Encontro Técnico AESABESP
37º Congresso Nacional de Saneamento e Meio Ambiente

AESABESP 40 Anos
Associação dos Engenheiros da Sabesp

AVALIAÇÃO DO IMPACTO DE VARIÁVEIS HIDROMETEOROLÓGICAS NA PREDIÇÃO DE DOSAGEM DE COAGULANTE EM ESTAÇÕES DE TRATAMENTO DE ÁGUA

37ETC-07420

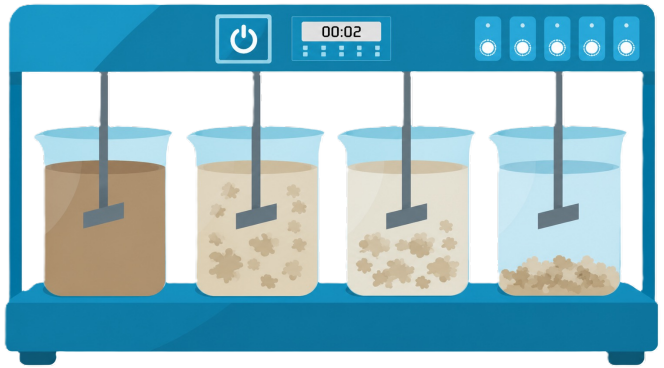
Juliana Neves
Rodrigo Campos de Andrade
Natan Padoin
Cíntia Soares

UFSC
labMAC+
Laboratório de Materiais e Computação Científica
Laboratory of Materials and Scientific Computing
CNPq
casan
HydroInfo
Smart Solutions

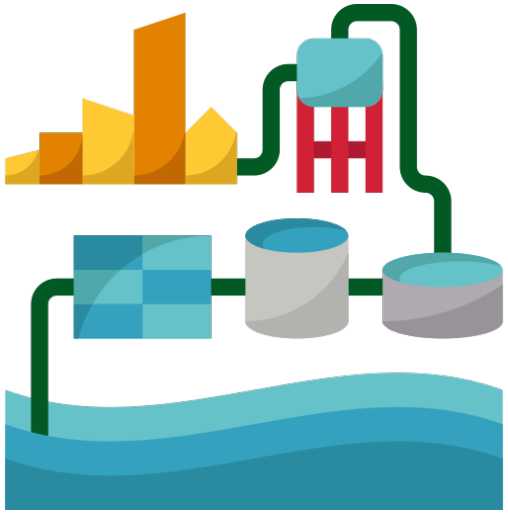
Contextualização

Encontro Técnico AESABESP
37º Congresso Nacional de Saneamento e Meio Ambiente

Estações de Tratamento de Água (ETAs) são exemplos de sistemas complexos que envolvem o monitoramento de variáveis com relações altamente não-lineares entre si



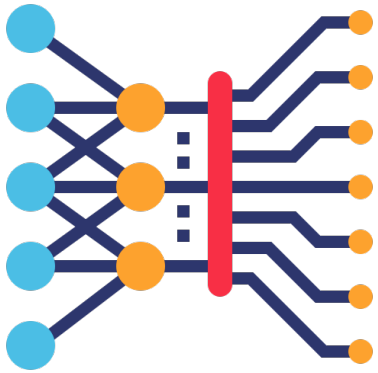
Cenário atual de digitalização, otimização de processos e aplicação de inteligência artificial



Gastos com produtos químicos em ETAs podem chegar a 57 % do total de despesas anuais da estação e há dificuldades para realização de Jar test com alta frequência



Contextualização



A precipitação pluviométrica é vista como um dos principais fatores ambientais determinantes para a antecipação de ajustes na dosagem química em processos de tratamento de água para abastecimento

Técnicas de *machine learning* têm grande potencial como ferramentas de apoio na definição da dosagem ideal para o processo



Casos práticos da literatura têm demonstrado a importância de se considerar este fator para a melhoria da acurácia dos modelos

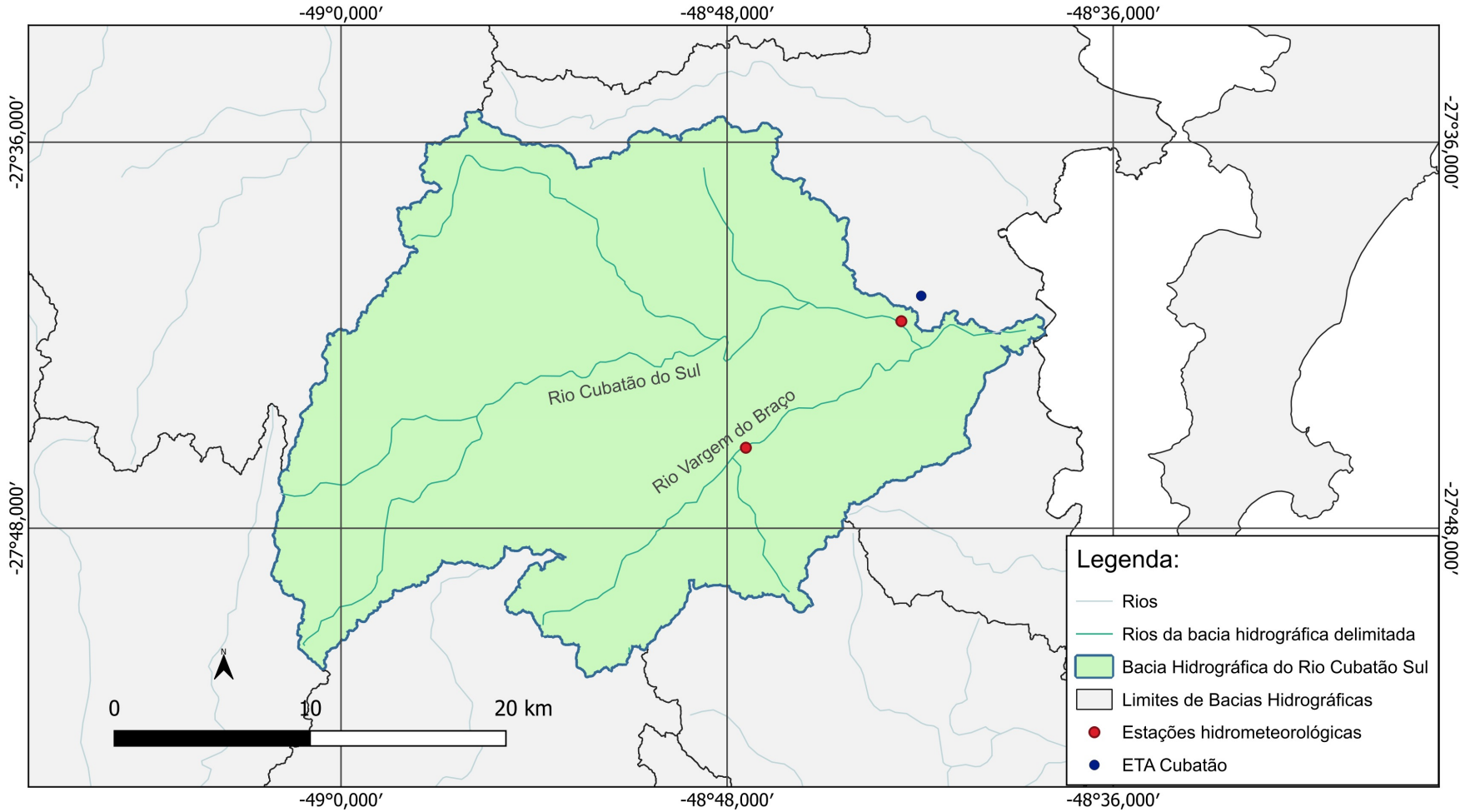
Objetivo

Encontro Técnico AESABESP
37º Congresso Nacional de Saneamento e Meio Ambiente

Avaliar o desempenho de dois modelos distintos de ***machine learning*** ao incorporar dados de qualidade da água bruta, decantada, vazão dos rios de captação e registros de precipitação pluviométrica na bacia hidrográfica para previsão da dosagem de coagulante em uma ETA

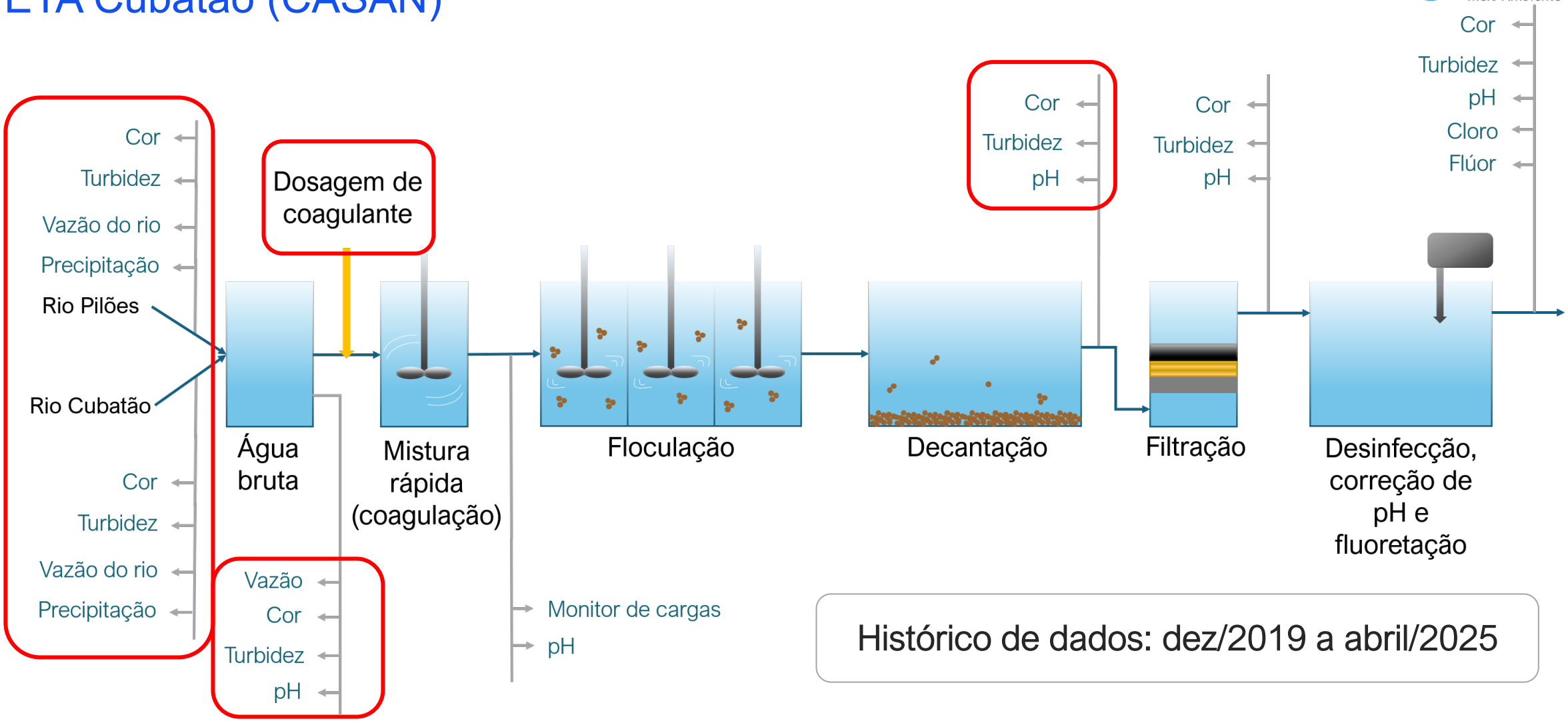
Metodologia

ETA Cubatão (CASAN)



Metodologia

ETA Cubatão (CASAN)



Histórico de dados: dez/2019 a abril/2025

Metodologia

Pré-processamento dos dados

Análise e tratamento de dados ausentes → Remoção de linhas com dados ausentes e preenchimento com registros de precipitação da base do INMET

Identificação e remoção de *outliers* → Critério de 3σ (três sigma) aliado à definição do *Z-score* (não aplicado aos dados de vazão dos rios e precipitação)

Feature engineering → Eventos extremos de precipitação, eficiência de remoção de turbidez, atraso temporal

Divisão de cenários

- Cenário 1: dados de monitoramento da ETA
- Cenário 2: dados de monitoramento + precipitação
- Cenário 3: dados de monitoramento + precipitação + atraso temporal

Metodologia

Modelos de *machine learning*

XGBoost
(*Extreme Gradient Boosting*)

LSTM
(*Long Short-Term Memory*)

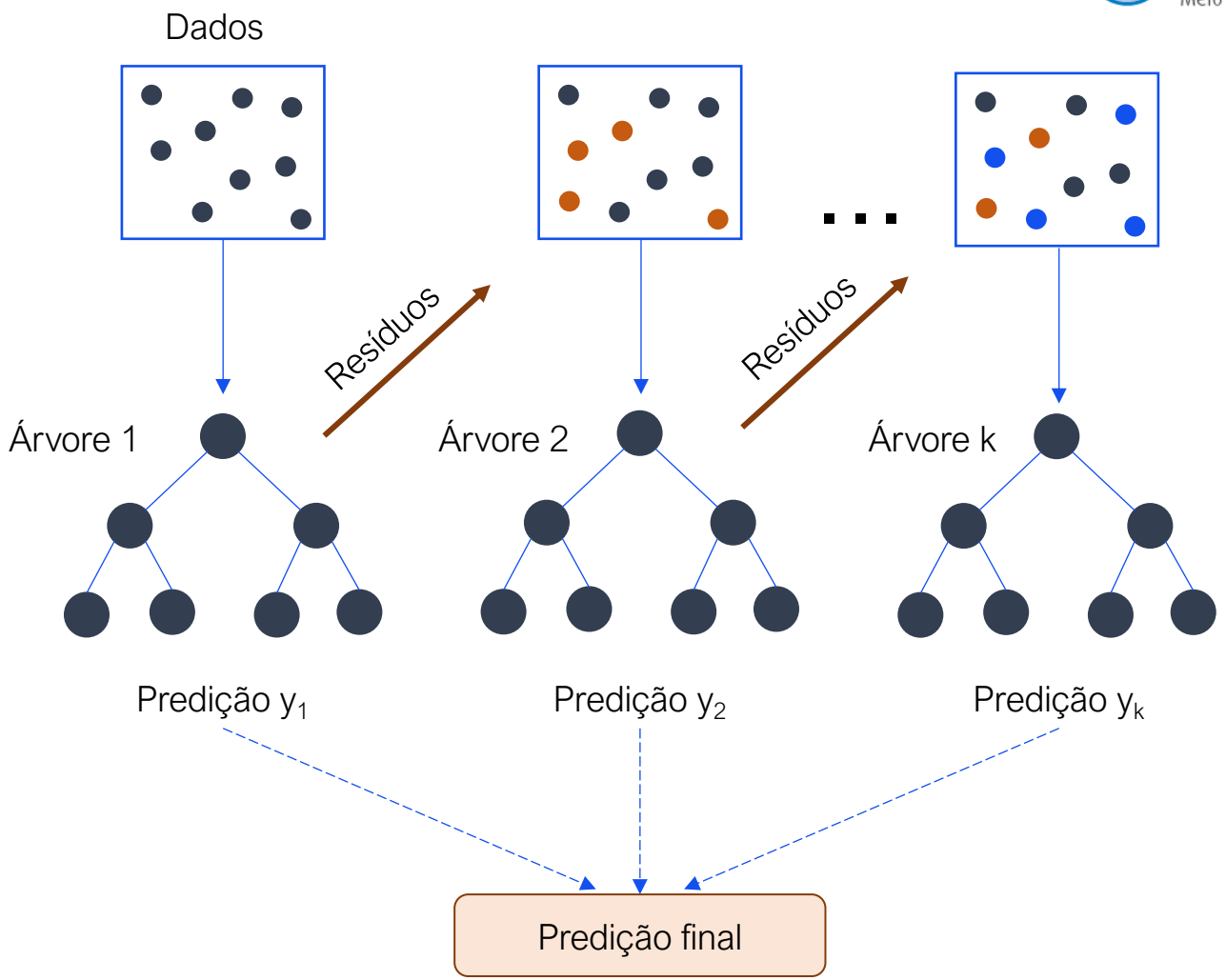
Algoritmo baseado em árvores de decisão que aplica um método sequencial de construção nomeado *boosting*, no qual novos modelos são adicionados iterativamente para corrigir os erros residuais dos preditores anteriores

Metodologia

Modelos de *machine learning*

XGBoost
(*Extreme Gradient Boosting*)

LSTM
(*Long Short-Term Memory*)



Metodologia

Modelos de *machine learning*

XGBoost
(*Extreme Gradient Boosting*)

LSTM
(*Long Short-Term Memory*)

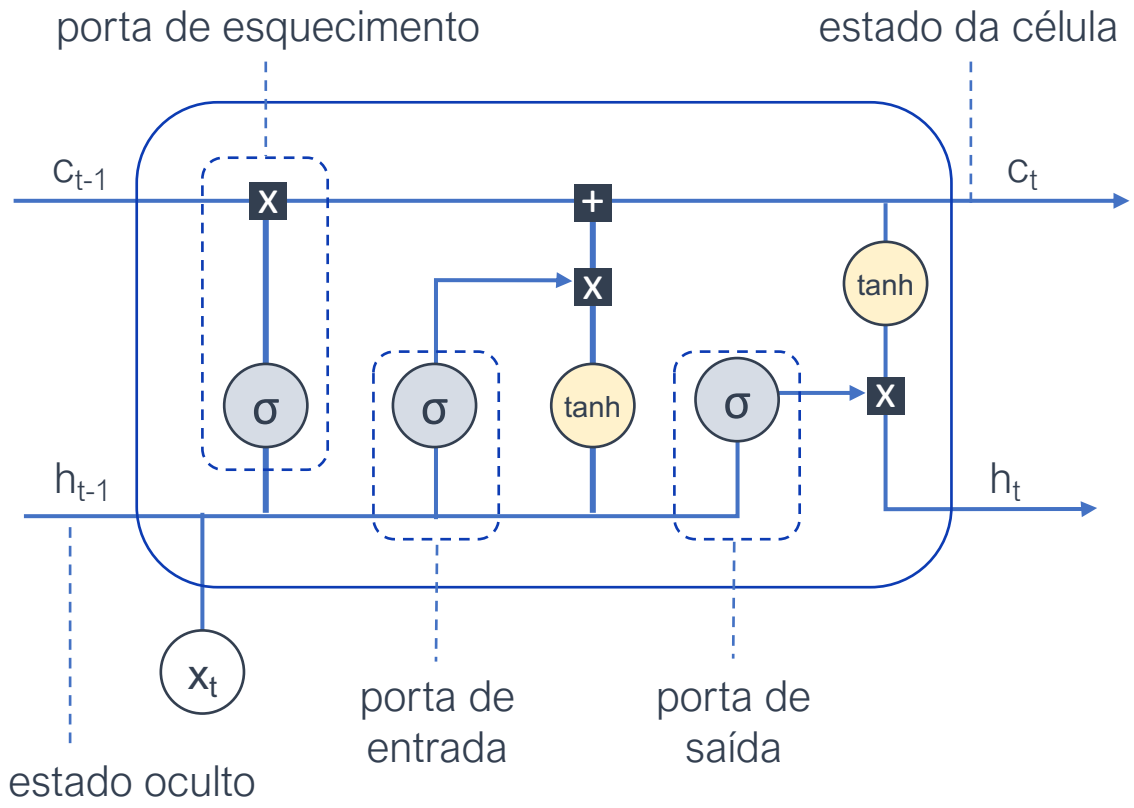
→ Rede neural especificamente projetada para dados sequenciais ou com dependência temporal

Metodologia

Modelos de *machine learning*

XGBoost
(*Extreme Gradient Boosting*)

LSTM
(*Long Short-Term Memory*)



Metodologia

Implementação e treinamento dos modelos

OPTUNA

Encontro Técnico AESABESP
37º Congresso Nacional de Saneamento e Meio Ambiente

XGBoost
(*Extreme Gradient Boosting*)

LSTM
(*Long Short-Term Memory*)

Divisão em dados de treino (80%) e teste (20%) em ordem cronológica

↓

Escalonamento do dados (para modelo LSTM)

↓

Otimização de hiperparâmetros via *Optuna*

↓

Validação cruzada com *TimeSeriesSplit*

↓

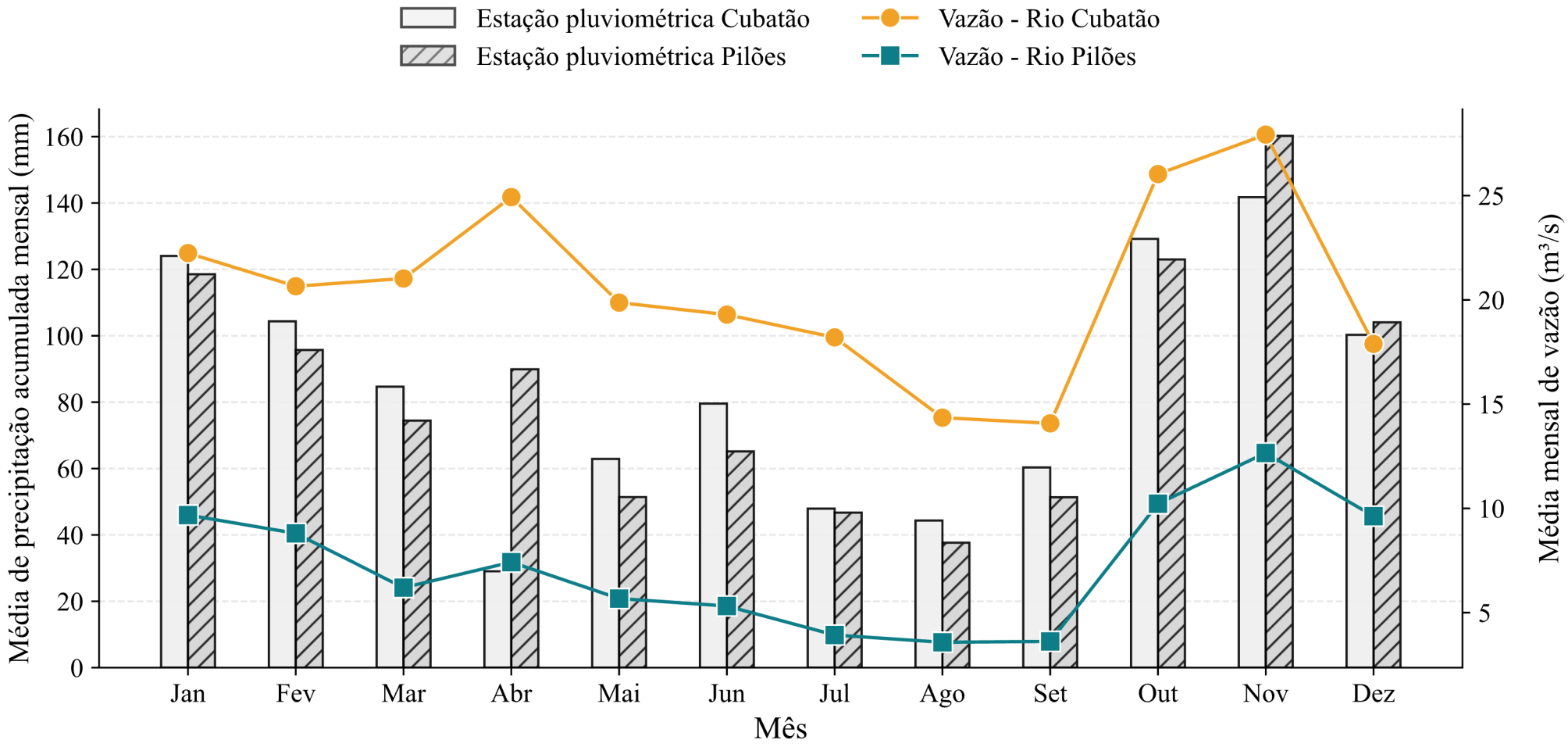
Treino final

↓

Permutation importance

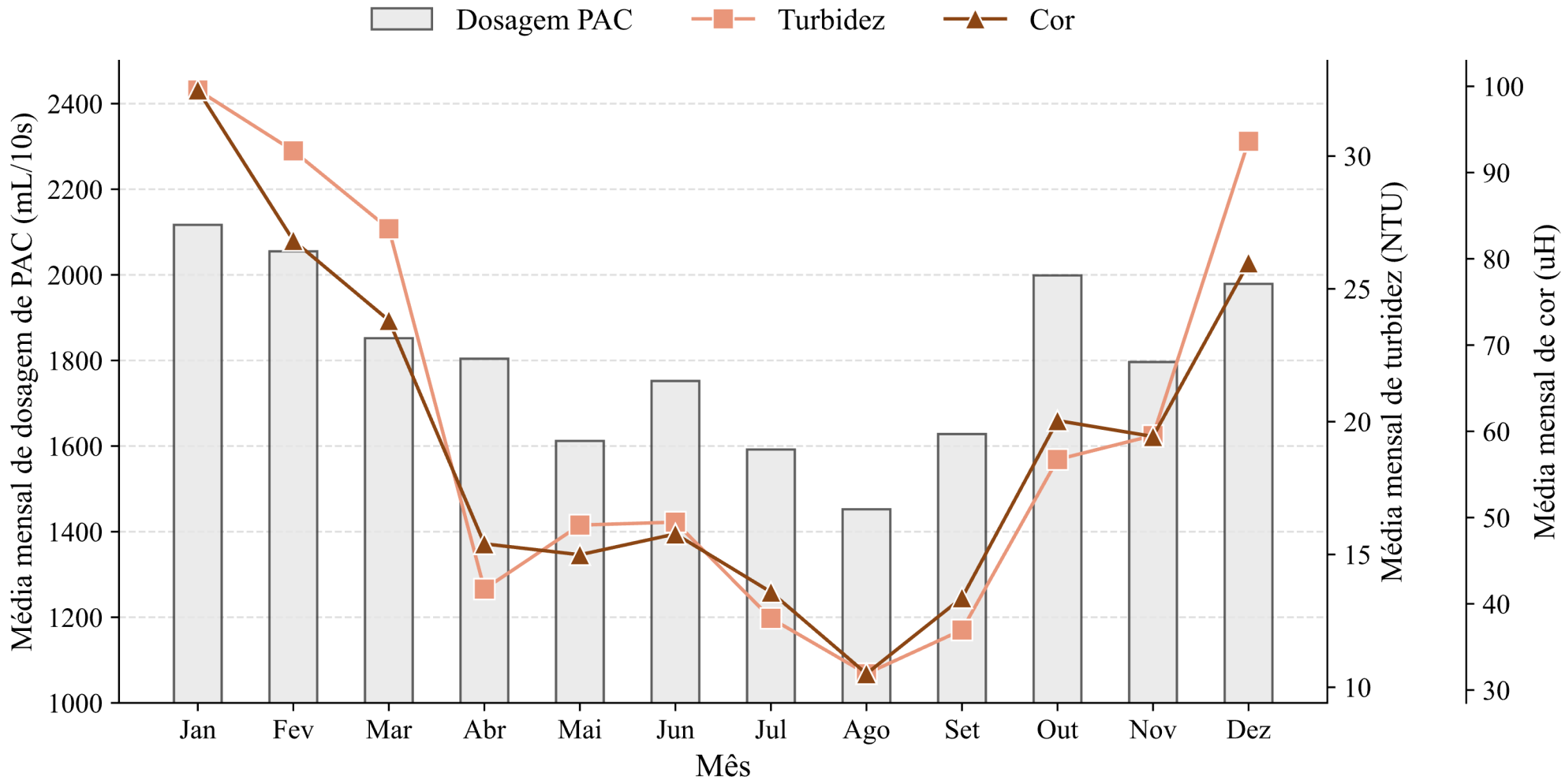
Resultados e discussão

Análise dos dados



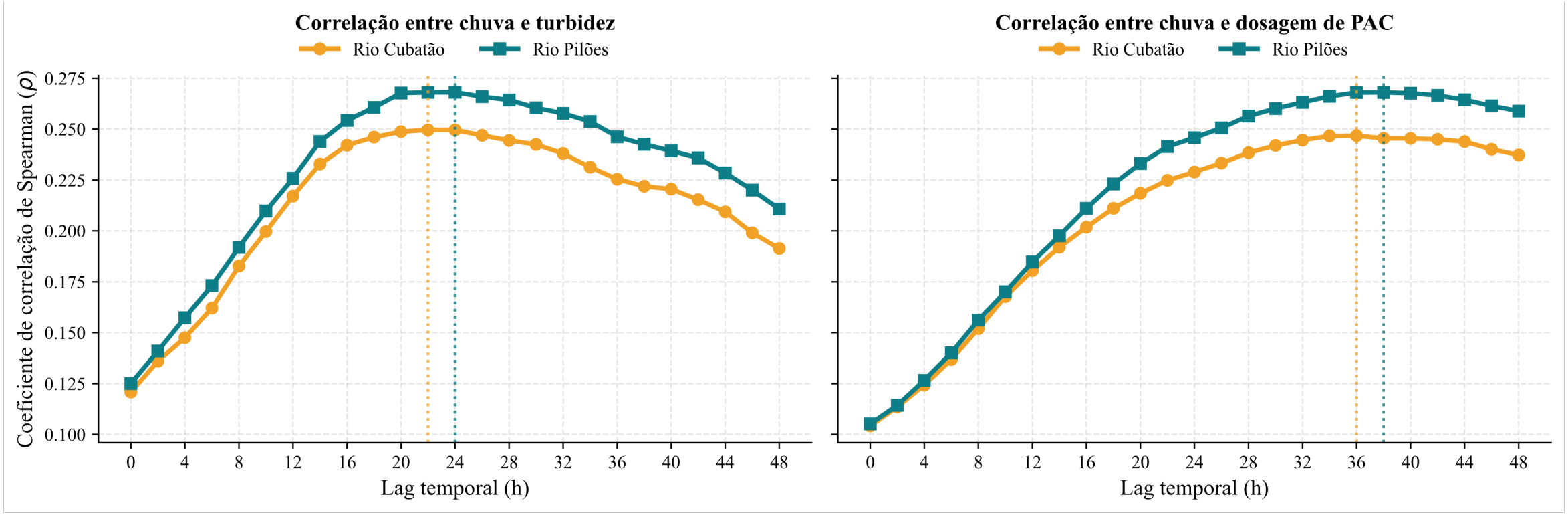
Resultados e discussão

Análise dos dados



Resultados e discussão

Análise dos dados



Resultados e discussão

Desempenho dos modelos de previsão

Modelo	Cenário	Métrica			
		RMSE	MAE	R^2	r
XGBoost	1	388,01	315,74	0,34	0,67
	2	446,47	369,35	0,18	0,66
	3	504,82	414,76	-0,05	0,63
LSTM	1	367,07	299,29	0,41	0,67
	2	490,84	392,46	0,01	0,59
	3	585,11	464,94	-0,41	0,47

Resultados e discussão

Desempenho dos modelos de previsão

Modelo	Cenário	Métrica			
		RMSE	MAE	R^2	r
XGBoost	1	388,01	315,74	0,34	0,67
	2	446,47	369,35	0,18	0,66
	3	504,82	414,76	-0,05	0,63
LSTM	1	367,07	299,29	0,41	0,67
	2	490,84	392,46	0,01	0,59
	3	585,11	464,94	-0,41	0,47

Resultados e discussão

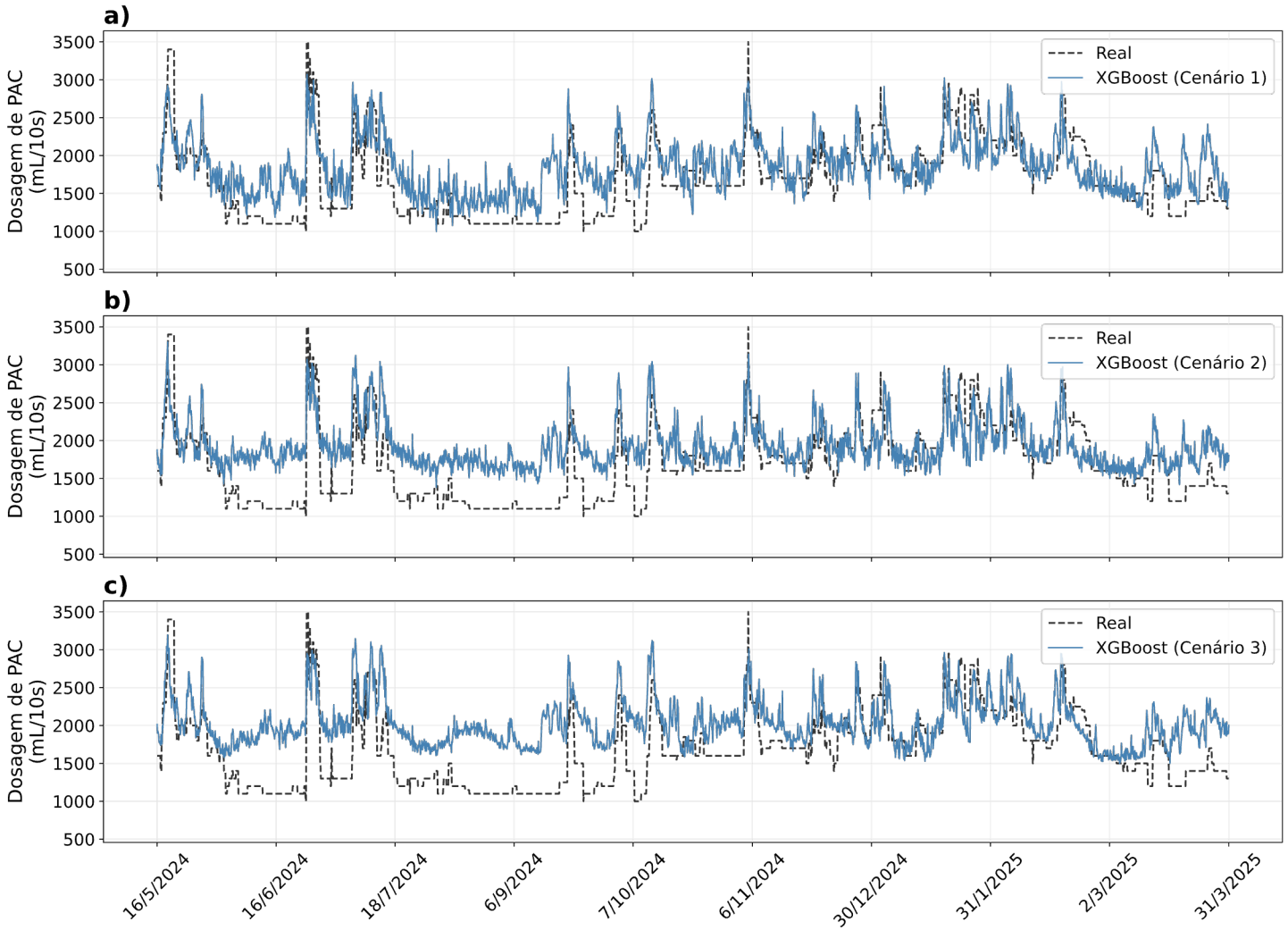
Desempenho dos modelos de previsão

Modelo	Cenário	Métricas	
		R^2	r
XGBoost	1	0,34	0,67
	2	0,18	0,66
	3	-0,05	0,63
LSTM	1	0,41	0,67
	2	0,01	0,59
	3	-0,41	0,47

Modelo	R^2	r	Referência
XGBoost	0,973	-	(Hu et al., 2023)
XGBoost	-	0,95	(Dunnington et al., 2021)
LSTM	0,62	-	(Lin et al., 2023b)
LSTM	0,70	-	(Wang; Asefa; Thornburgh, 2022)
ANN	0,85	-	(Haghiri et al., 2018)
GRU	0,88	0,94	(Lin et al., 2023a)
GRU	0,94	-	(Lin et al., 2023b)

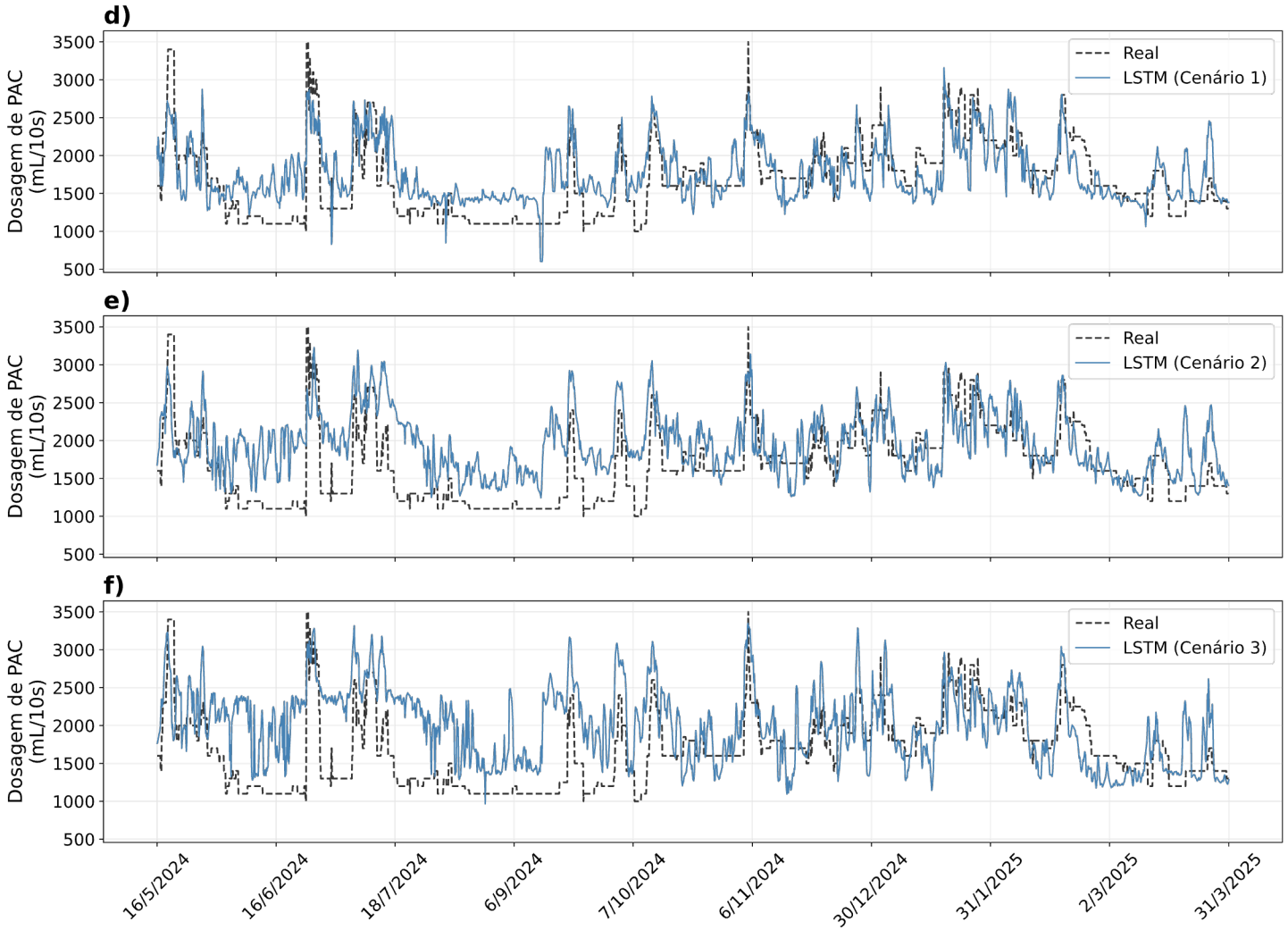
Resultados e discussão

Desempenho dos modelos de previsão



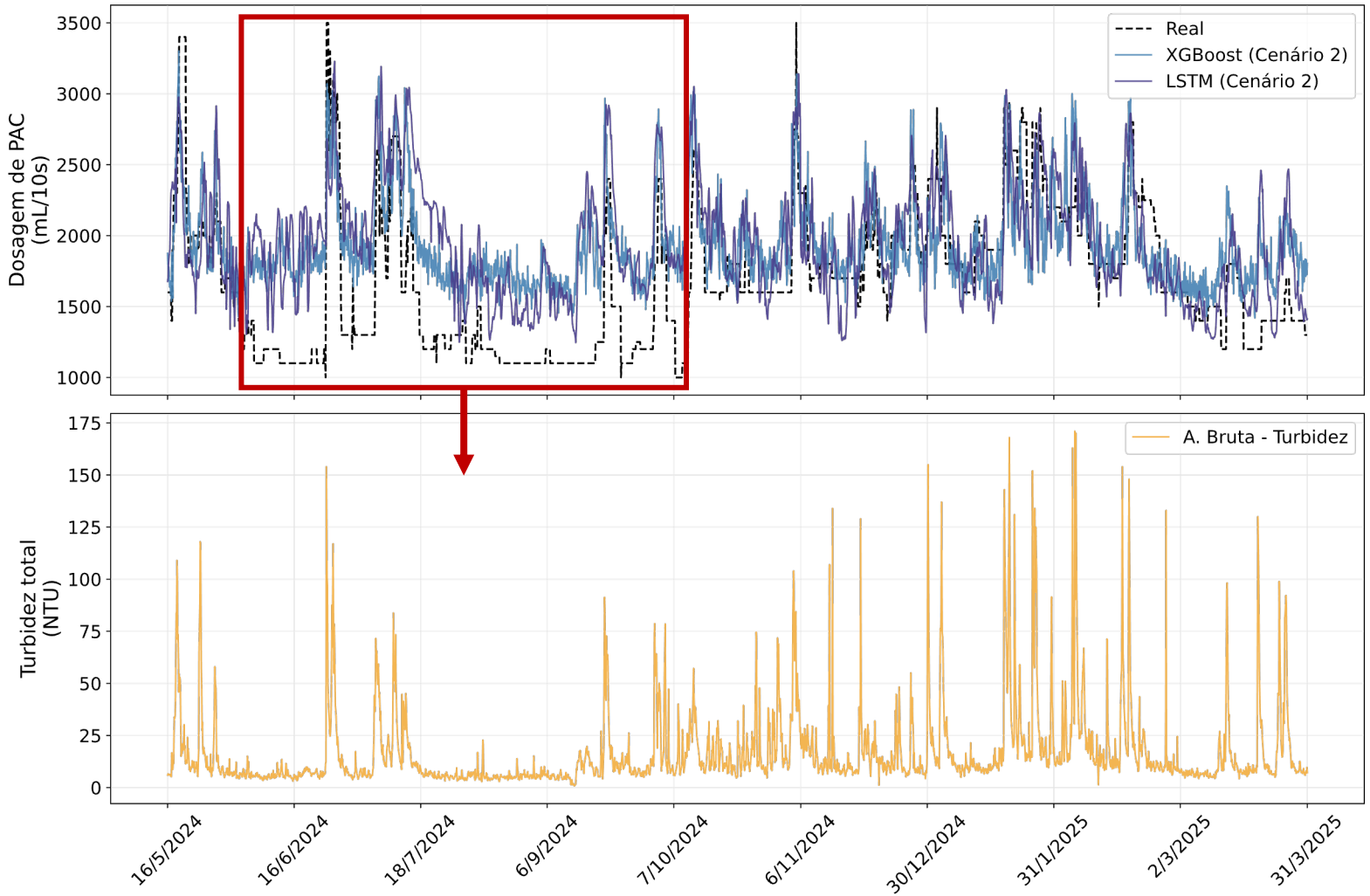
Resultados e discussão

Desempenho dos modelos de previsão



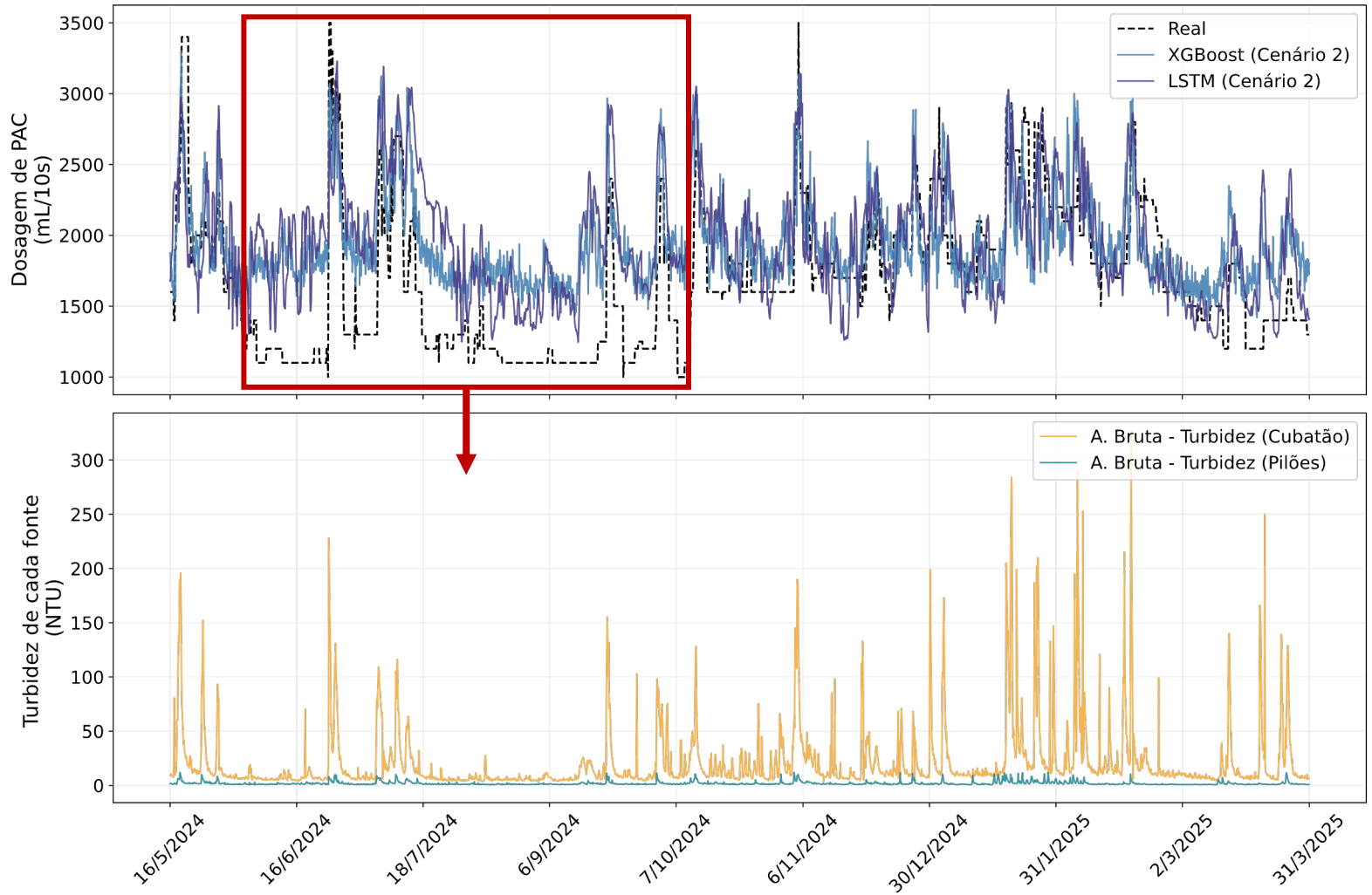
Resultados e discussão

Cenário 2: comparativo entre previsão e perfil de turbidez na água bruta



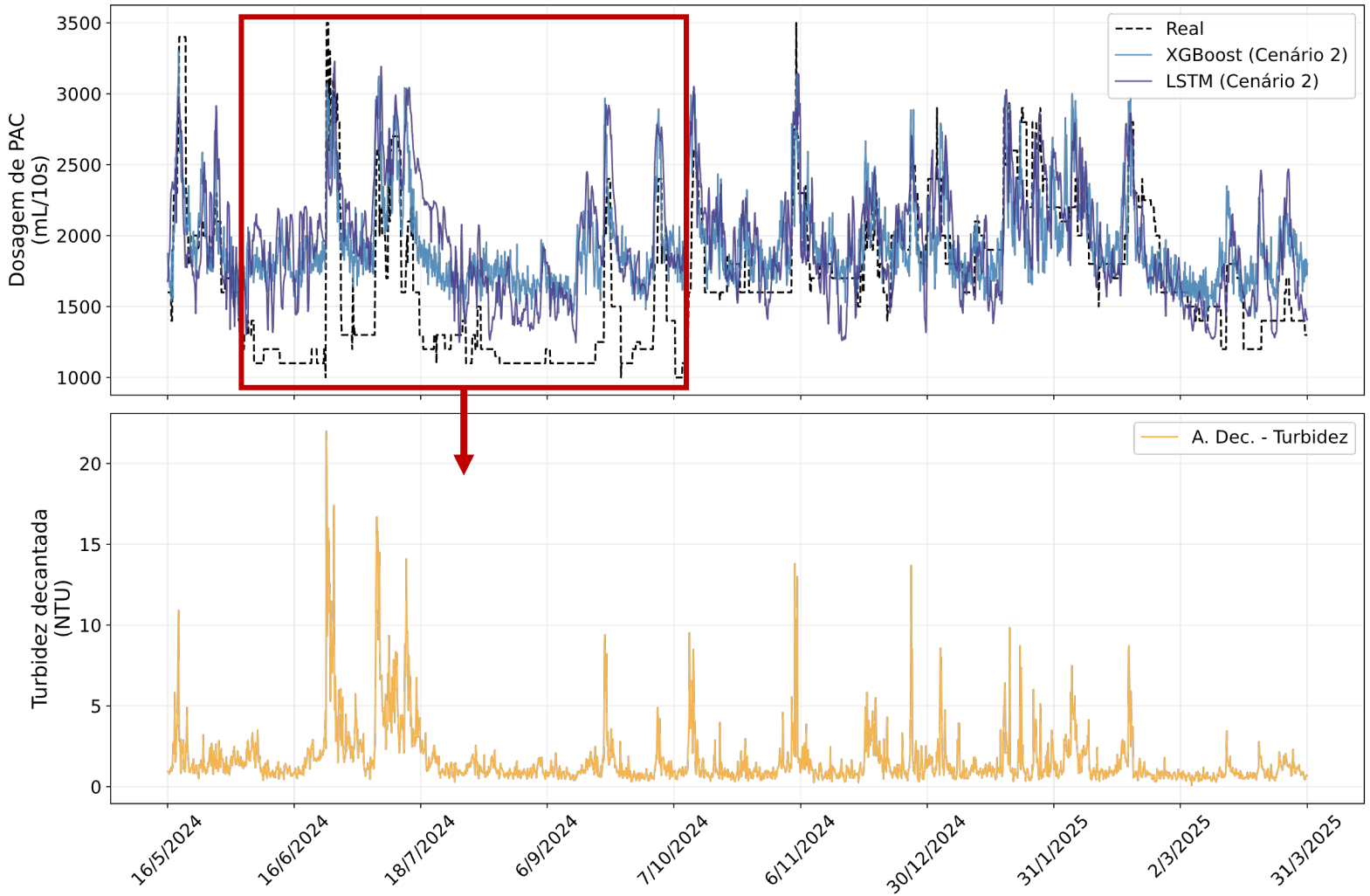
Resultados e discussão

Cenário 2: comparativo entre previsão e perfil de turbidez na água bruta



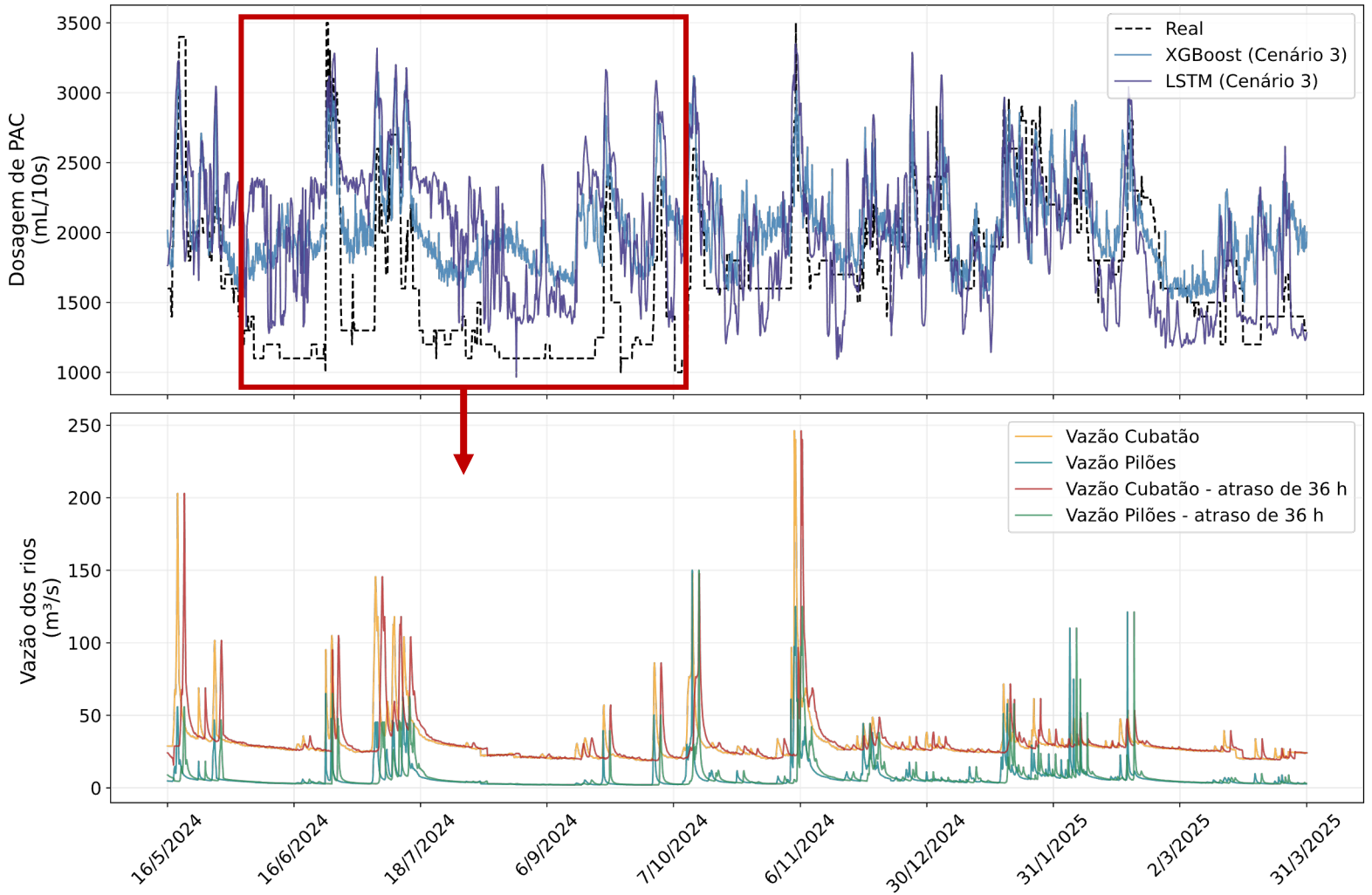
Resultados e discussão

Cenário 2: comparativo entre previsão e perfil de turbidez na água decantada



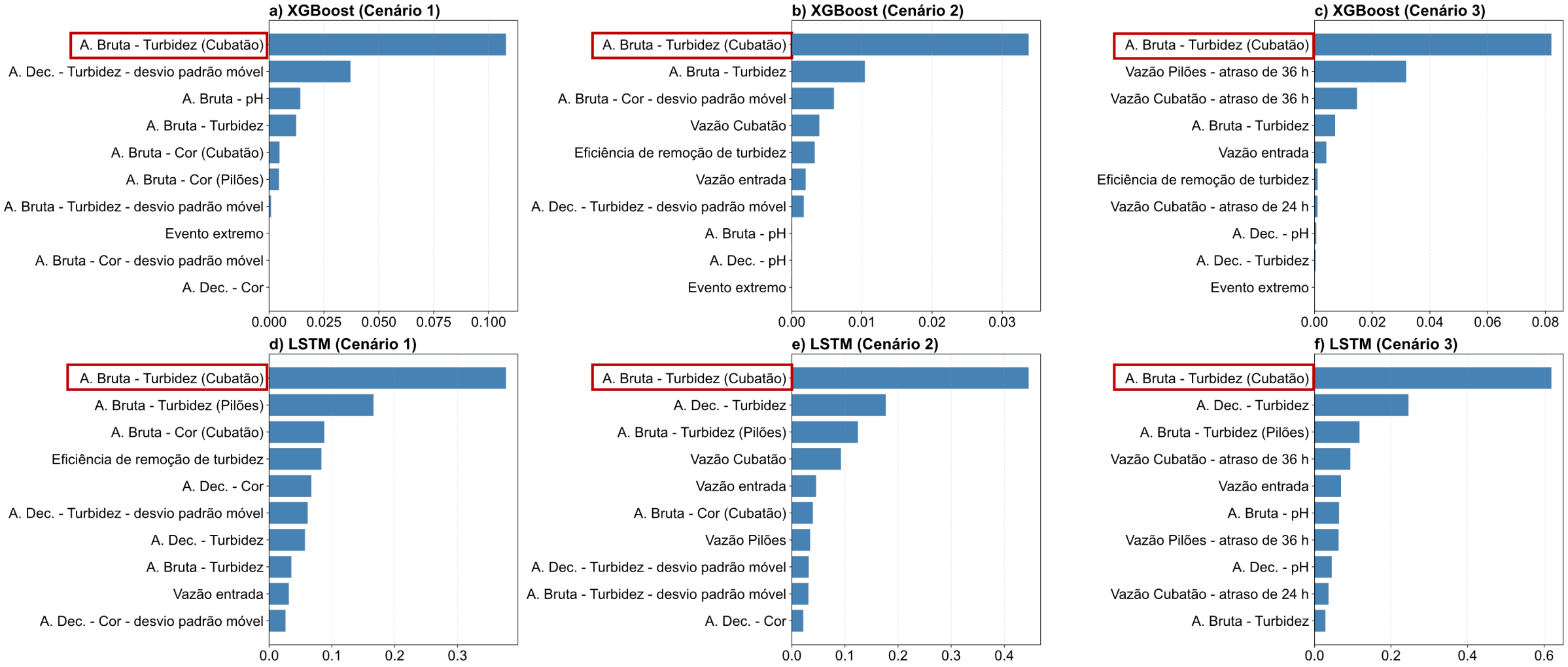
Resultados e discussão

Cenário 3: comparativo entre previsão e perfil de vazão dos rios



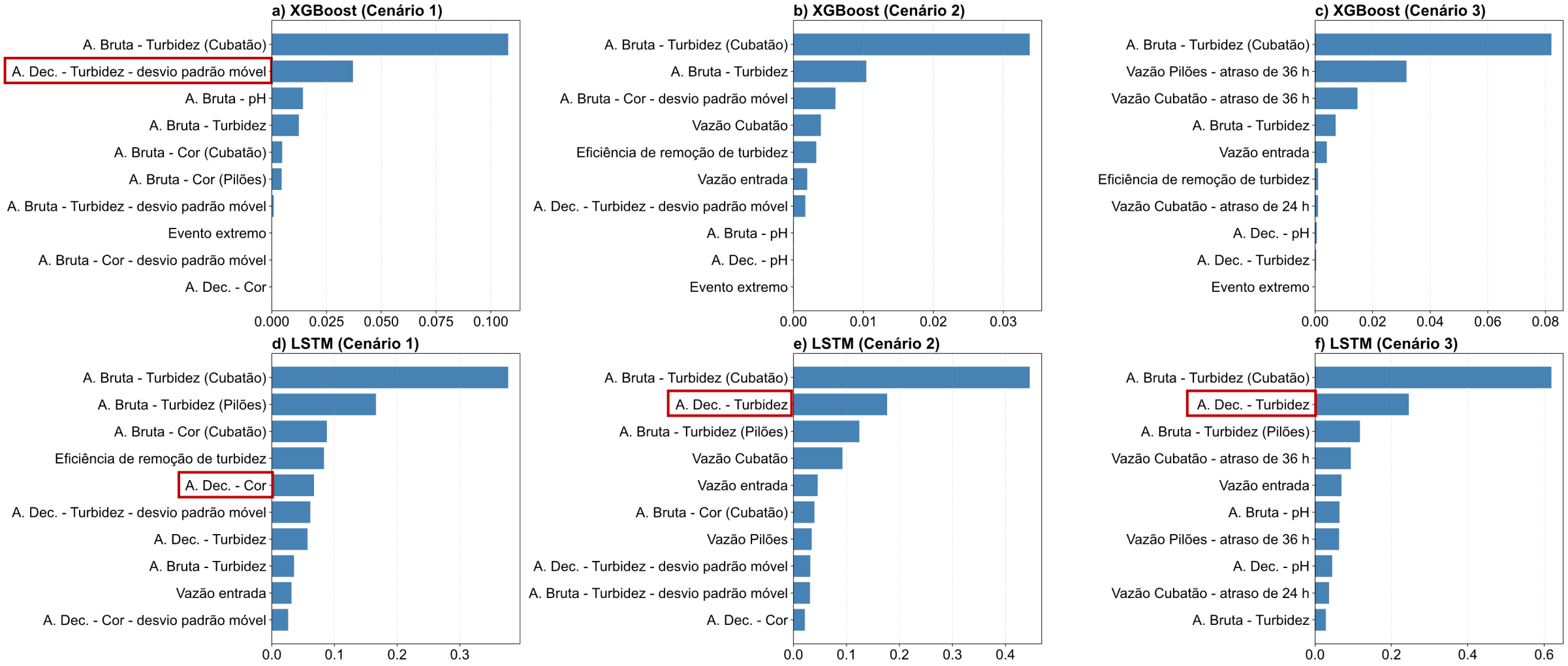
Resultados e discussão

Análise de importância das variáveis



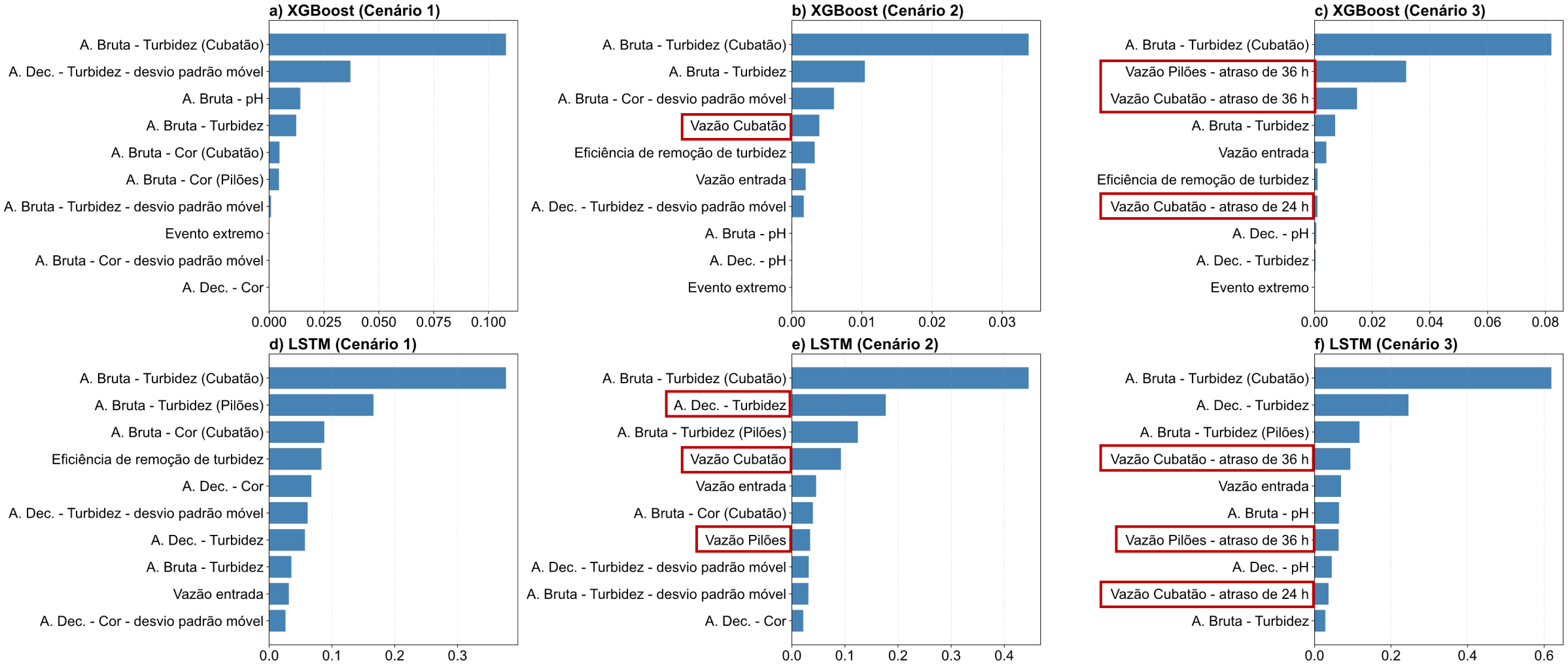
Resultados e discussão

Análise de importância das variáveis



Resultados e discussão

Análise de importância das variáveis



Conclusões

- Os resultados obtidos demonstraram uma queda na acurácia dos modelos quando comparados a cenários de previsão baseados exclusivamente em informações de qualidade da água;
- Contudo, há uma alta complexidade envolvida no padrão de causalidade do sistema;
- Além disso, a própria dinâmica de decisão de controle da dosagem realizada pelos operadores da estação também contribuiu para os resultados.

Conclusões

A relação entre as vazões dos rios e a qualidade da água de captação é multifatorial:

- qualidade do histórico de dados de monitoramento;
- características da bacia hidrográfica;
- influência de atividades antrópicas na região;
- demais fatores que influenciam os padrões hidrológicos observados.

Conclusões

Encontro Técnico AESABESP
37º Congresso Nacional de Saneamento e Meio Ambiente

Ademais, estabelecer uma relação entre essas variáveis com a dosagem de coagulante requer rigor no padrão técnico para decisão acerca do controle da dosagem, preferencialmente com embasamento experimental para determinar os requerimentos reais de coagulante no processo de forma dinâmica.

Obrigada!

Juliana Neves

nevesjulianajn@gmail.com

(48) 98811-9802

Encontro Técnico AESABESP
37º Congresso Nacional de Saneamento e Meio Ambiente

labMAC+
Laboratório de Materiais e Computação Científica
Laboratory of Materials and Scientific Computing

AESABESP 40 Anos
Associação dos Engenheiros da Sabesp

HydroInfo Smart Solutions

CNPq

UFSC

casan

Coautores

Rodrigo Campos de Andrade
rca@hydroinfo.com.br

Natan Padoin
natan.padoin@ufsc.br

Cíntia Soares
cintia.soares@ufsc.br