	Concentração DOC	Concentração DOC (mg/L)
8	Pacum - dentro do tc	Precipitação Acumulada - dentro do tempo de concentração (mm)
9	Pacum - desde chuva inicial	Precipitação Acumulada - desde chuva inicial (mm)
10	Pacum antecedente 1 dia	Precipitação Antecedente 1 dia (mm)
11	Pacum antecedente 3 dia	Precipitação Antecedente 3 dias (mm)
12	Pacum antecedente 5 dia	Precipitação Antecedente 5 dias (mm)
13	Pacum antecedente 7 dia	Precipitação Antecedente 7 dias (mm)
14	Qrio	Vazão no rio (m ³ /s)
15	Condição hidrograma	Hidrograma sobe (2) ou desce (1)
16	Evapotranspiração real	Evapotranspiração real
17	Área da bacia	Área da bacia (km ²)
18	Declividade média	Declividade média (m/m)
19	Tipo de Solo - LVA	Tipo de Solo - Latossolo vermelho-amarelo (%)
20	Tipo de Solo - CA	Tipo de Solo - Cambissolo Álico (%)
21	Tipo de Solo - PVAE	Tipo de Solo - Podzólico Vermelho-amarelo Eutrófico (%)
22	Tipo de Solo - LVED	Tipo de Solo - Latossolo Vermelho-escuro Distrófico (%)
23	Tipo de Solo - C	Tipo de Solo - Curitiba (%)
24	Tipo de Solo - PVAA	Tipo de Solo - Podzólico Vermelho-amarelo Álico (%)
25	Uso do Solo - AUB	Uso do Solo - Área Urbana Baixa (%)
26	Uso do Solo - VAN	Uso do Solo - Vegetação Arbórea Natural (%)
27	Uso do Solo - C	Uso do Solo - Campo (%)
28	Uso do Solo - A	Uso do Solo - Água (%)
29	Uso do Solo - VAP	Uso do Solo - Vegetação Arbórea Plantada (%)
30	Uso do Solo - VAN	Uso do Solo - Vegetação Arbustiva Natural (%)
31	Uso do Solo - M/O	Uso do Solo - Mineração/Outros (%)
32	Uso do Solo - AI	Uso do Solo - Área Industrial (%)
33	Uso do Solo - V	Uso do Solo - Vila (%)
34	Uso do Solo - CT	Uso do Solo - Cultura Temporária (%)
35	Uso do Solo - AA	Uso do Solo - Área AIBYada (%)
36	Uso do Solo - G	Uso do Solo - Granja (%)
37	Uso do Solo - CP	Uso do Solo - Cultura Permanente (%)
38	Uso do Solo - SE	Uso do Solo - Solo exposto (%)
39	Uso do Solo - AUM	Uso do Solo - Área Urbana Média (%)
40	Uso do Solo - Lo	Uso do Solo - Loteamentos (%)
41	Uso do Solo - Li	Uso do Solo - Lixão (%)
42	P antecedente 1 dia- Thiessen	Precipitação Antecedente 1 dia (mm) - Thiessen
43	P antecedente 3 dia- Thiessen	Precipitação Antecedente 3 dias (mm) - Thiessen
44	P antecedente 5 dia- Thiessen	Precipitação Antecedente 5 dias (mm) - Thiessen
45	P antecedente 7 dia- Thiessen	Precipitação Antecedente 7 dias (mm) - Thiessen

46	P antecedente 1 dia- 2549100	Precipitação Antecedente 1 dia (mm) - 2549100
47	P antecedente 3 dia- 2549100	Precipitação Antecedente 3 dias (mm) - 2549100
48	P antecedente 5 dia- 2549100	Precipitação Antecedente 5 dias (mm) - 2549100
49	P antecedente 7 dia- 2549100	Precipitação Antecedente 7 dias (mm) - 2549100
50	P antecedente 1 dia- 2549107	Precipitação Antecedente 1 dia (mm) - 2549107
51	P antecedente 3 dia- 2549107	Precipitação Antecedente 3 dias (mm) - 2549107
52	P antecedente 5 dia- 2549107	Precipitação Antecedente 5 dias (mm) - 2549107
53	P antecedente 7 dia- 2549107	Precipitação Antecedente 7 dias (mm) - 2549107
54	P antecedente 1 dia- 2549128	Precipitação Antecedente 1 dia (mm) - 2549128
55	P antecedente 3 dia- 2549128	Precipitação Antecedente 3 dias (mm) - 2549128
56	P antecedente 5 dia- 2549128	Precipitação Antecedente 5 dias (mm) - 2549128
57	P antecedente 7 dia- 2549128	Precipitação Antecedente 7 dias (mm) - 2549128
58	P antecedente 1 dia- 2549077	Precipitação Antecedente 1 dia (mm) - 2549077
59	P antecedente 3 dia- 2549077	Precipitação Antecedente 3 dias (mm) - 2549077
60	P antecedente 5 dia- 2549077	Precipitação Antecedente 5 dias (mm) - 2549077
61	P antecedente 7 dia- 2549077	Precipitação Antecedente 7 dias (mm) - 2549077
62	P antecedente 1 dia- 2549090	Precipitação Antecedente 1 dia (mm) - 2549090
63	P antecedente 3 dia- 2549090	Precipitação Antecedente 3 dias (mm) - 2549090
64	P antecedente 5 dia- 2549090	Precipitação Antecedente 5 dias (mm) - 2549090
65	P antecedente 7 dia- 2549090	Precipitação Antecedente 7 dias (mm) - 2549090
66	P antecedente 1 dia- 2549006	Precipitação Antecedente 1 dia (mm) - 2549006
67	P antecedente 3 dia- 2549006	Precipitação Antecedente 3 dias (mm) - 2549006
68	P antecedente 5 dia- 2549006	Precipitação Antecedente 5 dias (mm) - 2549006
69	P antecedente 7 dia- 2549006	Precipitação Antecedente 7 dias (mm) - 2549006