

UNIVERSIDADE FEDERAL DO CARIRI
CENTRO DE CIÊNCIAS AGRÁRIAS E DA BIODIVERSIDADE
PROGRAMA DE PÓS-GRADUAÇÃO EM DESENVOLVIMENTO
REGIONAL SUSTENTÁVEL

PREVISÃO DO NÍVEL DE ÁGUA SUBTERRÂNEA USANDO
ENSEMBLE DE MODELOS DE SÉRIES TEMPORAIS

RUBENS OLIVEIRA DA CUNHA JÚNIOR

CRATO - CE

2022

RUBENS OLIVEIRA DA CUNHA JÚNIOR

PREVISÃO DO NÍVEL DE ÁGUA SUBTERRÂNEA USANDO *ENSEMBLE* DE
MODELOS DE SÉRIES TEMPORAIS

Dissertação apresentada ao Programa de Pós-Graduação em Desenvolvimento Regional Sustentável, Centro de Ciências Agrárias e da Biodiversidade, Universidade Federal do Cariri, como requisito parcial à obtenção do título de mestre em Desenvolvimento Regional Sustentável.

Orientador: Prof. Dr. Paulo Renato
Alves Firmino

Coorientadora: Profa. Dra. Celme
Torres Ferreira da Costa

Dados Internacionais de Catalogação na Publicação.
Universidade Federal do Cariri.
Sistema de Bibliotecas

- C972p Cunha Júnior, Rubens Oliveira da.
Previsão do nível de água subterrânea usando *ensemble* de modelos de séries temporais /
Rubens Oliveira da Cunha Júnior. – 2022.
150 f.: il. color.30 cm
- Dissertação (mestrado) - Universidade Federal do Cariri, Mestrado em Desenvolvimento
Regional Sustentável (PRODER), Crato, 2022.
- Orientação: Prof. Dr. Paulo Renato Alves Firmino.
Coorientação: Profa. Dra. Celme Torres Ferreira da Costa.
1. Combinação de preditores. 2. Aprendizado de máquina. 3. Hidrogeologia. 4.
Desenvolvimento sustentável. 5. Semiárido. I. Título.

CDD 551.49

Bibliotecária: Glacínésia Leal Mendonça
CRB 3/ 925

RUBENS OLIVEIRA DA CUNHA JÚNIOR

PREVISÃO DO NÍVEL DE ÁGUA SUBTERRÂNEA USANDO *ENSEMBLE* DE
MODELOS DE SÉRIES TEMPORAIS

Dissertação apresentada ao Programa de Pós-Graduação em Desenvolvimento Regional Sustentável, Centro de Ciências Agrárias e da Biodiversidade, Universidade Federal do Cariri, como requisito parcial à obtenção do título de mestre em Desenvolvimento Regional Sustentável.

Aprovada em: 11 de Julho de 2022

BANCA EXAMINADORA



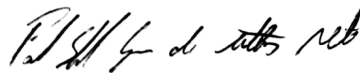
Prof. Dr. Paulo Renato Alves
Firmino (Orientador)
Universidade Federal do Cariri (UFCA)



Profa. Dra. Celme Torres Ferreira da
Costa (Coorientadora)
Universidade Federal do Cariri (UFCA)



Prof. Dr. Carlos Wagner Oliveira
Universidade Federal do Cariri (UFCA)



Prof. Dr. Paulo Salgado Gomes de Mattos Neto
Universidade Federal de Pernambuco (UFPE)



Prof. Dr. Marco Aurélio Holanda de Castro
Universidade Federal do Ceará (UFC)

AGRADECIMENTOS

Aos meus pais e irmão, pelo apoio e incentivo ao longo de toda a minha vida.

À minha namorada, pela compreensão, paciência e incentivo sempre.

Ao professor Paulo Renato Alves Firmino, pelos ensinamentos e por ter sido meu orientador, desempenhando tal função com dedicação e parceria.

À professora Celme Torres Ferreira da Costa, pela coorientação e importantes contribuições para o meu projeto.

Aos professores Carlos Wagner Oliveira, Paulo Salgado Gomes de Mattos Neto da Universidade Federal de Pernambuco (UFPE) e Marco Aurélio Holanda de Castro da Universidade Federal do Ceará (UFC), pelas considerações que muito contribuíram para a realização deste trabalho.

A todos os professores do Programa de Pós-Graduação em Desenvolvimento Regional Sustentável (PRODER), pelos ensinamentos.

Aos colegas de turma do PRODER que tive a oportunidade de conhecer e trabalhar em conjunto.

Aos colegas do grupo de pesquisa em Modelagem Estatística, Simulação e Otimização de Risco (MESOR), pelo compartilhamento de experiências e conhecimento durante os encontros síncronos semanais.

A todos que convivi ao longo desses anos de curso e que participaram, direta ou indiretamente, do desenvolvimento deste trabalho de pesquisa, contribuindo para o meu processo de aprendizado.

E ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), pelo período em que a pesquisa de mestrado foi financiada via bolsa de estudos.

RESUMO

A modelagem e a previsão dos níveis de água subterrânea são fundamentais para a avaliação da disponibilidade hídrica. Portanto, formalismos de séries temporais são uma alternativa robusta para a inferência sobre o estado deste importante recurso. Entre as estratégias de modelagem e previsão de séries temporais, as técnicas de combinação de preditores, também conhecidas como *ensemble*, podem melhorar o desempenho dos modelos. Contudo, este ainda é um tema pouco explorado em estudos sobre água subterrânea. Diante disso, o objetivo deste trabalho é desenvolver modelos individuais e de *ensemble* para previsão de séries temporais do nível de água subterrânea. Adotou-se como caso de estudo a bacia Sedimentar do Araripe, no Estado do Ceará, Brasil. Entre as estratégias individuais, foram usados modelos lineares tradicionais, como os modelos Autorregressivos Integrados de Médias Móveis (ARIMA) e os modelos de Suavização Exponencial (ETS), além de técnicas baseadas em Inteligência Artificial e Aprendizado de Máquina, tais como Redes Neurais Artificiais clássicas (ANN), Máquinas de Aprendizado Extremo (ELM), redes *Long Short-Term Memory* (LSTM) e *Support Vector Regression* (SVR). Para a combinação, optou-se por técnicas lineares tradicionais, como Média Simples (cSA), Mediana Simples (cSM) e Mínima Variância (cMV), além de alternativas não lineares, baseadas em Cópulas (cCP) e ANN (cANN) e SVR (cSVR). O desempenho dos modelos foi comparado e classificado segundo métricas de erro. A maioria das séries estudadas apresentou diminuição expressiva dos níveis de água subterrânea. Contudo, os modelos foram capazes de representar e prever o comportamento da água subterrânea nos aquíferos da bacia do Araripe de forma satisfatória. Os modelos ARIMA e LSTM foram superiores, mas os modelos combinados mostraram resultados atraentes, superando a maioria das alternativas individuais. As abordagens combinadas propostas podem ser aprimoradas.

Palavras-chave: Combinação de preditores. Aprendizado de máquina. Hidrogeologia. Desenvolvimento sustentável. Semiárido.

ABSTRACT

Modeling and forecasting groundwater levels are essential for water availability evaluation. Therefore, time series formalisms are a robust alternative to make inferences about this relevant resource. Among the time series modeling and forecasting strategies, the techniques for the combination of forecasts, also known as an ensemble, can improve the performance of the models. However, this is still a little explored subject in groundwater studies. This work aims to develop single and ensemble models for forecasting groundwater levels time series. The Araripe sedimentary basin in the Ceará state, Brazil, was adopted as a case study. Among the single strategies, traditional linear models were used, such as Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing (ETS), besides Artificial Intelligence and Machine Learning-based techniques, such as classic Artificial Neural Networks (ANN), Extreme Learning Machines (ELM), Long Short-Term Memory (LSTM), and Support Vector Regression (SVR). For the combination, traditional linear techniques were selected, such as Simple Average (cSA), Simple Median (cSM), and Minimal Variance (cMV), in addition to non-linear alternatives, based on Cópulas (cCP), ANN (cANN), and SVR (cSVR). Performance evaluation of the models was compared and classified according to error measures. Most of the studied time series showed a sharp drawdown in the groundwater levels. However, the models were able to satisfactorily model and predict the groundwater behavior in the Araripe basin's aquifers. The ARIMA and LSTM single models achieved the best performance, but the combined models showed interesting results, outperforming most of the single alternatives. The proposed combined approaches can be improved.

Keywords: Combination of forecasts. Machine learning. Hidrogeology. Sustainable development. Semiarid.

LISTA DE ILUSTRAÇÕES

Figura 1 – Representação esquemática da distribuição vertical da água no solo e subsolo, destacando diversas zonas de umidade em um aquífero livre . . .	28
Figura 2 – Representação esquemática de a. poço de monitoramento com a medida do nível estático (NE); e b. instalação da sonda submersa medidora de nível do tipo <i>levelogger</i> em um poço de monitoramento	31
Figura 3 – Arquitetura básica de uma rede MLP para previsão de séries temporais univariadas com uma camada oculta	37
Figura 4 – Aplicação de SVR como um modelo autorregressivo de ordem 1 para previsão um passo adiante de séries temporais	42
Figura 5 – Localização da porção leste da bacia Sedimentar do Araripe, com destaque para os aquíferos e poços selecionados	53
Figura 6 – Mapa geológico e estratigrafia da bacia Sedimentar do Araripe	54
Figura 7 – Resumo do percurso metodológico adotado. (i) Pré-processamento, (ii) Modelagem individual, (iii) Modelagem combinada e (iv) Avaliação do desempenho	57
Figura 8 – Gráficos das séries mensais de NAS na BSA, com destaque para os valores observados (cor preta) e imputados (cor vermelha). Diferentes períodos indicados nos gráficos	65
Figura 9 – Correlograma das séries mensais de NAS na BSA	68
Figura 10 – Gráfico de dispersão dos coeficientes de correlação linear de Pearson R entre pares de séries mensais de NAS e as distâncias (em km) entre os poços na BSA	68
Figura 11 – Parâmetros dos modelos baseados em ML considerando todas as séries. a. termos autorregressivos não sazonais (p), b. termos autorregressivos sazonais (P), c. número de nós nas camadas ocultas (nN), d. número de camadas ocultas (nL), e. f. g. parâmetros γ , C e ε da SVR, respectivamente, h. função de ativação (FA)	71
Figura 12 – Distribuição das medidas de desempenho dos modelos individuais (cor laranja) e combinados (cor azul) no conjunto de testes considerando todas as séries	76

Figura B.1 – Gráficos das Funções de Autocorrelação (FAC) e Autocorrelação Parcial (FACP) até o lag 36 das séries mensais de NAS (séries indicadas nos gráficos)	98
Figura B.1 – (Continuação) Gráficos das Funções de Autocorrelação (FAC) e Autocorrelação Parcial (FACP) até o lag 36 das séries mensais de NAS (séries indicadas nos gráficos)	99
Figura B.1 – (Continuação) Gráficos das Funções de Autocorrelação (FAC) e Autocorrelação Parcial (FACP) até o lag 36 das séries mensais de NAS (séries indicadas nos gráficos)	100
Figura D.1 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P01	115
Figura D.2 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P02	116
Figura D.3 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P03	117
Figura D.4 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P04	118
Figura D.5 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P05	119
Figura D.6 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P06	120
Figura D.7 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P07	121
Figura D.8 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P08	122
Figura D.9 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P09	123
Figura D.10 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P10	124
Figura D.11 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P11	125

Figura D.12–Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P12	126
Figura D.13–Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P13	127
Figura D.14–Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P14	128
Figura D.15–Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P15	129
Figura D.16–Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P16	130
Figura D.17–Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P17	131
Figura D.18–Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P18	132
Figura D.19–Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P19	133
Figura D.20–Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P20	134
Figura D.21–Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P21	135

LISTA DE TABELAS

Tabela 1 – Parâmetros otimizados pelo algoritmo genético e os seus possíveis valores	60
Tabela 2 – Medidas descritivas das séries mensais de NAS na BSA	66
Tabela 3 – Etapas de modelagem e previsão para as séries temporais (ST)	70
Tabela 4 – Tempos de modelagem (em segundos) dos modelos individuais e combinados para as séries temporais (ST). Os melhores resultados em cada caso estão destacados em negrito	74
Tabela 5 – Classificação dos modelos para as séries temporais (ST) de NAS na BSA. Os melhores resultados estão destacados em negrito	77
Tabela A.1 – Informações gerais e geológicas dos poços na BSA	95
Tabela A.2 – Funções de ativação disponíveis por padrão nos pacotes <code>neuralnet</code> e <code>elmNNRcpp</code> para ANN e ELM, respectivamente	96
Tabela B.1 – Matriz de distâncias (em km) entre os poços da BSA	101
Tabela C.1 – Descrição dos modelos construídos para a série P01	103
Tabela C.2 – Descrição dos modelos construídos para a série P02	103
Tabela C.3 – Descrição dos modelos construídos para a série P03	104
Tabela C.4 – Descrição dos modelos construídos para a série P04	104
Tabela C.5 – Descrição dos modelos construídos para a série P05	105
Tabela C.6 – Descrição dos modelos construídos para a série P06	105
Tabela C.7 – Descrição dos modelos construídos para a série P07	106
Tabela C.8 – Descrição dos modelos construídos para a série P08	106
Tabela C.9 – Descrição dos modelos construídos para a série P09	107
Tabela C.10 – Descrição dos modelos construídos para a série P10	107
Tabela C.11 – Descrição dos modelos construídos para a série P11	108
Tabela C.12 – Descrição dos modelos construídos para a série P12	108
Tabela C.13 – Descrição dos modelos construídos para a série P13	109
Tabela C.14 – Descrição dos modelos construídos para a série P14	109
Tabela C.15 – Descrição dos modelos construídos para a série P15	110
Tabela C.16 – Descrição dos modelos construídos para a série P16	110
Tabela C.17 – Descrição dos modelos construídos para a série P17	111
Tabela C.18 – Descrição dos modelos construídos para a série P18	111
Tabela C.19 – Descrição dos modelos construídos para a série P19	112

Tabela C.20–Descrição dos modelos construídos para a série P20	112
Tabela C.21–Descrição dos modelos construídos para a série P21	113
Tabela E.1 – Desempenho dos modelos de previsão para a série P01 no conjunto de teste (mar/2020 – fev/2021)	137
Tabela E.2 – Desempenho dos modelos de previsão para a série P02 no conjunto de teste (mar/2020 – fev/2021)	137
Tabela E.3 – Desempenho dos modelos de previsão para a série P03 no conjunto de teste (fev/2020 – jan/2021)	138
Tabela E.4 – Desempenho dos modelos de previsão para a série P04 no conjunto de teste (mar/2020 – fev/2021)	138
Tabela E.5 – Desempenho dos modelos de previsão para a série P05 no conjunto de teste (mar/2020 – fev/2021)	139
Tabela E.6 – Desempenho dos modelos de previsão para a série P06 no conjunto de teste (nov/2020 – out/2021)	139
Tabela E.7 – Desempenho dos modelos de previsão para a série P07 no conjunto de teste (dez/2020 – nov/2021)	140
Tabela E.8 – Desempenho dos modelos de previsão para a série P08 no conjunto de teste (mar/2020 – fev/2021)	140
Tabela E.9 – Desempenho dos modelos de previsão para a série P09 no conjunto de teste (nov/2020 – out/2021)	141
Tabela E.10–Desempenho dos modelos de previsão para a série P10 no conjunto de teste (nov/2020 – out/2021)	141
Tabela E.11–Desempenho dos modelos de previsão para a série P11 no conjunto de teste (mar/2020 – fev/2021)	142
Tabela E.12–Desempenho dos modelos de previsão para a série P12 no conjunto de teste (mar/2020 – fev/2021)	142
Tabela E.13–Desempenho dos modelos de previsão para a série P13 no conjunto de teste (mar/2020 – fev/2021)	143
Tabela E.14–Desempenho dos modelos de previsão para a série P14 no conjunto de teste (mar/2020 – fev/2021)	143
Tabela E.15–Desempenho dos modelos de previsão para a série P15 no conjunto de teste (mar/2020 – fev/2021)	144

Tabela E.16–Desempenho dos modelos de previsão para a série P16 no conjunto de teste (mar/2020 – fev/2021)	144
Tabela E.17–Desempenho dos modelos de previsão para a série P17 no conjunto de teste (mar/2020 – fev/2021)	145
Tabela E.18–Desempenho dos modelos de previsão para a série P18 no conjunto de teste (set/2020 – ago/2021)	145
Tabela E.19–Desempenho dos modelos de previsão para a série P19 no conjunto de teste (mar/2020 – fev/2021)	146
Tabela E.20–Desempenho dos modelos de previsão para a série P20 no conjunto de teste (mar/2020 – fev/2021)	146
Tabela E.21–Desempenho dos modelos de previsão para a série P21 no conjunto de teste (mar/2020 – fev/2021)	147
Tabela E.22–Classificação dos modelos por medida de desempenho (MD): série P01	148
Tabela E.23–Classificação dos modelos por medida de desempenho (MD): série P02	148
Tabela E.24–Classificação dos modelos por medida de desempenho (MD): série P03	148
Tabela E.25–Classificação dos modelos por medida de desempenho (MD): série P04	148
Tabela E.26–Classificação dos modelos por medida de desempenho (MD): série P05	148
Tabela E.27–Classificação dos modelos por medida de desempenho (MD): série P06	149
Tabela E.28–Classificação dos modelos por medida de desempenho (MD): série P07	149
Tabela E.29–Classificação dos modelos por medida de desempenho (MD): série P08	149
Tabela E.30–Classificação dos modelos por medida de desempenho (MD): série P09	149
Tabela E.31–Classificação dos modelos por medida de desempenho (MD): série P10	149
Tabela E.32–Classificação dos modelos por medida de desempenho (MD): série P11	150
Tabela E.33–Classificação dos modelos por medida de desempenho (MD): série P12	150
Tabela E.34–Classificação dos modelos por medida de desempenho (MD): série P13	150
Tabela E.35–Classificação dos modelos por medida de desempenho (MD): série P14	150
Tabela E.36–Classificação dos modelos por medida de desempenho (MD): série P15	150
Tabela E.37–Classificação dos modelos por medida de desempenho (MD): série P16	151
Tabela E.38–Classificação dos modelos por medida de desempenho (MD): série P17	151
Tabela E.39–Classificação dos modelos por medida de desempenho (MD): série P18	151
Tabela E.40–Classificação dos modelos por medida de desempenho (MD): série P19	151
Tabela E.41–Classificação dos modelos por medida de desempenho (MD): série P20	151

LISTA DE ABREVIATURAS E SIGLAS

AI	inteligência artificial
AIC	Critério de Informação de Akaike
AICc	Critério de Informação de Akaike corrigido
ANN	redes neurais artificiais
AR	autorregressivo
ARIMA	autorregressivo integrado de médias móveis
ARMA	autorregressivo de médias móveis
BIC	Critério de Informação Bayesiano
BSA	bacia Sedimentar do Araripe
cANN	combinação via ANN
cCP	combinação via cópulas
cMV	mínima variância
CPRM	Serviço Geológico do Brasil
cSA	média simples
cSM	mediana simples
cSVR	combinação via SVR
DNOCS	Departamento Nacional de Obras Contra a Seca
ELM	máquinas de aprendizado extremo
ETS	suavização exponencial
EWMA	média móvel exponencial
FA	função de ativação
FAC	função de autocorrelação
FACP	função de autocorrelação parcial
GA	algoritmos genéticos
IFM	Inferência por Marginais
LSTM	<i>long short-term memory</i>
LWMA	média móvel linear
MA	médias móveis
MAE	Erro Absoluto Médio
MAPE	Erro Percentual Absoluto Médio
MD	medidas de desempenho

ML	aprendizado de máquina
MLP	perceptron multicamadas
MSE	Erro Quadrático Médio
NA	valores ausentes
NAS	níveis de água subterrânea
ND	nível dinâmico
NE	nível estático
POCID	<i>Prediction Of Change In Direction</i>
RBF	<i>Radial Basis Function</i>
RIMAS	Rede Integrada de Monitoramento de Águas Subterrâneas
RMC	Região Metropolitana do Cariri
RMSE	Raiz do Erro Quadrático Médio
RNN	redes neurais recorrentes
SES	suavização exponencial simples
SLFN	redes neurais <i>feed-forward</i> com uma camada oculta
SMA	média móvel simples
STL	decomposição sazonal e de tendência usando <i>Loess</i>
SUDENE	Secretaria de Desenvolvimento do Nordeste
SVM	máquinas de vetores de suporte
SVR	regressão por vetores de suporte

LISTA DE SÍMBOLOS

p	Ordem da parcela autorregressiva dos modelos ARIMA
q	Ordem da parcela de médias móveis dos modelos ARIMA
d	Número de diferenciações dos modelos ARIMA
P	Ordem da parcela autorregressiva sazonal dos modelos ARIMA
Q	Ordem da parcela de médias móveis sazonal dos modelos ARIMA
D	Número de diferenciações sazonais dos modelos ARIMA
m	Sazonalidade da série temporal
y_t	Valor da série temporal no instante de tempo t
\mathbf{Y}	Representação matricial da série temporal y
\mathbf{Y}^*	Matriz de valores defasados de y
\mathbf{Y}_{comp}^*	Subconjunto com registros completos de \mathbf{Y}^*
\mathbf{Y}_{incomp}^*	Subconjunto com registros incompletos de \mathbf{Y}^*
c	Constante da equação dos modelos ARIMA
v	Viés (<i>bias</i>) da rede neural
ε_t	Ruído branco
ϕ	Coefficiente do modelo autorregressivo
θ	Coefficiente do modelo de médias móveis
$\dot{\mathbf{x}}$	Vetor de estados do modelo ETS
l	Componente de <i>level</i> do modelo ETS
b	Componente de tendência do modelo ETS
s	Componente de sazonalidade do modelo ETS
F, g, ψ	Coefficientes do modelo ETS
i	Índice referente à camada de entrada da rede neural
h	Índice referente à camada oculta da rede neural
o	Índice referente à camada de saída da rede neural
n_i	Número de nós na camada de entrada da rede neural

n_h	Número de nós na camada oculta da rede neural
\mathbf{w}	Vetor de pesos
\mathbf{W}	Matriz de pesos
f	Função de ativação
\dot{h}_t	Memória das redes neurais recorrentes RNN e LSTM no instante de tempo t
FG	<i>Forget gate</i> da rede neural LSTM
IG	<i>Input gate</i> da rede neural LSTM
OG	<i>Output gate</i> da rede neural LSTM
CG	Valores candidatos da rede neural LSTM
L	Número de nós na camada oculta da rede neural ELM
\mathbf{x}_i	Vetor de variáveis regressoras
k	Número de variáveis regressoras ou atributos
σ	Função de ativação sigmoide ou logística
\tanh	Função de ativação tangente hiperbólica
\mathbf{H}	Matriz de saída da rede neural ELM
\mathbf{H}^+	Matriz inversa generalizada de Moore-Penrose de \mathbf{H}
$\hat{\eta}$	Solução para os pesos da camada de saída da rede neural ELM
C	Constante de compensação da regressão por vetores de suporte SVR
ξ_i, ξ_i^*	Variáveis de folga da SVR
α_i, α_i^*	Multiplicadores de Lagrange
ϵ	Tolerância de erro da SVR
$\mathbf{K}(\cdot)$	Função <i>kernel</i>
Φ	Transformação
\mathbb{R}^n	Espaço de dimensão n
γ	Parâmetro da função <i>kernel radial basis</i>
N	Número de modelos individuais considerados na combinação
\hat{y}_t	Resultado da combinação de preditores no instante de tempo t

φ	Função de combinação de preditores
ω	Peso atribuído ao preditor na combinação
M^{-1}	Matriz de covariâncias inversa dos resíduos
a_{lj}	Elemento j da linha l da matriz M^{-1}
\mathbf{X}_t	Vetor de preditores individuais para y_t
\mathbf{E}_t	Vetor de erros dos preditores individuais
\mathbf{e}_t	Instâncias de \mathbf{E}_t
$F_{\mathbf{E}_t,i}$	Distribuição acumulada conjunta de \mathbf{E}_t
$F_{E_{t,i}}(e_{t,i})$	Distribuição acumulada marginal
$p_{E_{t,i}}(e_{t,i})$	Função densidade de probabilidade
$C(\cdot)$	Função cópula
u_i	Distribuição acumulada marginal dos erros do modelo i
α_i	Parâmetro da distribuição acumulada marginal i
p_n	Polinômio de grau n
$g(\cdot)$	Função
$\hat{g}(\cdot)$	Função interpoladora de g
M	Janela de média móvel
S_t	Componente de sazonalidade da decomposição STL
T_t	Componente de tendência da decomposição STL
R_t	Componente de resíduo da decomposição STL
nN	Número de nós na camada oculta de uma rede neural
nL	Número de camadas ocultas de uma rede neural
R^2	Coefficiente de determinação
R	Coefficiente de correlação de Pearson
ρ	Coefficiente de correlação de postos de Spearman
K	Número de parâmetros do modelo
L^*	Máximo da função de verossimilhança do modelo
ln	Logaritmo natural

SUMÁRIO

1	INTRODUÇÃO	22
1.1	Objetivos	24
1.1.1	Objetivo geral	24
1.1.2	Objetivos específicos	24
1.2	Estrutura do trabalho	24
2	REFERENCIAL TEÓRICO	25
2.1	Desenvolvimento sustentável e recursos hídricos	25
2.2	Recursos hídricos para o semiárido	25
2.3	Água subterrânea	27
2.3.1	Sistemas aquíferos	27
2.3.2	Recarga da água subterrânea	29
2.3.3	Monitoramento da água subterrânea	30
2.4	Séries temporais	32
2.4.1	ARIMA	33
2.4.2	ETS	35
2.4.3	Redes Neurais Artificiais	36
2.4.3.1	<i>Multilayer Perceptron</i>	36
2.4.3.2	<i>Extreme Learning Machine</i>	37
2.4.3.3	<i>Redes Neurais Recorrentes</i>	38
2.4.3.4	<i>Long Short-Term Memory</i>	39
2.4.4	Support Vector Regression	40
2.4.5	Modelos combinados	42
2.4.5.1	<i>Média Simples e Mediana Simples</i>	43
2.4.5.2	<i>Mínima Variância</i>	43
2.4.5.3	<i>Cópulas</i>	43
2.4.5.4	<i>Combinação baseada em aprendizado de máquina</i>	45
2.5	Algoritmos genéticos	45
2.6	Análise de dados ausentes	47
2.6.1	Imputação de dados ausentes em séries temporais	48
2.6.1.1	<i>Interpolação</i>	49

2.6.1.2	<i>Médias móveis</i>	49
2.6.1.3	<i>Imputação com decomposição</i>	50
2.6.1.4	<i>Regressão em valores defasados</i>	51
3	METODOLOGIA	53
3.1	Área de estudo	53
3.2	Conjunto de dados	55
3.3	Arquitetura do estudo	55
3.4	Pré-processamento dos dados	56
3.5	Modelagem individual	58
3.6	Modelagem combinada	59
3.7	Otimização dos parâmetros	60
3.8	Previsão	60
3.9	Avaliação do desempenho	61
3.10	Implementação computacional	63
4	RESULTADOS E DISCUSSÃO	64
4.1	Resultados gerais	64
4.2	Etapas de modelagem	69
4.3	Etapas de previsão	75
5	CONSIDERAÇÕES FINAIS	80
5.1	Limitações	80
5.2	Trabalhos futuros	80
	REFERÊNCIAS	82
	APÊNDICES	94
	APÊNDICE A – Informações suplementares	94
	APÊNDICE B – Análise exploratória dos dados	97
	APÊNDICE C – Descrição dos modelos	102
	APÊNDICE D – Gráficos das previsões	114
	APÊNDICE E – Avaliação do desempenho dos modelos	136

1 INTRODUÇÃO

A água subterrânea é uma das principais fontes de água para diversos usos em atividades industriais, agrícolas e abastecimento humano, sobretudo em regiões semiáridas e áridas, onde as águas superficiais são escassas e má distribuídas (SILVA *et al.*, 2021). Por conta disso, o uso racional deste importante recurso é fundamental para o desenvolvimento sustentável. No entanto, os recursos hídricos subterrâneos estão expostos a graves problemas de qualidade e quantidade em diversas regiões do mundo (ABDELKARIM *et al.*, 2022). Fatores relacionados ao crescimento econômico, ocupação desordenada de espaços urbanos, crescimento populacional, aumento da demanda por alimentos, expansão de áreas agrícolas e mudanças climáticas contribuem para o problema (ADIMALLA; WU, 2019; SAHU *et al.*, 2020). Neste contexto, a gestão sustentável da água subterrânea requer meios para a avaliação da disponibilidade hídrica e para a previsão de cenários futuros (AHMADI *et al.*, 2022). Portanto, a estimativa dos níveis de água subterrânea (NAS) é essencial para a tomada de decisão na gestão de recursos hídricos (WU *et al.*, 2021).

A modelagem de água subterrânea tem sido tradicionalmente realizada por meio de modelos físicos ou conceituais (ENEMARK *et al.*, 2019). Contudo, tais técnicas requerem um profundo conhecimento hidrogeológico dos sistemas aquíferos. Diante disso, a análise de séries temporais constitui uma alternativa confiável para a obtenção de estimativas dos níveis freáticos (RAJAEI *et al.*, 2019). Para a modelagem e previsão de séries temporais de NAS, pesquisadores têm empregado formalismos estatísticos, tais como o modelo autorregressivo integrado de médias móveis (ARIMA) e os modelos de suavização exponencial (ETS) (SAKIZADEH *et al.*, 2019). Outros autores também têm recorrido a modelos não lineares, tais como técnicas baseadas em aprendizado de máquina (ML) (TAO *et al.*, 2022). Entre elas, destacam-se a regressão por vetores de suporte (SVR) (BOZORG-HADDAD *et al.*, 2019; MOZAFFARI *et al.*, 2022) e as redes neurais artificiais (ANN) (LEE *et al.*, 2019; MALIK; BHAGWAT, 2021; KIM; LEE, 2022), como as máquinas de aprendizado extremo (ELM) (MALEKZADEH *et al.*, 2019; WU *et al.*, 2021; AZIZPOUR *et al.*, 2021) e as redes recorrentes *long short-term memory* (LSTM) (WUNSCH *et al.*, 2021; SOLGI *et al.*, 2021).

A escolha da abordagem mais apropriada para a modelagem e previsão de séries temporais de NAS tem sido discutida por Zanotti *et al.* (2019). No geral, alguns

autores destacam que ambos modelos, lineares ou não lineares, apresentam vantagens e desvantagens (DE OLIVEIRA *et al.*, 2020). Diante disso, estratégias de modelagem combinada têm sido propostas com o objetivo de melhorar o desempenho dos modelos de previsão de séries temporais (HAJIRAHIMI; KHASHEI, 2019). Uma alternativa consiste em combinar modelos individuais de uma mesma série, em uma técnica conhecida como *ensemble*. Existem também estratégias híbridas, em que os resíduos da previsão de um modelo são usados sequencialmente como entradas em um segundo modelo. Esta dissertação se dedicará somente aos combinadores que se classificam como *ensemble*.

Os métodos da média simples (cSA), mediana simples (cSM) e mínima variância (cMV) são exemplos de combinadores lineares. Por outro lado, entre as estratégias não lineares, destacam-se a combinação baseada em ML, tais como a combinação via ANN (cANN) (DE MATTOS NETO *et al.*, 2021) e a combinação via SVR (cSVR) (HAN *et al.*, 2017), além da combinação via cópulas (cCP) (DE OLIVEIRA *et al.*, 2017). Tais abordagens combinadas têm mostrado resultados atraentes.

Em uma revisão da literatura sobre água subterrânea, considerando os últimos cinco anos, nota-se que a maior parte dos artigos tem adotado técnicas individuais de modelagem de séries temporais. Por outro lado, um número menor de estudos têm recorrido à combinação de preditores. Neste caso, a combinação tem sido realizada usando preditores de um mesmo modelos de ML, como em trabalhos de Afan *et al.* (2021), Roy *et al.* (2021) e Wunsch *et al.* (2021), ou usando preditores de diferentes modelos individuais, assim como Xing *et al.* (2018), Iqbal *et al.* (2021), Jiang *et al.* (2022) e Najafabadipour *et al.* (2022). No geral, tais estudos investigaram um número limitado de alternativas de modelos individuais e combinados.

Diante do exposto, este trabalho busca expandir a literatura referente à combinação de preditores de séries temporais de níveis de água subterrânea. Ao todo, doze formalismos diferentes, entre individuais e combinados, do estado-da-arte da modelagem e previsão de séries temporais foram investigados. Adotou-se como caso de estudo a água subterrânea na bacia Sedimentar do Araripe (BSA), no Estado do Ceará, Brasil.

1.1 Objetivos

1.1.1 Objetivo geral

Desenvolver modelos individuais e de *ensemble* para previsão de séries temporais do nível de água subterrânea na bacia Sedimentar do Araripe.

1.1.2 Objetivos específicos

1. Obter séries históricas dos níveis estáticos de poços de monitoramento na bacia Sedimentar do Araripe;
2. Realizar o pré-processamento dos dados, incluindo o preenchimento de falhas nas séries temporais;
3. Construir modelos individuais de previsão das séries temporais;
4. Implementar modelos combinados de previsão das séries temporais;
5. Prever o nível da água subterrânea da bacia Sedimentar do Araripe;
6. Avaliar o desempenho dos modelos de previsão.

1.2 Estrutura do trabalho

O presente trabalho está organizado como segue. O Capítulo 2 apresenta um referencial teórico sobre desenvolvimento sustentável, água subterrânea, modelagem e previsão de séries temporais segundo as abordagens individuais e combinadas, e as técnicas para lidar com o problema dos dados ausentes. O Capítulo 3 discorre sobre os procedimentos metodológicos adotados para a execução do presente trabalho. Os resultados obtidos são apresentados e discutidos no Capítulo 4. Considerações finais são feitas no Capítulo 5.

2 REFERENCIAL TEÓRICO

2.1 Desenvolvimento sustentável e recursos hídricos

As ideias do desenvolvimento sustentável estão fundamentadas no estabelecimento de um equilíbrio harmonioso entre desenvolvimento econômico, social e meio ambiente, de forma que os avanços das gerações atuais não comprometam a disponibilidade de recursos para as gerações futuras (SACHS, 2000). As ações neste sentido devem ocorrer de forma descentralizada e integrada, atendendo às características e necessidades distintas locais e regionais. Portanto, a abordagem em escala regional é essencial, uma vez que provê mecanismos para intervenções de maneira integrada. A partir desta ótica interdisciplinar do desenvolvimento sustentável, é evidente que a sustentabilidade ambiental sofre influência de ações oriundas de vários setores da sociedade. Logo, a água, um recurso natural renovável e finito, é afetada pela maioria das atividades humanas. Neste contexto, justifica-se a importância de estratégias de gestão integrada a nível regional dos recursos hídricos (COUNCIL, 1999). A água está incluída nas agendas internacionais ambientais através do Objetivo de Desenvolvimento Sustentável 6, que visa assegurar a disponibilidade e gestão sustentável da água e saneamento para todos. Para permitir que todas as pessoas tenham acesso à água potável, a Agenda 2030 prevê como meta uma gestão mais responsável dos recursos hídricos, incluindo a implementação de saneamento básico em todas as regiões vulneráveis e a proteção dos ecossistemas relacionados à água (ONU, 2015).

2.2 Recursos hídricos para o semiárido

Em regiões semiáridas, os desafios impostos pelas características hidroclimáticas demandam soluções sustentáveis para a gestão de recursos hídricos. As regiões são marcadas por altas temperaturas, frágil hidrografia e irregular distribuição pluviométrica temporal e espacial. Por conta disso, a água subterrânea é a principal fonte de água em regiões semiáridas (WHEATER *et al.*, 2010).

Particularmente no Brasil, o semiárido brasileiro abrange todos os estados da região Nordeste, além de parte do estado de Minas Gerais, na região Sudeste. A precipitação média anual é igual ou inferior a 800 mm e distribuída em uma única estação chuvosa de três a cinco meses. Os índices de evaporação da água superficial são altos,

o que causa em média uma perda de 2.500 a 3.000 mm por ano (SUDENE, 2017). As condições físico-climáticas do semiárido brasileiro, aliadas às secas periódicas incidentes na região, constituem um problema estrutural com implicações econômicas e sociais que é enfrentado historicamente pela população da região (BEZERRA, 2002; CAMPOS, 2014).

A problemática hídrica no semiárido brasileiro é ponto de intervenções públicas desde o século XIX. Entre as ações mais comumente adotadas, a construção de açudes para armazenamento de água superficial é historicamente o tipo de obra mais executado na região, além da perfuração de poços de captação de água subterrânea. Durante o século XX, órgãos como o Departamento Nacional de Obras Contra a Seca (DNOCS) e a Secretaria de Desenvolvimento do Nordeste (SUDENE) atuaram neste sentido. Entretanto, tais ações possuíam caráter centralizador e isolado, não sendo capazes de explorar o potencial hídrico da região para obter sucesso na busca de uma solução contra seca (REBOUÇAS, 1997; GARJULLI, 2003; CIRILO, 2008). Avanços puderam ser observados após a criação da Lei Federal nº 9.433 de 1997, conhecida como Lei das Águas (BRASIL, 1997). Alinhada aos conceitos de desenvolvimento sustentável, a lei fundamenta-se na ideia da gestão descentralizada, integrada e participativa da água, e fornece os instrumentos legais pelos quais os órgãos e entidades devem atuar, para assegurar a disponibilidade e a qualidade adequada da água, e garantir que os múltiplos usos da água se deem de forma racional e integrada (SOUZA FILHO, 2011).

O desenvolvimento de soluções hídricas sustentáveis no semiárido requer ações de caráter estruturante, que busquem viabilizar uma convivência normal e sustentável com a seca; através da gestão participativa e integrada dos recursos hídricos, levando em consideração as particularidades regionais (ambientais, econômicas e sociais) no processo de tomada de decisão, e que permitam a construção da infraestrutura hídrica necessária para a garantia do abastecimento da população (REBOUÇAS, 1997; CAMPOS, 2014).

A hidrogeologia do semiárido brasileiro é caracterizada por dois contextos distintos: o embasamento cristalino, cujas formações rochosas de baixa permeabilidade ensejam o uso eficiente dos açudes como solução hídrica; e rochas sedimentares, com potenciais reservas de água subterrânea (REBOUÇAS, 1997). No Estado do Ceará, predomina o embasamento cristalino, correspondendo a cerca de 75% do território, e localizado na sua região central. Contudo, na extremidade sul e no litoral do Estado encontram-se formações sedimentares com maior potencial em recursos hídricos subterrâneos (TEIXEIRA, 2003).

Em especial, ao sul do Estado, na divisa com os Estados de Pernambuco e Piauí, está localizada a bacia Sedimentar do Araripe (BSA), área de relevante interesse hidrogeológico pelo seu potencial de qualidade e quantidade de água subterrânea (CPRM, 2009).

2.3 Água subterrânea

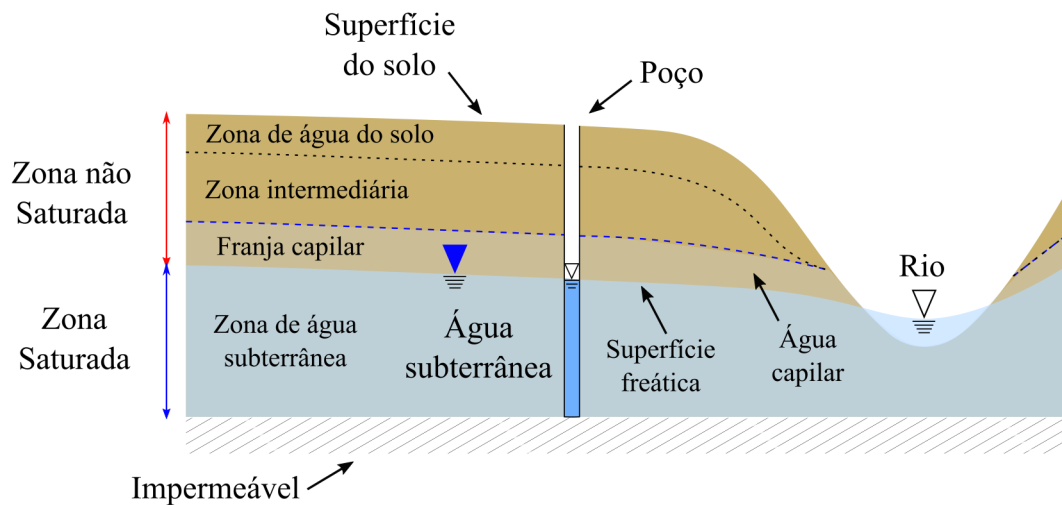
O ciclo hidrológico é o sistema complexo e interligado que representa o permanente percurso da água à medida em que ela se transforma e circula através da terra, oceanos e atmosfera. A água subterrânea é a parcela de água que infiltra e sofre a ação da gravidade até atingir uma profundidade no solo onde todos os poros são preenchidos com água. Neste ponto, o solo ou rocha fica saturado e o nível de água resultante é chamado de lençol freático. Representando cerca de 30% de toda água potável disponível no planeta, a água subterrânea é utilizada como fonte de água para atividades domésticas, industriais ou agrícolas, por meio da perfuração de poços, sobretudo em regiões onde águas superficiais são de difícil acesso. Isso se deve à boa qualidade da água subterrânea e do custo de acessibilidade relativamente baixo. Diante da importância de tal recurso, o estudo da água subterrânea deve ser realizado a partir de uma visão holística em escala regional, abordando os diversos fatores que influenciam as dinâmicas dos sistemas de armazenamento da água subterrânea (FEITOSA, 2008; MANZIONE, 2015).

2.3.1 Sistemas aquíferos

Um aquífero é uma unidade geológica formada por rochas ou sedimentos que contém água e permitem que quantidades significativas dessa água se movimentem no seu interior em condições naturais. Os aquíferos podem ser classificados de acordo com a pressão das águas nas suas superfícies limítrofes superior e inferior, e, também, em função da capacidade de transmissão de água dessas camadas limítrofes. Entre as classificações dos aquíferos, destacam-se os aquíferos livres e confinados (MANZIONE, 2015).

Os aquíferos livres, também conhecidos como freáticos ou não confinados, possuem como limite superior a superfície de saturação ou freática, em que todos os pontos se encontram à pressão atmosférica (FEITOSA, 2008). A Figura 1 mostra a representação esquemática da distribuição vertical da água no solo e subsolo, destacando as diversas zonas de umidade em um aquífero livre.

Figura 1 – Representação esquemática da distribuição vertical da água no solo e subsolo, destacando diversas zonas de umidade em um aquífero livre



Fonte: Elaborado pelo autor (2022), adaptado de Feitosa (2008).

Abaixo da superfície, a água contida no solo pode ser dividida verticalmente em duas camadas: a zona saturada e a zona não saturada. A zona saturada situa-se abaixo da superfície freática, e nela, todos os vazios do solo estão preenchidos com água. Na hidrogeologia, denomina-se por água subterrânea apenas a água que circula na zona saturada. Na zona não saturada, os poros do solo estão parcialmente preenchidos por ar, vapor d'água e água. Essa zona é dividida em três partes: capilar, intermediária e de água do solo. Na zona capilar, em sua parte mais inferior, próxima à superfície freática, os poros encontram-se praticamente saturados, enquanto que no limite de ascensão capilar da água, somente os poros menores estão preenchidos com água. Na zona intermediária, a umidade presente no solo se deve à água capilar isolada. Na zona de água do solo, também conhecida como zona de evapotranspiração, a água capilar isolada ou suspensa é usada pelas plantas para as suas funções de transpiração e nutrição (FEITOSA, 2008).

Por sua vez, os aquíferos confinados são formações geológicas que armazenam água e cujas camadas superiores são impermeáveis. Neste tipo de aquífero, a parcela da recarga proveniente da chuva ocorre apenas nas áreas onde a formação aflora à superfície. Nos aquíferos confinados, o nível da água encontra-se sob pressão, dando origem aos poços artesianos (MANZIONE, 2015).

Além das formações aquíferas, merecem destaque ainda as camadas confinantes, a exemplo dos aquícludes e aquitardos. Tais formações estão presentes geralmente entre

formações aquíferas importantes, conectando-as ou impedindo o fluxo entre os aquíferos. Os aquícludes são formações que podem conter água, inclusive em quantidades significativas, contudo, não possuem a capacidade de transmissão da água em condições naturais. Formações impermeáveis, como camadas de argila, são exemplos de aquícludes. Aquitardos são formações ou camadas semi-permeáveis, delimitadas superior e/ou inferiormente por camadas de permeabilidade muito maior (FEITOSA, 2008; MANZIONE, 2015).

2.3.2 Recarga da água subterrânea

A recarga é o fluxo descendente da água que atinge o lençol freático. Trata-se de uma importante variável hidrológica, cujo estudo é fundamental para o processo de tomada de decisão em recursos hídricos. A estimativa das taxas, épocas e locais de recarga em sistemas aquíferos é essencial para a gestão eficiente dos recursos hídricos subterrâneos, o planejamento do uso e ocupação do solo, e o uso sustentável da água. Diversos elementos do ciclo hidrológico estão envolvidos no processo de recarga dos aquíferos, tais como fatores climáticos, de uso do solo e geológicos (MANZIONE, 2015).

A precipitação é principal variável envolvida. Quando as taxas de precipitação excedem as taxas de evapotranspiração, tem-se condições favoráveis para a recarga. Nos aquíferos livres, a recarga acontece em toda a extensão da formação geológica que os constituem. Os níveis da água subterrânea em aquíferos livres e em áreas aflorantes são indicados pela posição da superfície freática, e tendem a se elevar após eventos de chuva, quando ocorre a recarga, e diminuem durante períodos de seca (MANZIONE, 2015).

A vegetação e a geologia também afetam o processo de recarga. O tipo e densidade de vegetação na superfície definem os padrões de evapotranspiração e influenciam o processo de infiltração da água no solo. Por sua vez, as propriedades físicas do solo e subsolo afetam as condições de fluxo e escoamento da água subterrânea, que se dá através de tubos capilares de diâmetros reduzidos e a velocidades baixas (MANZIONE, 2015).

Propriedades do solo como transmissividade, permeabilidade, porosidade e heterogeneidade são importantes nesse contexto. A transmissividade representa a habilidade de um aquífero de uma determinada espessura transmitir água, já a permeabilidade mede a transmissão do meio e a interconexão dos poros. A velocidade e o fluxo da água no solo são respectivamente funções da porosidade do meio e da heterogeneidade do solo, isto é, das variações na permeabilidade de ponto a ponto (FEITOSA, 2008; MANZIONE, 2015).

Sendo assim, um bom aquífero compreende formações com alta porosidade e permeabilidade. Geralmente, tais características podem ser encontradas em meios porosos, como formações com areia, cascalho, arenito, calcário, entre outros, que favorecem a infiltração, o fluxo e a transmissão da água subterrânea (MANZIONE, 2015).

A água subterrânea está predominantemente armazenada em formações sedimentares. Estas formações originam-se a partir do intemperismo, que decompõe e desintegra rochas pré-existentes. Uma bacia sedimentar é uma feição geográfica localizada em áreas de depressões, que acumulam espessas camadas, ou estratificações, formadas pela deposição de materiais (sedimentação). Geralmente, possuem alta porosidade e permeabilidade, permitindo a exploração de vazões significativas (MANZIONE, 2015).

2.3.3 Monitoramento da água subterrânea

Os múltiplos usos da água subterrânea e a vulnerabilidade das reservas requerem meios para a sua avaliação e acompanhamento. Neste sentido, o monitoramento é um instrumento essencial, uma vez que permite o controle qualitativo e quantitativo da água subterrânea. O processo de monitoramento engloba a coleta, o armazenamento, a análise e a interpretação de dados de água subterrânea (FEITOSA, 2008).

O monitoramento quantitativo tem o objetivo de avaliar a condição quantitativa dos corpos de água subterrânea. Através dos registros de níveis da água em poços de monitoramento, pode-se realizar o acompanhamento da evolução da diminuição dos níveis, além da avaliação da recarga e do balanço hídrico. Sendo assim, o monitoramento quantitativo fornece informações acerca do risco a que se encontram expostas as reservas e auxilia na caracterização hidrogeológica dos aquíferos (FEITOSA, 2008).

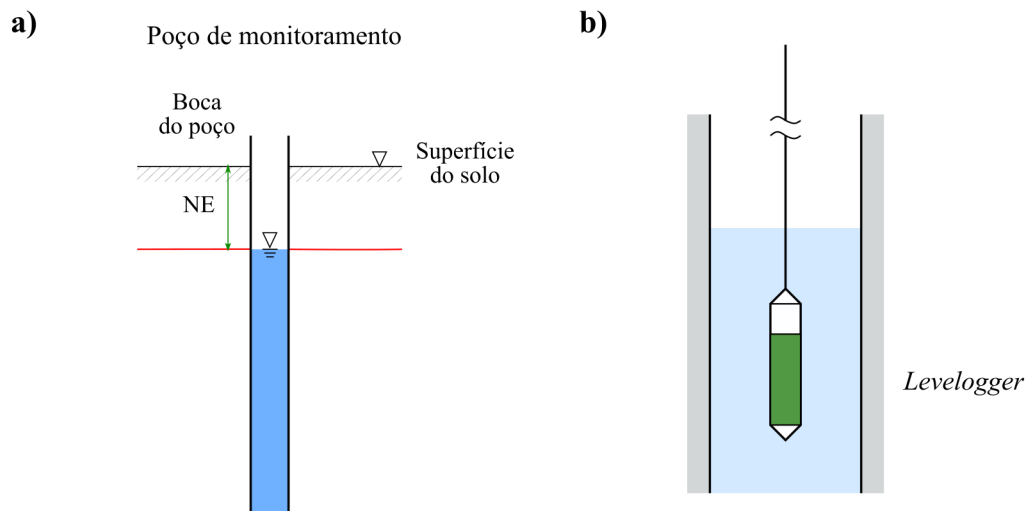
Os programas de monitoramento podem ter abrangência de nível nacional, regional ou local (DIRECTIVE, 2003). São elaborados levando-se em consideração aspectos hidrogeológicos, de uso e ocupação do solo e as necessidades hídricas específicas da região a ser monitorada. Geralmente são aplicados em regiões de interesse hidrogeológico ou em áreas com grande densidade de poços de captação (FEITOSA, 2008; CPRM, 2009).

A nível nacional, os programas englobam os principais sistemas aquíferos de um país, atendendo às políticas nacionais de gerenciamento de recursos hídricos. Os programas regionais atendem aos planos regionais de gestão e complementam os programas nacionais. Por sua vez, os programas de monitoramento realizados em âmbito local são independentes

aos de nível nacional e regional, e têm foco na avaliação da degradação e do risco de contaminação da água subterrânea (MARTINEZ, 2004 apud CPRM, 2009).

Os poços utilizados pelas redes de monitoramento para acompanhar a qualidade e a quantidade da água subterrânea dos aquíferos são denominados poços de monitoramento. A sua finalidade é proporcionar meios de avaliação dos estados qualitativos e quantitativos das reservas, de modo que não é realizado bombeamento, isto é, retirada de água em poços de monitoramento. Denomina-se nível estático (NE) a profundidade da água dentro do poço em estado de repouso, medida em metros a partir do solo. Por outro lado, em poços de captação, onde há bombeamento, o nível dinâmico (ND) corresponde ao nível da água quando o poço está sendo bombeado (FEITOSA, 2008; KARAMOUZ *et al.*, 2012). A Figura 2.a mostra a representação esquemática de um poço de monitoramento, com destaque para a medida do nível estático.

Figura 2 – Representação esquemática de a. poço de monitoramento com a medida do nível estático (NE); e b. instalação da sonda submersa medidora de nível do tipo *levellogger* em um poço de monitoramento



Fonte: Elaborado pelo autor (2022), adaptado de Feitosa (2008).

A medição do nível d'água em poços de monitoramento pode ser realizada por diversos equipamentos. As técnicas usadas incluem métodos manuais e de registro contínuo dos níveis. Encontram-se disponíveis atualmente sistemas telemétricos e automáticos para a medida dos níveis. O sistema é composto por uma sonda submersa, chamada *levellogger*, acoplada a um *datalogger*, que faz a medição periódica e o registro automático dos dados em uma memória. Os medidores do tipo *levellogger* determinam o nível d'água por meio da

pressão exercida sobre eles. São, na verdade, medidores de pressão, que, quando submersos, convertem as variações de pressão registradas em medidas de nível, tomando como base uma medida de nível de referência, realizada no momento da instalação do equipamento. A pressão sobre o sensor é a soma da pressão atmosférica e da pressão da coluna de água (FEITOSA, 2008; MANZIONE, 2015). A Figura 2.b mostra a representação esquemática da instalação do *levelogger* no interior de um poço de monitoramento.

O monitoramento dá suporte ao planejamento, gestão e tomada de decisão em recursos hídricos, uma vez que fornece meios para o diagnóstico e prognóstico do estado quantitativo das reservas de água subterrânea. Sendo assim, a modelagem e previsão dos níveis de água subterrânea é fundamental para avaliação da disponibilidade hídrica (DIRECTIVE, 2003; FEITOSA, 2008; CPRM, 2009). Diante disso, abordagens baseadas na análise de séries temporais constituem importante ferramenta para a modelagem da água subterrânea (MANZIONE, 2015).

2.4 Séries temporais

Séries temporais são coleções de observações realizadas sequencialmente no domínio do tempo. A sua análise permite identificar padrões e prever valores futuros a partir de um histórico conhecido (BOX *et al.*, 2015). Devido à sua natureza estocástica, a modelagem estatística de séries temporais é amplamente utilizada em diversas áreas do conhecimento, inclusive em ciências hidrológicas (MAITY, 2018).

Os diferentes fenômenos que determinam a distribuição espacial e temporal da água no ciclo hidrológico podem ser descritos através de variáveis aleatórias, conhecidas como variáveis hidrológicas. Os níveis freáticos são um exemplo de variável hidrológica. A descrição matemática do comportamento de tais fenômenos hidrológicos, em função do tempo, do espaço ou de ambos dá origem aos chamados processos hidrológicos. Devido às componentes aleatórias presentes nos fenômenos hidrológicos, estes são considerados processos estocásticos (KARAMOUZ *et al.*, 2012). Pode-se definir um processo estocástico como uma coleção de variáveis aleatórias indexadas no domínio do tempo. Uma vez que todos os fenômenos passíveis de acompanhamento histórico originam uma série temporal, tem-se que as séries temporais são instâncias de processos estocásticos. Apesar das variáveis hidrológicas apresentarem variações contínuas no domínio do tempo, os processos de mensuração ou observação de tais fenômenos resultam em registros realizados em

intervalos discretos de tempo. Por isso, as séries temporais de variáveis hidrológicas são consideradas em sua forma discreta (MAITY, 2018).

Entre os formalismos estatísticos para modelagem e previsão de séries temporais mais tradicionais, destacam-se os modelos ARIMA (BOX *et al.*, 2015) e ETS (HYNDMAN *et al.*, 2008). Quanto às abordagens não lineares, técnicas baseadas em inteligência artificial (AI) têm produzido resultados atraentes nas últimas décadas. O aprendizado de máquina (ML), ou *machine learning*, em inglês, é um ramo da AI que dispõe de uma coleção de algoritmos destinados a aprender através de um processo de treinamento e, baseado nisto, prever cenários ou eventos futuros. Entre estes algoritmos, destacam-se os baseados em redes neurais artificiais e máquinas de vetores de suporte (PALIT; POPOVIC, 2005).

2.4.1 ARIMA

O modelo autorregressivo (AR) permite prever valores de uma série temporal a partir da combinação linear de valores anteriores desta série. Este tipo de modelo se baseia na autocorrelação da série temporal, isto é, na relação de dependência linear existente entre os pontos passados da série. Um modelo AR pode ser visto como um modelo de regressão múltipla em que os valores passados da própria série são tomados como variáveis preditoras ou independentes. Um modelo autorregressivo de ordem p considera p valores passados da série, é indicado por $AR(p)$ e expresso pela Equação (1).

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (1)$$

em que $y_{t-1}, y_{t-2}, \dots, y_{t-p}$ são os valores da série nos instantes de tempos $t - 1, t - 2, \dots, t - p$; $\phi_1, \phi_2, \dots, \phi_p$ são coeficientes; c é a constante da equação de regressão e ε_t um ruído branco, isto é, uma série normalmente distribuída sem autocorrelação e com média igual a zero (CHAN; CRYER, 2008; HYNDMAN; ATHANASOPOULOS, 2018).

Além da autocorrelação da série temporal, pode-se também considerar a associação linear existente entre os valores da série e os resíduos provenientes de suas previsões passadas, por meio dos modelos de médias móveis (MA). Assim como o modelo AR, o modelo MA também é semelhante a um modelo de regressão múltipla. Este tipo de modelo considera um número q de resíduos passados. Um modelo de médias móveis de ordem q , indicado por $MA(q)$, pode ser expresso pela Equação (2).

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (2)$$

em que y_t é o valor da série no tempo t ; $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$ são ruídos brancos correspondentes aos resíduos do modelo ao prever $y_t, y_{t-1}, y_{t-2}, \dots, y_{t-q}$; $\theta_1, \theta_2, \dots, \theta_q$ são coeficientes e c é a constante da equação de regressão (CHAN; CRYER, 2008; HYNDMAN; ATHANASOPOULOS, 2018).

O modelo autorregressivo de médias móveis (ARMA) considera tanto a relação de dependência entre os p termos anteriores de uma série temporal, quanto os q resíduos anteriores. Portanto, um modelo ARMA(p, q) é uma combinação entre modelos autorregressivos e modelos de médias móveis. A aplicação dos modelos ARMA requer a estacionariedade da série temporal. Uma série temporal é dita estacionária se as propriedades estatísticas que descrevem o seu comportamento se desenvolvem de forma constante ao longo do tempo. A operação de diferenciação é uma das maneiras mais comuns de transformar uma série temporal não-estacionária em uma série estacionária, e consiste em se tomar diferenças sucessivas entre os termos da série, até se obter a estacionariedade. A primeira diferença é definida por $\Delta y_t = y_t - y_{t-1}$, a segunda diferença é $\Delta^2 y_t = \Delta y_t - \Delta y_{t-1}$, e assim por diante, até que $y' = \Delta^d y_t$ seja uma série estacionária, em que d é o número de vezes que a série foi diferenciada (MORETTIN; TOLOI, 1981).

Após a diferenciação de uma série temporal, esta pode ser representada por um modelo autorregressivo integrado de médias móveis (ARIMA), indicado por ARIMA(p, d, q), e expresso pela Equação (3).

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (3)$$

em que y'_t é a série diferenciada; p é a ordem da parcela autorregressiva; d é o número de diferenciações e q é a ordem da parcela de médias móveis (CHAN; CRYER, 2008; HYNDMAN; ATHANASOPOULOS, 2018).

A definição das ordens p e q dos modelos ARIMA pode ser auxiliada pela análise da função de autocorrelação (FAC) e da função de autocorrelação parcial (FACP), para a definição dos valores passados da série temporal mais relevantes. Contudo, pode-se recorrer à critérios de informação para auxiliar na seleção das ordens do modelo (HYNDMAN; ATHANASOPOULOS, 2018). A estimação dos parâmetros dos modelos AR, MA e ARMA pode ser realizada por meio de diversos métodos, incluindo o método dos momentos, mínimos quadrados e máxima verossimilhança (CHAN; CRYER, 2008).

Os modelos ARIMA também podem ser usados para a modelagem de séries temporais que possuam sazonalidade. Um modelo ARIMA sazonal pode ser cons-

truído incluindo-se termos sazonais adicionais ao modelo ARIMA e pode ser indicado por $ARIMA(p,d,q)(P,D,Q)[m]$, em que p , d e q são as ordens da parcela não sazonal, P , D e Q são as ordens da parcela sazonal, e m é a sazonalidade da série (HYNDMAN; ATHANASOPOULOS, 2018).

2.4.2 ETS

Os métodos de suavização exponencial são uma classe de métodos de previsão de séries temporais, em que as previsões são realizadas a partir de combinações ponderadas de valores passados da série, de modo que valores mais recentes recebem pesos relativamente maiores que valores mais antigos (HYNDMAN *et al.*, 2008). O método de suavização exponencial simples (SES) é o método de suavização exponencial mais simples, utilizado para séries temporais sem tendência e sazonalidade. Uma extensão do método SES, para previsão de séries com tendência, é o método de Holt. A versão sazonal do método de Holt é conhecida como método de Holt-Winters (HYNDMAN; ATHANASOPOULOS, 2018).

Os modelos de previsão referentes aos métodos de suavização exponencial são conhecidos pela tríade (E,T,S), que representa as componentes de erro, tendência e sazonalidade, respectivamente. A notação $ETS(\cdot, \cdot, \cdot)$ indica a manifestação de cada componente no modelo, podendo assumir valores aditivos (letra A), multiplicativos (letra M), ou sem determinada componente (letra N) (HYNDMAN *et al.*, 2008).

Para a representação das variações dos modelos ETS, utiliza-se o conceito de Modelos de Estados de Espaço, mostrado nas Equações (4) e (5).

$$y_t = \psi \dot{\mathbf{x}}_{t-1} + \varepsilon_t \quad (4)$$

$$\dot{\mathbf{x}}_t = F \dot{\mathbf{x}}_{t-1} + g \varepsilon_t \quad (5)$$

em que y_t é o valor da série no instante de tempo t , ε_t é um ruído branco e F , g e ψ são coeficientes do modelo. $\dot{\mathbf{x}}_t$ é um “vetor de estado“, contendo componentes da série, tais como *level* l_t , tendência b_t e sazonalidade s_t . De forma geral, $\dot{\mathbf{x}}_t = (l_t, b_t, s_t, s_{t-1}, \dots, s_{t-m+1})$. No modelo ETS, o *level* é o valor da série suavizada, isto é, o valor resultante da suavização exponencial. A Equação (4) descreve a relação entre os estados $\dot{\mathbf{x}}_{t-1}$ e a série y_t , enquanto a Equação (5) expressa a evolução temporal dos estados (HYNDMAN *et al.*, 2008).

A estimação dos parâmetros dos modelos ETS pode ser realizada pelo método da máxima verossimilhança. Para a seleção do modelo mais apropriado para uma dada

série temporal, pode-se recorrer à seleção por meio de critérios de informação (HYNDMAN; ATHANASOPOULOS, 2018).

2.4.3 Redes Neurais Artificiais

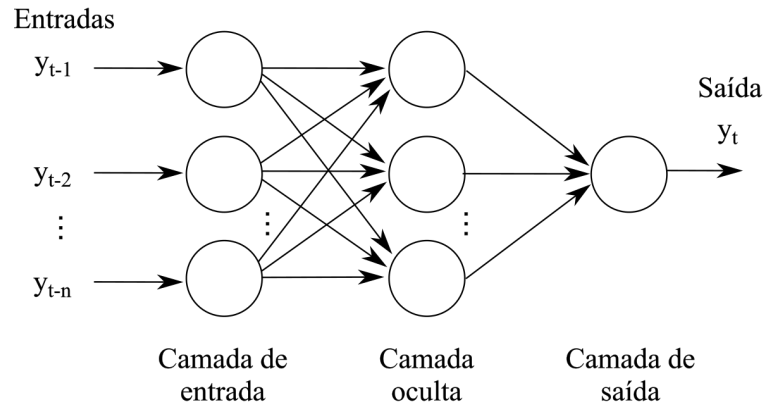
As redes neurais artificiais (ANN), ou *artificial neural networks*, em inglês, são sistemas de processamento baseados no funcionamento do cérebro humano, que possuem a capacidade de aprender a partir do armazenamento de conhecimento experimentado, e de generalizar tal conhecimento adquirido, a fim de realizar previsões de cenários ou eventos futuros. O funcionamento das ANN se dá através de uma rede de conexões entre unidades de processamento, análogas aos neurônios (HAYKIN, 1999; AWAD; KHANNA, 2015).

2.4.3.1 *Multilayer Perceptron*

As redes *feed-forward* de multicamadas, também conhecidas como perceptron multicamadas (MLP), ou *multilayer perceptron*, em inglês, são amplamente utilizadas em problemas de previsão. A sua arquitetura consiste em nós interconectados e dispostos em diversas camadas, de entrada, ocultas e de saída, de modo que cada camada se conecta à camada posterior (HAYKIN, 1999; AWAD; KHANNA, 2015). Para a previsão de séries temporais univariadas, geralmente são usados valores passados da série como entradas da rede. Portanto, neste caso, a MLP pode ser vista como um modelo autorregressivo não linear. O número de nós na camada de entrada é determinado pelo número de valores passados utilizados no modelo autorregressivo. Entretanto, a determinação do número de nós nas camadas ocultas é um processo de tentativa e erro, que pode ser facilitada por técnicas de otimização probabilística. Para a previsão um passo adiante, é necessário apenas um nó na camada de saída (ZHANG, 2001). A Figura 3 mostra uma rede MLP para previsão um passo adiante de séries temporais univariadas, em que y_t representa o valor da série no instante t .

A rede na Figura 3 pode ser expressa matematicamente pela Equação (6), em que os índices i , h e o se referem às camadas de entrada, oculta e de saída, respectivamente. y_i e y_t são as entradas e a saída, e n_i e n_h são o número de nós de entrada e na camada oculta, respectivamente. \mathbf{w} são os pesos das conexões entre os nós e v são constantes de viés (*bias*). A função f é chamada função de ativação (FA), que permite a aplicação da rede para processos não lineares. O objetivo da FA é realizar um mapeamento não linear

Figura 3 – Arquitetura básica de uma rede MLP para previsão de séries temporais univariadas com uma camada oculta



Fonte: Elaborado pelo autor (2022), adaptado de Haykin (1999) e Palit e Popovic (2005).

sobre as operações de soma ponderada calculada anteriormente a ela (LEE *et al.*, 2019).

$$y_t = f\left(\sum_h^{n_h} \mathbf{w}_{ho} \cdot f\left(\sum_i^{n_i} \mathbf{w}_{ih} \cdot y_i + v_h\right) + v_o\right) \quad (6)$$

O processo de aprendizagem de uma ANN se dá através de uma etapa de treinamento, onde o conjunto de dados disponível é dividido em dois subconjuntos, um de treinamento e outro de validação. O subconjunto de treinamento é usado para construção do modelo, enquanto que o de validação é usado para mensurar a capacidade de generalização do modelo. Neste processo, os pesos da rede são alterados de acordo com os dados de entrada, de modo a otimizar uma função objetiva de interesse. Técnicas baseadas no método do Gradiente são amplamente usadas na etapa de treinamento de redes neurais, a exemplo do algoritmo de retropropagação do erro, ou *backpropagation*, em inglês (HAYKIN, 1999; PALIT; POPOVIC, 2005).

2.4.3.2 Extreme Learning Machine

As máquinas de aprendizado extremo (ELM), ou *extreme learning machine*, em inglês, são um algoritmo de treinamento para redes neurais *feed-forward* com uma camada oculta (SLFN), ou *single hidden-layer feed-forward neural networks*, em inglês, originalmente proposto por Huang *et al.* (2006). Diferentemente das ANN tradicionais, cujos algoritmos de treinamento são baseados em gradientes, na estrutura do ELM, os pesos de entradas (conexões entre a camada de entrada e oculta) e os valores das constantes de viés na camada oculta são definidos aleatoriamente. Em seguida, o ELM determina

os pesos da rede por meio de operações matriciais, reduzindo o tempo de treinamento (YADAV *et al.*, 2017; XIONG *et al.*, 2018).

Considere um conjunto composto por n pares $(\mathbf{x}_i, y_i) = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$, onde cada dado \mathbf{x}_i possui k atributos, ou seja, $\mathbf{x}_i = (x_{i1}, \dots, x_{ik})$. No contexto de séries temporais, pode-se considerar $\mathbf{x}_i = (y_{i-1}, \dots, y_{i-k})$. A ELM com L nós na camada oculta e função de ativação f pode ser modelada matematicamente pela Equação (7).

$$y_t = \sum_{j=1}^L \mathbf{w}_{oj} \cdot f(\mathbf{w}_{ij} \cdot \mathbf{x}_t + v_j) \quad (7)$$

em que \mathbf{w}_{ij} é o vetor de pesos entre a camada de entrada e o nó j da camada oculta, \mathbf{w}_{oj} é o vetor de pesos conectando o nó j da camada oculta e o nó de saída, e v_j é a constante de viés do nó oculto j . A Equação (7) pode ser reescrita matricialmente como $\mathbf{H}\eta = \mathbf{Y}$, em que $\eta = [\mathbf{w}_{o1}, \dots, \mathbf{w}_{oL}]^T$, $\mathbf{Y} = [y_1, \dots, y_n]^T$, e \mathbf{H} é chamada matriz de saída da camada oculta, mostrada na Equação (8) (HUANG *et al.*, 2006).

$$\mathbf{H} = \begin{bmatrix} f(\mathbf{w}_1 \cdot \mathbf{x}_1 + v_1) & \cdots & f(\mathbf{w}_L \cdot \mathbf{x}_1 + v_L) \\ \vdots & \ddots & \vdots \\ f(\mathbf{w}_1 \cdot \mathbf{x}_n + v_1) & \cdots & f(\mathbf{w}_L \cdot \mathbf{x}_n + v_L) \end{bmatrix} \quad (8)$$

A determinação dos pesos de saída η é equivalente a solucionar o sistema linear $\mathbf{H}\eta = \mathbf{Y}$. Dessa forma, a solução da Equação (7) é $\hat{\eta} = \mathbf{H}^+ \mathbf{Y}$, em que \mathbf{H}^+ é a matriz inversa generalizada de Moore-Penrose de \mathbf{H} (HUANG *et al.*, 2006).

2.4.3.3 Redes Neurais Recorrentes

As redes neurais recorrentes (RNN), ou *recurrent neural networks*, em inglês, são uma extensão das ANN *feed-forward*, que possuem a capacidade de memorizar temporariamente informações (MANDIC; CHAMBERS, 2001). Ao contrário das redes MLP, que utilizam apenas os dados da camada de entrada para realizar as previsões, as RNN incorporam também a saída gerada pela rede no instante de tempo anterior juntamente com os dados de entrada. A habilidade de memorização deste tipo de rede é dada pelo estado oculto (*hidden state*) (MÜLLER *et al.*, 2021). Seja $y = (y_1, y_2, \dots, y_n)$ uma sequência de comprimento n , e \hat{h}_t a memória da RNN no instante de tempo t , uma RNN atualiza a sua informação de memória através da Equação (9).

$$\hat{h}_t = f(\mathbf{W}_y \cdot y_t + \mathbf{W}_h \cdot \hat{h}_{t-1} + v_t) \quad (9)$$

em que f é a função de ativação, \mathbf{W}_y e \mathbf{W}_h são matrizes de pesos e v_t representa uma constante de viés (SIAMI-NAMINI *et al.*, 2019).

Por outro lado, uma desvantagem das RNN trata-se da perda de informações através das camadas da rede, conhecido como o problema da dissipação do gradiente (*vanishing gradient*). O operador matemático Gradiente mensura quanto a saída de uma função se altera em relação a mudanças ocorridas na entradas desta função. Na RNN, a dissipação do gradiente leva o algoritmo de treinamento da rede a atribuir valores inferiores à matriz de pesos, e assim a rede para de aprender (SIAMI-NAMINI *et al.*, 2019).

2.4.3.4 Long Short-Term Memory

As redes *long short-term memory* (LSTM) são um tipo especial de RNN, capazes de armazenar informações a longo prazo. Inicialmente propostas por Hochreiter e Schmidhuber (1997), as LSTM estendem a memória das RNN através de portões (*gates*), que possibilitam preservar ou descartar informações a partir da sua memória. No geral, a arquitetura de uma rede LSTM é formada por três portões: *Forget*, *Input* e *Output*. Dessa forma, a LSTM é capaz de capturar características relevantes dos dados de entrada por um longo período de tempo, decidindo qual informação deve ser preservada ou descartada.

O *Forget gate* é responsável pela decisão de preservar ou remover informações armazenadas. A partir dos valores de \dot{h}_{t-1} e y_t , a função de ativação sigmoide σ é aplicada, obtendo-se como saída deste portão um valor FG_t , tal que $0 \leq FG_t \leq 1$, em que $FG_t = 0$ ($FG_t = 1$) indica remover (manter) o valor de forma integral. Seja b_{FG} uma constante de viés, a saída do *Forget gate* é dada pela Equação (10) (SIAMI-NAMINI *et al.*, 2019; MÜLLER *et al.*, 2021).

$$FG_t = \sigma \left(\mathbf{W}_{FG_h} [\dot{h}_{t-1}], \mathbf{W}_{FG_y} [y_t], b_{FG} \right) \quad (10)$$

O *Input gate* decide se novas informações devem ser adicionadas à rede. É composto por uma camada com uma função de ativação sigmoide σ , que decide quais valores devem ser atualizados, e uma camada com uma função de ativação tangente hiperbólica \tanh , que cria um vetor com novos candidatos a serem adicionados à memória da LSTM. As saídas destas camadas são mostradas nas Equações (11) e (12), respectivamente.

$$IG_t = \sigma \left(\mathbf{W}_{IG_h} [\dot{h}_{t-1}], \mathbf{W}_{IG_y} [y_t], b_{IG} \right) \quad (11)$$

$$CG_t^* = \tanh \left(\mathbf{W}_{CG_h^*} [\dot{h}_{t-1}], \mathbf{W}_{CG_y^*} [y_t], b_{CG^*} \right) \quad (12)$$

em que IG_t indica em que magnitude o valor deve ser atualizado ou não, e CG_t^* é o vetor de magnitudes de novos candidatos. Assim, a memória da LSTM é atualizada considerando o estado de célula anterior CG_{t-1} e os novos candidatos, segundo a Equação (13) (SIAMI-NAMINI *et al.*, 2019; MÜLLER *et al.*, 2021).

$$CG_t = FG_t \cdot CG_{t-1} + IG_t \cdot CG_t^* \quad (13)$$

O *Output gate* controla se um valor armazenado na memória da LSTM contribui para a saída da rede. As funções de ativação sigmoide e tangente hiperbólica são aplicadas, segundo mostrado na Equação (14).

$$OG_t = \sigma \left(\mathbf{W}_{OG_h} [\dot{h}_{t-1}], \mathbf{W}_{OG_y} [y_t], b_{OG} \right) \cdot \tanh (CG_t) \quad (14)$$

em que OG_t é o valor de saída da rede (SIAMI-NAMINI *et al.*, 2019).

2.4.4 Support Vector Regression

As máquinas de vetores de suporte (SVM), ou *support vector machines*, em inglês, são uma técnica de aprendizado supervisionado usada em problemas de classificação e regressão. A generalização dos conceitos de SVM, utilizada em problemas de regressão, é a regressão por vetores de suporte (SVR), ou *support vector regression*, em inglês (AWAD; KHANNA, 2015; STEINWART; CHRISTMANN, 2008).

Seja um conjunto de dados de treinamento composto por n pares $(\mathbf{x}_i, y_i) = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$, onde cada dado \mathbf{x}_i possui k atributos, ou seja, $\mathbf{x}_i = (x_{i1}, \dots, x_{ik})$, a SVR busca definir uma função $f(\mathbf{x})$ que aproxime da melhor forma os valores de saída y_i , com uma tolerância de erro ϵ . Esta função de regressão é dada pela Equação (15).

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b \quad (15)$$

em que \mathbf{x} são as entradas, \mathbf{w} é um vetor de pesos, b é um parâmetro da regressão e $\langle \cdot, \cdot \rangle$ denota a operação de produto interno. Os parâmetros \mathbf{w} e b são definidos de modo a otimizar o seguinte problema:

$$\begin{aligned} & \text{minimizar} \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i - \xi_i^*) \\ & \text{sujeito a} \quad \begin{cases} y_i - f(\mathbf{x}_i) \leq \epsilon + \xi_i \\ f(\mathbf{x}_i) - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, \end{cases} \end{aligned}$$

em que $C > 0$ é uma constante de compensação que indica a quantidade de erros no conjunto de treinamento, e ξ_i, ξ_i^* são variáveis de folga que medem desvios que excedam ϵ .

Aplicando o método dos multiplicadores de Lagrange a este problema de otimização, a Equação (15) pode ser reescrita segundo a Equação (16).

$$f(\mathbf{x}) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \langle \mathbf{x}_i, \mathbf{x} \rangle + b \quad (16)$$

em que α_i e α_i^* são os multiplicadores de Lagrange e \mathbf{x}_i são os vetores de suporte (SMOLA; SCHOLKOPF, 2003; OLIVEIRA, 2006; YU *et al.*, 2006; LORENA; CARVALHO, 2007).

As SVR são utilizadas tanto para regressão linear quanto não linear. Para problemas não lineares, as SVR mapeiam o conjunto de dados de entrada em um novo espaço de dimensão maior, transformando-os em dados linearmente separáveis, através de funções *kernel*. Desse modo, é possível estabelecer neste novo espaço um hiperplano que faça a separação linear dos dados. A versão não linear da expansão mostrada na Equação (16) é mostrada na Equação (17).

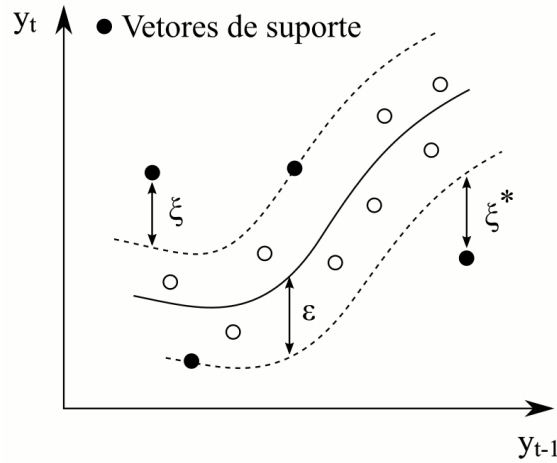
$$f(\mathbf{x}) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \mathbf{K}(\mathbf{x}_i, \mathbf{x}) + b \quad (17)$$

Seja uma transformação $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^{n+h}$, com $\mathbf{x} \rightarrow \Phi(\mathbf{x})$, a função *kernel* pode ser definida como $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$. As SVR não lineares usam as funções *kernels* para calcular produtos escalares entre os dados de entrada em um espaço de dimensão maior. Entre os *kernels* mais comumente utilizados, destaca-se o *Radial Basis Function* (RBF), $\mathbf{K}(x, y) = \exp(-\gamma(x - y)^2)$, onde γ é um parâmetro a ser definido (STEINWART; CHRISTMANN, 2008; AWAD; KHANNA, 2015).

Para a previsão um passo adiante de séries temporais, a SVR usualmente faz uso da autocorrelação da série. Desse modo, os dados de entrada \mathbf{x}_i são os valores anteriores da série, ou seja, $\mathbf{x}_i = (y_{i-1}, \dots, y_{i-p})$ (SILVA, 2020). Sendo assim, a SVR pode ser entendida como um modelo autorregressivo de ordem p . A Figura 4 mostra uma aplicação de SVR para previsão de séries temporais utilizando um termo autorregressivo.

Todos os pontos no interior do ϵ -tubo, isto é, valores $|f(\mathbf{x}) - y_i| \leq \epsilon$ (região tracejada na Figura 4), possuem multiplicadores de Lagrange $\alpha_i = 0$ e $\alpha_i^* = 0$. Por outro lado, apenas para valores $|f(\mathbf{x}) - y_i| \geq \epsilon$ os multiplicadores de Lagrange podem assumir valores diferentes de zero. Os pontos com multiplicadores não nulos são chamados de vetores de suporte (SMOLA; SCHOLKOPF, 2003).

Figura 4 – Aplicação de SVR como um modelo autorregressivo de ordem 1 para previsão um passo adiante de séries temporais



Fonte: Elaborado pelo autor (2022), adaptado de Silva (2020).

2.4.5 Modelos combinados

Abordagens de modelagem combinada têm estado na vanguarda dos modelos de previsão de séries temporais (BATES; GRANGER, 1969). A combinação de preditores proporciona um aumento da qualidade dos formalismos de modelagem, em comparação aos modelos individuais. Encontram-se na literatura diversos métodos de combinação (HAJIRAHIMI; KHASHEI, 2019). Entre eles, destacam-se a combinação pela média simples (cSA), ou *simple average*, em inglês; mediana simples (cSM), ou *simple median*, em inglês, e mínima variância (cMV), além das combinações via ANN (cANN), SVR (cSVR) e cópulas (cCP) (SALES, 2019; SILVA, 2020).

A combinação de preditores leva em consideração as estimativas geradas por N modelos individuais, $y_{t,N}$, para um determinado valor a que se deseja prever, diga-se y_t , no instante de tempo t . O resultado da combinação é então obtido pela soma dos preditores ponderados, e o modelo é construído através de funções de combinação linear ou não linear. De maneira geral, a combinação pode ser expressa pela Equação (18).

$$\hat{y}_t = \varphi(y_{t,1}, y_{t,2}, \dots, y_{t,N}) \quad (18)$$

em que $y_{t,i}$ é a estimativa proveniente do preditor i para o valor da série no instante t (y_t) e φ é uma função de combinação linear ou não linear dos preditores.

2.4.5.1 Média Simples e Mediana Simples

O combinador cSA calcula a média aritmética simples dos N modelos individuais utilizados, e pode ser expresso segundo a Equação (19).

$$\hat{y}_t = \frac{\sum_{i=1}^N y_{t,i}}{N}. \quad (19)$$

De modo semelhante, o combinador cSM calcula a mediana simples dos N modelos individuais considerados. A mediana é o valor que separa um conjunto de dados ordenados em duas metades (SALES, 2019; SILVA, 2020).

2.4.5.2 Mínima Variância

Para o modelo cMV, os pesos são determinados em função das covariâncias dos erros dos modelos individuais. Este combinador envolve a variância do erro de cada preditor individual e a sua covariância em relação aos demais modelos individuais. Desse modo, quanto maior as covariâncias referentes a um dado preditor individual, menor será o peso atribuído (SALES, 2019; SILVA, 2020). A Equação (20) mostra o combinador cMV.

$$\hat{y}_t = \sum_{i=1}^N \omega_i y_{t,i} \quad \text{em que} \quad \omega_i = \frac{\sum_{l=1}^N a_{li}}{\sum_{l=1}^N \sum_{j=1}^N a_{lj}}, \quad (20)$$

em que a_{lj} é o elemento j da linha l da matriz de covariâncias inversa M^{-1} referente aos resíduos dos modelos individuais, e ω_i são os pesos associados ao modelo i . Destaque-se que nesta classe de modelos, os pesos obtidos através do método de mínima variância coincidem com aqueles provenientes do método de máxima verossimilhança quando os erros seguirem uma distribuição normal conjunta (FIRMINO *et al.*, 2014).

2.4.5.3 Cópulas

Para a combinação via cópulas, considere que $\mathbf{X}_t = (X_{t,1}, \dots, X_{t,i}, \dots, X_{t,N})$ é o vetor de N preditores individuais para as observações y_t da série temporal, $\mathbf{x}_t = (x_{t,1}, \dots, x_{t,i}, \dots, x_{t,N})$ é o vetor das estimativas correspondentes aos preditores de y_t , e $\mathbf{E}_t = (E_{t,1}, \dots, E_{t,i}, \dots, E_{t,N})$ corresponde ao vetor de erros dos preditores individuais. As Equações (21) e (22) mostram o cálculo do erro $E_{t,i}$ segundo uma estrutura aditiva, e do

resíduo, isto é, a instância, associado a $E_{t,i}$, respectivamente.

$$E_{t,i} = y_t - X_{t,i} \quad (21)$$

$$e_{t,i} = y_t - x_{t,i} \quad (22)$$

A estrutura de erro multiplicativa pode ser adotada de maneira semelhante, mas utilizando-se a operação de divisão em vez da subtração na Equação (21). Desse modo, $\mathbf{e}_t = (e_{t,1}, \dots, e_{t,i}, \dots, e_{t,N})$ é o conjunto de resíduos observados das estimativas para y_t , isto é, uma instância de \mathbf{E}_t (OLIVEIRA, 2014; ASSIS, 2017; SILVA, 2020).

A distribuição acumulada conjunta de $\mathbf{E}_{t,i}$ é dada por $F_{\mathbf{E}_{t,i}}(e_{t,1}, \dots, e_{t,N})$ com distribuição acumulada marginal $F_{E_{t,i}}(e_{t,i})$. Considerando a estrutura de erros aditiva, a distribuição acumulada marginal de $E_{t,i}$ e a função de densidade de probabilidade são dadas respectivamente segundo as Equações (23) e (24) (SILVA, 2020).

$$F_{E_{t,i}}(e_{t,i}) = P(E_{t,i} < y_t - x_{t,i}) \quad (23)$$

$$p_{E_{t,i}}(e_{t,i}) = p_{E_{t,i}}(y_t - x_{t,i}) \quad (24)$$

Uma cópula $C(\cdot)$ é uma função que combina distribuições univariadas para construir uma distribuição multivariada, incorporando a interdependência entre essas distribuições univariadas (NELSEN, 2013). As distribuições marginais de $C(\cdot)$ são uniformes no intervalo $[0, 1]$, de modo que u_i ($i = 1, \dots, N$) é uma instância de N distribuições acumuladas marginais unitárias, tais que $u_i \in [0, 1]$. Seja u_i a distribuição acumulada marginal dos erros do modelo i diante do resíduo $e_{t,i}$ (Equação 25), a cópula é dada pela distribuição acumulada conjunta mostrada na Equação (26).

$$u_i = F_{E_{t,i}}(e_{t,i} | \alpha_i) \quad (25)$$

$$C(u_1, \dots, u_N) \quad (26)$$

em que α_i são os parâmetros das respectivas distribuições acumuladas marginais univariadas. Por meio da Equação (25), as cópulas transformam $(e_{t,1}, \dots, e_{t,i}, \dots, e_{t,k})$ em $(u_1, \dots, u_i, \dots, u_k)$. Logo, a função de densidade de probabilidade conjunta de \mathbf{E}_t é dada pela Equação (27).

$$p_{\mathbf{E}_t}(e_{t,1}, \dots, e_{t,k}) = c_{\mathbf{E}_t}(u_1, \dots, u_k) \prod_{i=1}^k p_{E_{t,i}}(e_{t,i}) \quad (27)$$

Diante das estimativas $(x_{t,1}, \dots, x_{t,N})$ dos N modelos individuais para y_t , a incógnita da Equação (27) passa a ser apenas y_t . Sendo assim, a combinação via cópulas

define um espaço de possibilidades para o valor de y_t , representado por 1000 pontos igualmente espaçados e, deles, o valor mais verossímil estimado para y_t , segundo a Equação (27), é adotado como estimativa do combinador baseado em cópulas (cCP) (OLIVEIRA, 2014; ASSIS, 2017; SILVA, 2020).

Aqui, supõem-se que $c_{\mathbf{E}_t}$ e $p_{E_{t,i}}(e_{t,i})$ são invariantes no tempo, podendo-se suprimir o índice t . Ainda, destaca-se que a função distribuição acumulada multivariada $C(\cdot)$ poderá assumir qualquer tipo de cópula, sendo $c(\cdot)$ a sua respectiva função densidade de probabilidade (ASSIS, 2017; SILVA, 2020).

2.4.5.4 Combinação baseada em aprendizado de máquina

Os diversos preditores individuais podem exibir relações lineares ou não lineares entre si. Para contornar esta situação, uma alternativa é a adoção de funções não lineares de combinação (YU *et al.*, 2005). Destacam-se estratégias como a combinação via redes neurais artificiais (cANN) e *support vector regression* (cSVR). Desse modo, as entradas dos modelos são as previsões dos k modelos individuais, e a saída é o respectivo valor da série temporal em questão. De forma semelhante à modelagem individual baseada em ML, os parâmetros dos modelos podem ser estimados segundo técnicas de otimização probabilística. A diferença está nas camadas de entrada, uma vez que o número de variáveis de entrada na modelagem combinada será dado em função do número de modelos individuais utilizados (SALES, 2019; SILVA, 2020).

2.5 Algoritmos genéticos

Os algoritmos genéticos (GA) são algoritmos de busca estocástica utilizados para a solução de problemas complexos de otimização. Os GA permitem explorar um espaço de busca para identificar soluções tão ótimas quanto possível. Baseados nos princípios da evolução biológica e da seleção natural, os GA simulam o processo de evolução de organismos vivos em um ambiente competitivo ao reproduzir mecanismos biológicos, como mutação, *crossover* e seleção (WHITLEY, 1994).

A implementação do GA tem início com uma população aleatória de indivíduos (cromossomos), que são definidos segundo um conjunto de características (genes). Em problemas de otimização, deseja-se otimizar um conjunto de variáveis a fim de maximizar

ou minimizar alguma função objetiva. Os GA adotam uma função objetiva como a função de aptidão para avaliar o desempenho dos indivíduos. Na seleção, o GA seleciona os indivíduos mais aptos para participar do processo de reprodução para originar a próxima geração. O *crossover* gera descendentes por meio da combinação das informações genéticas de dois ou mais parentes. Por fim, um ou mais genes dos descendentes são alterados por meio da mutação, para garantir a diversidade da população (MIRJALILI, 2019).

Os algoritmos genéticos (GA) têm sido usados em diversas áreas do conhecimento (LEE, 2018), inclusive nas ciências hidrológicas (HAN *et al.*, 2020; TEGEGNE; KIM, 2020). Na modelagem e previsão de séries temporais, os GA podem ser usados em conjunto com técnicas baseadas em aprendizado de máquina (ML) (KATOCH *et al.*, 2021). Particularmente, na área de água subterrânea, autores como Mirarabi *et al.* (2019) e Seifi *et al.* (2020) têm empregado os GA para a otimização de parâmetros dos modelos de ML.

Na literatura referente à ML, costuma-se denominar como hiper-parâmetros aqueles parâmetros de um modelo que não podem ser estimados diretamente durante o treinamento, mas que devem ser definidos previamente, pois configuram e definem a arquitetura do modelo. São exemplos de hiper-parâmetros a função de ativação (FA) e o número de nós de uma rede neural. Diante das diferentes opções de valores para os hiper-parâmetros, pesquisadores têm buscado desenvolver técnicas para selecionar a melhor arquitetura dos modelos. Este processo, conhecido como otimização de hiper-parâmetros, é importante pois reduz os esforços humanos, aprimora o desempenho dos modelos e contribui para a reprodutibilidade dos experimentos em ML (YANG; SHAMI, 2020).

O procedimento para otimização de hiper-parâmetros de modelos de ML usando GA é descrito a seguir. (i) inicializar aleatoriamente população (espaço de busca), cromossomos (hiper-parâmetros) e genes (valores dos hiper-parâmetros); (ii) avaliar o desempenho de cada indivíduo da geração atual segundo a função de ajuste (função objetiva); (iii) realizar as operações de seleção, *crossover* e mutação no cromossomos, para produzir uma nova geração, contendo as configurações de hiper-parâmetros que serão avaliadas em seguida; (iv) repetir as etapas (ii) e (iii) até que os critérios de parada sejam atendidos; e (v) finalizar e retornar a configuração ótima de hiper-parâmetros (YANG; SHAMI, 2020).

Nas seções seguintes do presente texto, os hiper-parâmetros serão referidos simplesmente como parâmetros. Na Seção 3.7 serão apresentados com maiores detalhes os parâmetros otimizados com GA dos modelos de ML usados neste trabalho.

2.6 Análise de dados ausentes

Em processos de mensuração ou observação é comum a ocorrência de períodos sem informações ou com falhas nas observações. Isso se deve a diversos fatores, tais como erros humanos, falhas em equipamentos de monitoramento, entre outros. Os dados ausentes, comumente referidos como falhas, são um dos principais problemas em estudos hidrológicos (DA CUNHA JÚNIOR; FIRMINO, 2022). Geralmente, denotam-se por valores ausentes (NA), ou *not available*, em inglês (MORITZ *et al.*, 2015).

Existem diversas abordagens para lidar com o problema de dados ausentes. Entre elas, estratégias baseadas em se descartar os registros com falhas e analisar apenas os dados que não possuam valores ausentes podem levar a resultados enviesados, além de diminuir o tamanho da amostra. Nesse contexto, as técnicas de imputação constituem uma alternativa confiável para a solução do problema (ENDERS, 2010).

As falhas nos dados são um problema antigo. Na hidrologia, alguns dos métodos clássicos desenvolvidos para lidar com problemas relacionados à falhas nos dados datam do século XX (PAULHUS; KOHLER, 1952; SEARCY; HARDISON, 1960). Por outro lado, a teoria dos dados ausentes foi construída fundamentando-se na Estatística, sobretudo pela contribuição dos trabalhos de Little e Rubin, a partir da década de 1970 (LITTLE; RUBIN, 2019). Embora o termo “preenchimento de falhas” seja o mais tradicionalmente usado em ciências hidrológicas, nota-se que, nos últimos anos, tem sido cada vez mais comum pesquisadores da hidrologia se referirem ao problema por “imputação”, como, por exemplo, Feng *et al.* (2014), Zhang *et al.* (2017) e Teegavarapu (2020). Além disso, é comum que estudos hidrológicos que envolvam modelagem de séries temporais também utilizem o termo “imputação” para se referir ao processo de preenchimento de falhas na etapa de pré-processamento dos dados, que precede a modelagem (SAHU *et al.*, 2020; MÜLLER *et al.*, 2021; NOORI; SINGH, 2021), uma vez que o termo “imputação” é o mais usado em estudos de séries temporais (MORITZ *et al.*, 2015; PHAN *et al.*, 2020). Diante disso, no presente trabalho, termos como “falhas”, “dados ausentes” e “dados faltantes” serão adotados para se referir à ausência de valores em um conjunto de dados ou série temporal, enquanto que “imputação” e “preenchimento de falhas” serão usados para se referir ao processo de se estimar valores em substituição àqueles ausentes.

2.6.1 Imputação de dados ausentes em séries temporais

A imputação de valores ausentes é uma técnica que consiste em se estimar de maneira razoável valores plausíveis para substituir as falhas ou os pontos ausentes em um conjunto de dados, baseando-se nas informações disponíveis. Compreende um extenso campo de estudo, que possui aplicações em diversas áreas do conhecimento (ENDERS, 2010; LITTLE; RUBIN, 2019).

A modelagem de processos hidrológicos geralmente leva em consideração as informações disponíveis em pontos ou estações de monitoramento vizinhas. Assim, é comum se trabalhar com conjuntos de dados multivariados (KARAMOUZ *et al.*, 2012). Conseqüentemente, é importante o estudo de métodos de imputação para séries temporais multivariadas (JUNNINEN *et al.*, 2004). Nesse sentido, encontram-se disponíveis na literatura diversos métodos de imputação de valores ausentes comumente aplicados em estudos hidrológicos, tais como os baseados em interpolação espacial (HARMAN *et al.*, 2016; ASGHARINIA; PETROSELLI, 2020), regressão linear múltipla e aprendizado de máquina (KASHANI; DINPASHOH, 2012; MEKANIK *et al.*, 2013; MAITY, 2018).

De maneira geral, muitas das técnicas de análise de dados ausentes difundidas na literatura foram desenvolvidas com foco em conjuntos de dados multivariados. Portanto, os métodos de imputação mais tradicionalmente usados geralmente se fundamentam nas relações existentes entre as variáveis de um conjunto de dados (ENDERS, 2010; LITTLE; RUBIN, 2019). Contudo, o problema dos dados ausentes também é recorrente em estudos que envolvam séries temporais de apenas uma variável (DA CUNHA JÚNIOR *et al.*, 2022). Para o caso especial das séries temporais univariadas, essas abordagens não podem ser aplicadas diretamente. Diante disso, autores têm se dedicado ao desenvolvimento de alternativas de algoritmos para a solução desse problema (BOKDE *et al.*, 2018; PHAN *et al.*, 2020; SAVARIMUTHU; KARESIDDIAIAH, 2021). As estratégias variam desde algoritmos simples de imputação, como a imputação pela média ou a imputação aleatória, até métodos mais sofisticados, que exploram explicitamente as características da série, como os métodos baseados em interpolação e em médias móveis. Ainda, pode-se recorrer à aplicação de métodos multivariados em dados defasados (MORITZ *et al.*, 2015).

2.6.1.1 Interpolação

Métodos de imputação baseados em interpolação são amplamente usados em séries temporais, uma vez que a interpolação permite estimar novos pontos de dados dentro de um intervalo conhecido. O objetivo é ajustar uma função relativamente mais simples, tais como funções lineares, polinomiais ou *splines*, a um conjunto de observações. Seja um conjunto de $n + 1$ pontos $(x_i, g(x_i)) = \{(x_0, g(x_0)), \dots, (x_n, g(x_n))\}$, com $g(x_i) = y_i$, a interpolação é a aproximação da função g por uma outra função \hat{g} , onde \hat{g} é então usada em substituição de g . Seja $\hat{g} = p_n(x)$ um polinômio de grau n que interpola g nos pontos x_0, \dots, x_n , então existe um e apenas um polinômio de grau menor ou igual a n que atravessa os $n + 1$ pontos. Assim, a determinação de $p_n(x)$ permite calcular valores intermediários. A Equação (28) mostra a representação de Lagrange de $p_n(x)$ (CHAPRA; CANALE, 2016).

$$p_n(x) = \sum_{k=0}^n y_k L_k(x) \quad \text{em que} \quad L_k(x) = \prod_{\substack{j=0 \\ j \neq k}}^n \frac{x - x_j}{x_k - x_j} \quad (28)$$

A interpolação linear é caso especial para a interpolação entre dois pontos distintos $\{(x_0, g(x_0)), (x_1, g(x_1))\}$, e assim $n = 1$ e $p_n(x)$ é uma função linear, mostrada na Equação (29).

$$p_1(x) = y_0 \frac{x - x_1}{x_0 - x_1} + y_1 \frac{x - x_0}{x_1 - x_0} \quad (29)$$

Funções polinomiais de graus mais altos podem capturar os ruídos dos dados e, como consequência, levar a erros de estimativa. Uma outra abordagem para interpolação é determinar polinômios de graus inferiores para subconjuntos dos pontos, através de funções polinomiais definidas por trechos. Adotando-se polinômios de grau três nos intervalos entre cada pontos, tem-se as chamadas *splines* cúbicas (CHAPRA; CANALE, 2016).

2.6.1.2 Médias móveis

Para imputação de valores ausentes, o método das médias móveis calcula médias a partir de um número igual de observações em ambos os lados de um valor ausente central. Seja um valor ausente y_t em uma série temporal, considerando uma janela de média móvel M , os valores y_{t-M}, \dots, y_{t-1} e y_{t+1}, \dots, y_{t+M} são usados para calcular a média. Diversas configurações de pesos podem ser adotadas. Na média móvel simples (SMA), ou *simple moving average*, em inglês, os valores são calculados pela média aritmética simples das

observações, isto é, todos os valores recebem o mesmo peso. Podem ser aplicados ainda pesos com decaimento linear, por meio da média móvel linear (LWMA), ou *linear weighted moving average*, em inglês (MORITZ; BARTZ-BEIELSTEIN, 2017a). A média móvel exponencial (EWMA), ou *exponentially weighted moving average*, em inglês, calcula uma média aritmética ponderada, cujos pesos decaem exponencialmente segundo a idade das observações. Em outras palavras, a EWMA atribui pesos menores para os dados quanto mais antigos estes forem (HUNTER, 1986).

2.6.1.3 Imputação com decomposição

Métodos de decomposição de séries temporais têm o objetivo de identificar e isolar as componentes que representam os padrões existentes nas séries temporais, como tendência e sazonalidade. A tendência representa mudanças a longo prazo no valor da série, e a sazonalidade são flutuações periódicas de frequência fixa. A diferença entre o efeito combinado destas duas componentes e os dados originais resulta em uma componente de erro, conhecida como componente irregular ou resíduo. A decomposição do tipo aditiva assume que $y_t = S_t + T_t + R_t$, onde S_t , T_t e R_t são as componentes sazonal, tendência e resíduo, respectivamente (HYNDMAN; ATHANASOPOULOS, 2018).

Existem diversos métodos de decomposição de séries temporais. Em especial, a decomposição sazonal e de tendência usando *Loess* (STL) é um método de decomposição aditiva versátil e robusto, desenvolvido por Cleveland *et al.* (1990). O método STL decompõe a série nas componentes S_t , T_t e R_t através de uma sequência de aplicações do suavizador *Loess*. Trata-se de um método de suavização de séries temporais que realiza regressões polinomiais locais ponderadas em cada ponto da série, usando como variáveis explicativas os valores mais próximos (CLEVELAND; DEVLIN, 1988).

O método STL permite o controle da suavidade da tendência e das variações da sazonalidade. A robustez à *outliers* permite que valores incomuns afetem apenas a componente de resíduo. Uma vantagem do uso da função *Loess* para a decomposição é que esta pode ser definida em todos os pontos da série, diferentemente de outros métodos de decomposição, como as médias móveis. Ainda, a STL é capaz de lidar com séries sazonais com qualquer frequência sazonal maior que um (CLEVELAND *et al.*, 1990).

É importante destacar que, de acordo com Bergmeir *et al.* (2016), utilizar a modelagem ETS para decomposição pode não levar a bons resultados, uma vez que

os componentes de um modelo ETS são definidos com base nos termos de ruído e se desenvolvem dinamicamente com o ruído. Por outro lado, na STL, as componentes de tendência e sazonalidade são suaves e o seu comportamento no domínio do tempo não depende da componente de resíduo diretamente.

Uma vez decomposta a série, pode-se analisar as componentes de forma separada e usar a decomposição STL para previsão ou imputação (THEODOSIOU, 2011). A imputação com STL consiste em determinar a componente sazonal (S_t) e aplicar algum método de imputação à série ajustada sazonalmente ($T_t + R_t$). Em seguida, a componente sazonal é adicionada de volta, e os valores estimados são usados para preencher as falhas na série original (MORITZ; BARTZ-BEIELSTEIN, 2017a).

A implementação computacional da decomposição STL em linguagem R requer uma série completa sem falhas, portanto, um processamento prévio é necessário. Nesta etapa, pode-se utilizar algum método de imputação, como uma interpolação linear (MORITZ; BARTZ-BEIELSTEIN, 2017a). Destaque-se que para casos com longos intervalos de falhas, o processamento prévio pode comprometer a determinação da componente sazonal, e por conseguinte, diminuir a qualidade da imputação final (LIU *et al.*, 2020).

Moritz *et al.* (2015) compararam o desempenho de diversos métodos tradicionais de imputação de séries temporais univariadas. Seus resultados apontam que a interpolação linear na série ajustada sazonalmente mostrou resultados superiores em relação aos outros métodos, principalmente para séries com forte sazonalidade. De maneira geral, o método apresenta bom desempenho para séries com sazonalidade.

2.6.1.4 Regressão em valores defasados

No estudo de séries temporais univariadas, o tempo pode ser entendido como uma variável implícita. Portanto, ao se incorporar informações de tempo às análises, é possível a aplicação de algoritmos multivariados. Isso pode ser feito por meio dos valores defasados da série, assim como nos problemas de previsão. Assim, converte-se o problema univariado em um problema de regressão, onde o modelo toma como preditores os valores passados da própria série. Essa metodologia pode ser estendida ao problema de imputação (MORITZ *et al.*, 2015). As vantagens dos modelos baseados em ML em problemas de regressão também podem ser aproveitadas neste sentido (GUPTA; LAM, 1996; HONGHAI *et al.*, 2005; LI *et al.*, 2009; AYDILEK; ARSLAN, 2013).

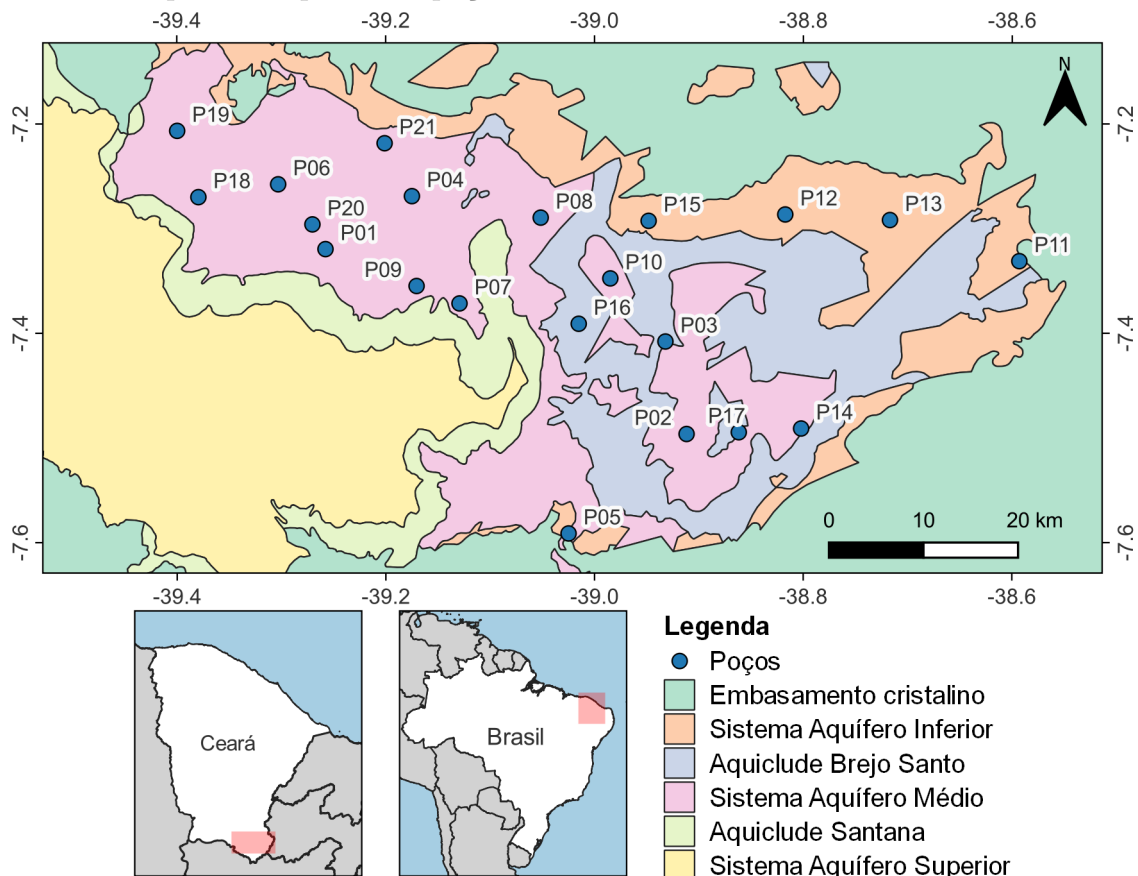
A partir de uma série temporal incompleta, isto é, com valores ausentes, pode-se construir uma matriz de valores defasados $\mathbf{Y}^* = (y_t, y_{t-1}, \dots, y_{t-p})$. Um registro incompleto é definido como uma linha da matriz \mathbf{Y}^* em que uma ou mais colunas possuem dados ausentes, enquanto que um registro completo refere-se a uma linha de \mathbf{Y}^* em que nenhuma coluna possui dados ausentes. A ideia da utilização de modelos de ML para imputação de dados ausentes é usar os registros completos para treinar o algoritmo, e baseado nisto, prever os registros incompletos. Neste processo, a matriz \mathbf{Y}^* é dividida em dois subconjuntos: \mathbf{Y}_{comp}^* e \mathbf{Y}_{incomp}^* . O subconjunto \mathbf{Y}_{comp}^* contém todos os registros completos, e \mathbf{Y}_{incomp}^* é composto pelos registros incompletos (HONGHAI *et al.*, 2005; LI *et al.*, 2009).

3 METODOLOGIA

3.1 Área de estudo

A área de estudo do presente trabalho foi a porção leste da bacia Sedimentar do Araripe (BSA), no Estado do Ceará, Brasil. Está inserida no Cariri, uma região de planejamento do Estado do Ceará, composta por 29 municípios, abrangendo uma área de 17.390,30 km², e uma população total estimada em 1.014.610 habitantes segundo o IPECE (2017). Engloba os três principais municípios da Região Metropolitana do Cariri (RMC): Juazeiro do Norte, Crato e Barbalha, além dos municípios de Missão Velha, Brejo Santo, Mauriti, Milagres, Porteiras, Jardim e Abaiara (COGERH, 2009). A Figura 5 mostra a localização da área de estudo, os aquíferos da BSA e os poços selecionados.

Figura 5 – Localização da porção leste da bacia Sedimentar do Araripe, com destaque para os aquíferos e poços selecionados



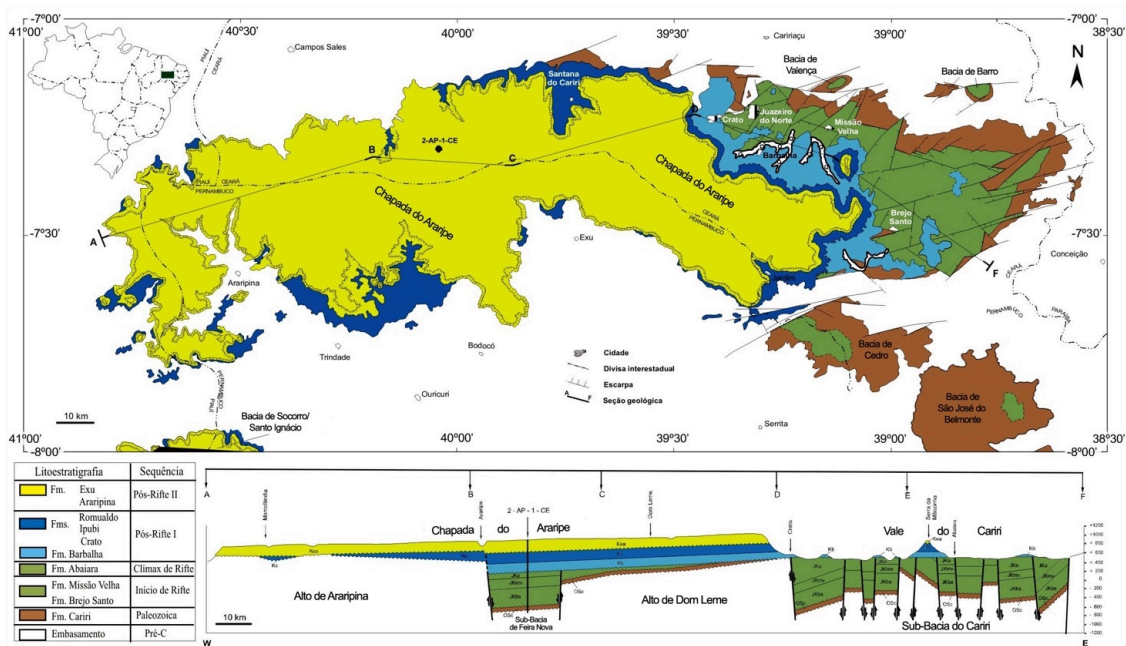
Fonte: Elaborado pelo autor (2022), adaptado de CPRM (2009) e COGERH (2009).

A região tem em seu relevo dois domínios principais: planalto e depressão, conhecidos respectivamente como a Chapada do Araripe, que ocupa 73% da área total da

Bacia, e o Vale do Cariri. O clima é tropical quente semiárido brando, tropical quente e tropical quente sub-úmido (COGERH, 2009). O regime pluviométrico é caracterizado pela grande irregularidade interanual e pela variabilidade temporal e espacial das chuvas. A precipitação média anual na região do Cariri é 700,0 mm, com chuvas concentradas na chamada estação chuvosa, compreendida entre os meses de dezembro a maio. Nos meses de dezembro e janeiro, ocorrem as chamadas chuvas de pré-estação, sendo o início da quadra chuvosa no mês de fevereiro de cada ano (TEIXEIRA, 2003; SILVA *et al.*, 2021).

A bacia Sedimentar do Araripe (BSA) é a maior dentre as bacias interiores da região Nordeste do Brasil, com mais de 9.000 km². A sua hidrogeologia é caracterizada pela diversidade litoestratigráfica, que propicia a ocorrência de uma alternância de aquíferos, aquíferos e aquícludes, com variações e descontinuidades espaciais (FAMBRINI *et al.*, 2020). Dessa forma, é comum se adotar a seguinte divisão hidrogeológica da BSA: (i) Sistema Aquífero Superior, representado pelas formações Exu e Arajara; (ii) Aquíclode Santana, caracterizado pela formação homônima; (iii) Sistema Aquífero Médio, representado pelas formações Rio da Batateira, Abaiara e Missão Velha; (iv) Aquíclode Brejo Santo, definido pela formação homônima; e (v) Sistema Aquífero Inferior, determinado pela Formação Mauriti e parte basal da Formação Brejo Santo (COGERH, 2009). A Figura 6 mostra o mapa geológico e estratigráfico da BSA.

Figura 6 – Mapa geológico e estratigrafia da bacia Sedimentar do Araripe



Fonte: Fambrini *et al.* (2020).

Segundo a COGERH (2009), o Sistema Aquífero Médio é predominantemente livre. A Formação Missão Velha aflora única e exclusivamente no Vale do Cariri e, no geral, a exploração do sistema ocorre em condições livres (CPRM, 2009). Neste sistema, a maior parcela da recarga provém diretamente da chuva nas áreas aflorantes, havendo ainda contribuição de fontes no sopé da Chapada do Araripe (COGERH, 2009). Trata-se do aquífero mais importante em termos de potencial de exploração e o mais explorado da região do Cariri (CPRM, 2009).

O Sistema Aquífero Inferior é confinado e silicificado na sua maior parte. Neste sistema, a base da Formação Brejo Santo corresponde à uma zona confinada, sendo predominantemente argilosa. Esta camada impede a circulação descendente da água do Sistema Aquífero Médio para o Sistema Aquífero Inferior. A recarga do Sistema Aquífero Inferior é realizada por drenagens existentes e pela precipitação pluviométrica direta na área livre em que aflora a Formação Mauriti (COGERH, 2009).

Os poços selecionados na área de estudo monitoram as águas do Sistema Aquífero Médio e do Sistema Aquífero Inferior. Apesar disso, é importante destacar que os poços estão instalados nas áreas aflorantes das Formações Missão Velha e Mauriti, onde os sistemas se comportam como aquíferos livres. A Tabela A.1 do apêndice A mostra informações gerais e geológicas dos poços selecionados.

3.2 Conjunto de dados

O conjunto de dados utilizado consiste em 21 séries temporais da evolução do nível estático (NE) medido em poços de monitoramento instalados sobre os aquíferos da BSA. Os poços selecionados fazem parte da Rede Integrada de Monitoramento de Águas Subterrâneas (RIMAS), coordenada e operada pelo Serviço Geológico do Brasil (CPRM). Os dados foram coletados do portal RIMAS *Web* (<<http://rimasweb.cprm.gov.br/layout/>>) (RIMAS, 2021). A partir dos dados obtidos, foram construídas séries dos níveis médios mensais para cada poço.

3.3 Arquitetura do estudo

No presente estudo, 12 formalismos, entre estratégias individuais e combinadas, foram avaliados para a modelagem e previsão de 21 séries temporais mensais de níveis de

água subterrânea. Os 6 formalismos individuais aplicados foram ARIMA, ETS, ANN, ELM, SVR e LSTM, e a combinação de preditores foi realizada por 6 métodos, a saber, cSA, cSM, cMV, cCP, cANN e cSVR. Foi utilizado o ambiente de programação estatística R (R CORE TEAM, 2022) para a implementação computacional. Todos resultados apresentados como tabelas e figuras foram elaborados em linguagem R. As etapas do estudo estão mostradas graficamente na Figura 7 e são descritas nas seções seguintes.

3.4 Pré-processamento dos dados

Os dados foram tratados quanto a eventuais valores ausentes. As falhas nos dados foram identificadas e preenchidas por meio de métodos univariados de imputação de séries temporais. Utilizando o pacote `imputeTS` (MORITZ; BARTZ-BEIELSTEIN, 2017b), o método da imputação com decomposição sazonal (`na_seadec`) foi aplicado, conforme apresentado na Seção 2.6.1.3. Este método tem sido aplicado em estudos semelhantes, a exemplo de Sahu *et al.* (2020).

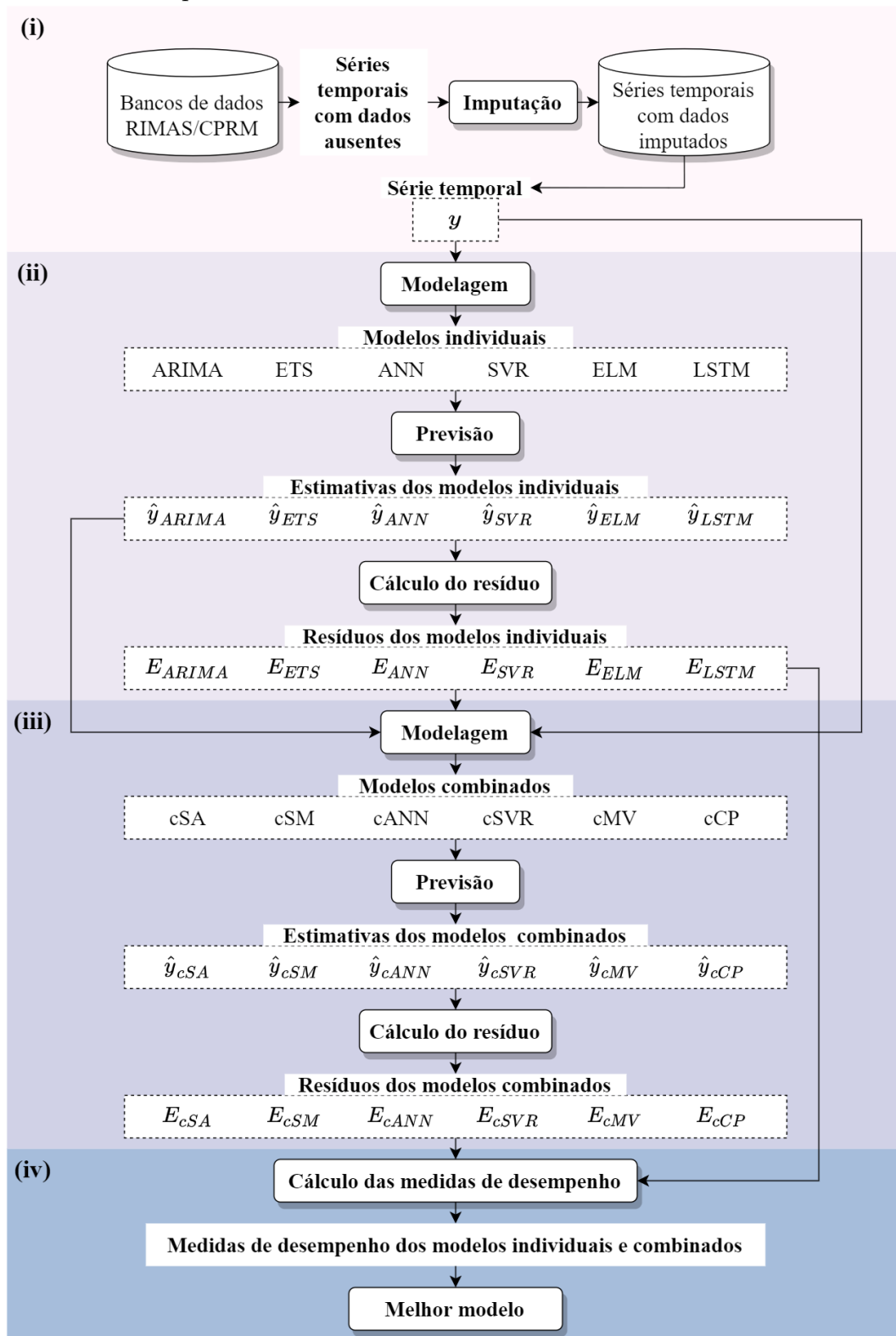
O conjunto de dados foi particionado em subconjuntos de treinamento e teste. O último ano de cada série foi dedicado ao conjunto de teste, correspondendo às últimas 12 observações, e os pontos restantes foram usados como conjunto de treinamento. Todos os modelos foram construídos utilizando-se o conjunto de treinamento, e o conjunto de testes foi usado para realizar previsões e avaliar os desempenhos dos modelos.

O conjunto de treinamento foi normalizado por meio de uma transformação do tipo Min-Max para o intervalo $[-1; 1]$, para se adequar às entradas dos modelos e minimizar o tempo de treinamento (AYDILEK; ARSLAN, 2013). A Equação 30 mostra a transformação aplicada.

$$z_t = \frac{y_t - \min(y)}{\max(y) - \min(y)} \times (\text{Max} - \text{Min}) + \text{Min} \quad (30)$$

em que z_t é o valor transformado no instante de tempo t , y_t o correspondente valor original, $\min(y)$ e $\max(y)$ são respectivamente os valores mínimo e máximo observados na série y , e Min e Max são os limites mínimo e máximo do novo intervalo, respectivamente (PATRO; SAHU, 2015). No presente estudo, adotou-se $\text{Min} = -1$ e $\text{Max} = 1$.

Figura 7 – Resumo do percurso metodológico adotado. (i) Pré-processamento, (ii) Modelagem individual, (iii) Modelagem combinada e (iv) Avaliação do desempenho



Fonte: Elaborado pelo autor (2022).

3.5 Modelagem individual

Os modelos ARIMA e ETS foram construídos usando pacote `forecast` (HYNDMAN; KHANDAKAR, 2008), através das funções `auto.arima` e `ets`, respectivamente. Para a seleção do modelo ARIMA, a função `auto.arima` usa o algoritmo proposto por Hyndman e Khandakar (2008), que combina testes de raiz unitária e o método da máxima verossimilhança para minimizar o Critério de Informação de Akaike corrigido (AICc) (HYNDMAN; ATHANASOPOULOS, 2018). Utilizou-se o parâmetro `nmodels = 500` como o número máximo de modelos ARIMA a ser considerado no algoritmo. Para o modelo ETS, a função `ets` seleciona o modelo mais apropriado para a série temporal segundo a minimização do AICc (HYNDMAN; ATHANASOPOULOS, 2018). O Critério de Informação de Akaike (AIC) e o AICc são apresentados na Seção 3.9.

Os pacotes usados para a implementação dos modelos baseados em aprendizado de máquina (ML) são descritos a seguir. ANN: `neuralnet` (FRITSCH *et al.*, 2019), SVR: `e1071` (MEYER *et al.*, 2019), ELM: `elmNMRcpp` (MOUSELIMIS, 2022), e LSTM: `keras` (ALLAIRE; CHOLLET, 2022) e `tensorflow` (ALLAIRE; TANG, 2022). A otimização dos parâmetros dos modelos ANN, SVR e ELM foi realizada através de algoritmos genéticos (GA), usando o pacote `GA` (SCRUCCA, 2013), conforme será descrito na Seção 3.7.

A camada de entrada dos modelos individuais ANN, ELM, LSTM e SVR foi constituída de forma similar à metodologia adotada por Hyndman e Khandakar (2008) para a construção de redes neurais para séries temporais, em que dois parâmetros são usados: o número de termos autorregressivos não sazonais, p , e não sazonais, P , análogos aos termos p e P dos modelos ARIMA e ARIMA sazonal, respectivamente. Assim, as entradas do modelo são os *lags* no intervalo entre 1 e p , e entre m e $m \times P$, em que m é a sazonalidade da série. A sazonalidade é dada em uma frequência fixa e conhecida. Por exemplo, para séries cujas observações são realizadas mensalmente, tem-se $m = 12$ (HYNDMAN; ATHANASOPOULOS, 2018).

Para o modelo ANN, os pesos das conexões entre os nós da rede neural foram otimizados usando o algoritmo de treinamento *backpropagation* resiliente (`'rprop+'`), com os parâmetros `stepmax = 1e + 6`, `threshold = 1e - 2` e `err.fct = 'sse'`. Quanto aos demais parâmetros da rede, foram otimizados através de GA o número de nós na camada de entrada (p e P), o número de camadas ocultas (nL), o número de nós nas camadas ocultas (nN) e a função de ativação (FA).

Para o modelo ELM, foram otimizados através de GA a camada de entrada (p e P), o número de nós na camada oculta (nN) e a função de ativação (FA). A Tabela A.2 no Apêndice A mostra um resumo das funções de ativação disponíveis nos pacotes R usados para a construção dos modelos ANN e ELM.

A SVR foi construída usando `type = 'eps-regression'`, `scale = FALSE` e *Radial Basis Function* (RBF) como função *kernel*. Foram otimizados através de GA a camada de entrada (p e P), e os parâmetros C e ε da SVR e γ do *kernel* RBF.

Para cada série temporal, foram gerados 2 modelos LSTM candidatos, usando as camadas de entrada (p e P) e oculta (nN) das redes ANN e ELM otimizados pelo GA, respectivamente. A seleção do melhor modelo LSTM se deu pela minimização da Raiz do Erro Quadrático Médio (RMSE), ou *Root Mean Squared Error*, em inglês, no conjunto de validação. As redes LSTM consistiram em uma camada LSTM seguida por uma camada do tipo *dense* com um nó de saída. Para a compilação da LSTM, adotou-se o Erro Quadrático Médio (MSE) ou *Mean Squared Error*, em inglês, como função de perda (`loss`), otimizador ADAM e um número de `epochs` igual a 400. As medidas de erro MSE e RMSE são apresentadas na Seção 3.9.

3.6 Modelagem combinada

As combinações por cSA e cSM levaram em consideração as previsões dos modelos individuais, calculando respectivamente a sua média aritmética e mediana.

Os combinadores cMV e cCP consideraram os resíduos dos modelos individuais. Para a inferência dos parâmetros da cópula, utilizou-se o método da Inferência por Marginais (IFM) (JOE; XU, 1996), que permite a inferência dos parâmetros das marginais separadamente para cada modelo individual. As cópulas adotadas foram Clayton, Frank, Gumbel e Normal. Para as marginais, optou-se pelas distribuições Normal, Normal assimétrica, Laplace e Cauchy. O melhor ajuste para cada marginal e para a conjunta foi realizado através da minimização do Critério de Informação Bayesiano (BIC), apresentado na Seção 3.9. Utilizou-se o pacote `copula` (HOFERT *et al.*, 2020).

Os modelos combinados cANN e cSVR foram construídos de maneira semelhante aos modelos individuais ANN e SVR, mas as camadas de entrada foram previamente definidas como as previsões dos modelos individuais. Foi usado GA para a otimização de parâmetros dos modelos combinados cANN e cSVR.

3.7 Otimização dos parâmetros

Para cada série temporal, o último ano do subconjunto de treinamento foi dedicado à etapa de validação, correspondendo a 12 observações. A função de aptidão dos algoritmos genéticos (GA) buscou a minimização do MSE no conjunto de validação. O GA foi executado com os seguintes parâmetros: população = 10, elitismo = 1, probabilidade de *crossover* = 0,90, probabilidade de mutação = 0,10, número de gerações = 50. Assim, realizando 50 iterações com uma população formada por 10 indivíduos, o GA construiu no total 500 modelos para cada formalismo cujos parâmetros foram otimizados. A Tabela 1 mostra os parâmetros selecionados para serem otimizados pelo GA em cada formalismo e os seus possíveis valores.

Tabela 1 – Parâmetros otimizados pelo algoritmo genético e os seus possíveis valores

Parâmetro	ANN	ELM	SVR	cANN	cSVR
p	[1, 5]	[1, 5]	[1, 5]	N/A	N/A
P	[1, 3]	[1, 3]	[1, 3]	N/A	N/A
nL	[1, 4]	N/A	N/A	[1, 4]	N/A
nN	[1, 40]	[1, 40]	N/A	[1, 40]	N/A
FA	logistic, tanh	sig, radbas, hardlim, hardlims, satlins, tansig, tribas, relu, purelin	sin, N/A	N/A	N/A
C	N/A	N/A	[1e-0, 1e+3]	N/A	[1e-0, 1e+3]
γ	N/A	N/A	[1e-3, 1e-0]	N/A	[1e-3, 1e-0]
ϵ	N/A	N/A	[1e-3, 1e-1]	N/A	[1e-3, 1e-1]

Nota: N/A indica que um modelo não possui dado parâmetro ou que o parâmetro não foi otimizado, [a, b] indica o intervalo de busca entre a e b. p, P – termos autorregressivos não sazonais e sazonais, respectivamente, nL – número de camadas ocultas, nN – número de nós nas camadas ocultas, FA – função de ativação.

Fonte: Elaborado pelo autor (2022).

3.8 Previsão

Na etapa de previsão, os modelos baseados em ML foram aplicados de forma iterativa, conforme descrevem Hyndman e Athanasopoulos (2018). Para a previsão um

passo adiante, todos os valores disponíveis da série são usados como entradas dos modelos. Para a previsão dois passos adiante, o valor previsto um passo a diante é usado como entrada nos modelos, juntamente com os valores da série. Este processo segue até que todas as previsões sejam realizadas. Esta estratégia tem sido adotada em estudos semelhantes, a exemplo de Sahu *et al.* (2020) e Müller *et al.* (2021). No presente estudo, esta abordagem iterativa foi usada para realizar previsões em um horizonte de 12 meses (conjunto de teste). Por fim, as previsões foram transformadas de volta ao intervalo original dos dados por meio da operação inversa à Equação (30), e o desempenho dos modelos foi avaliado.

3.9 Avaliação do desempenho

Diversas medidas podem ser usadas para a avaliação do desempenho de modelos de previsão de séries temporais (HYNDMAN; ATHANASOPOULOS, 2018). Cada métrica enfatiza um determinado aspecto do erro. Portanto, costuma-se utilizar um conjunto de diferentes métricas (CHAI; DRAXLER, 2014). No presente trabalho, a avaliação do desempenho dos modelos de previsão se deu por meio das seguintes métricas de erro: Erro Quadrático Médio (MSE), Raiz do Erro Quadrático Médio (RMSE), Erro Absoluto Médio (MAE), ou *Mean Absolute Error*, em inglês, Erro Percentual Absoluto Médio (MAPE), ou *Mean Absolute Percetile Error*, em inglês, Coeficiente de Determinação (R^2), estatística U de Theil e a medida *Prediction Of Change In Direction* (POCID), cujas expressões são mostradas nas Equações (31), (32), (33), (34), (35), (36), e (37), respectivamente.

$$\text{MSE} = \frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{n}, \quad (31)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{n}}, \quad (32)$$

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |\hat{y}_t - y_t|, \quad (33)$$

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{\hat{y}_t - y_t}{y_t} \right|, \quad (34)$$

$$U \text{ de Theil} = \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - y_{t-1})^2}, \quad (35)$$

$$R^2 = 1 - \frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{\sum_{t=1}^n (y_t - \bar{y})^2} \quad (36)$$

e

$$\text{POCID} = 100 \cdot \frac{\sum_{t=1}^n D_t}{n} \text{ onde } D_t = \begin{cases} 1, & \text{se } (y_t - y_{t-1})(\hat{y}_t - \hat{y}_{t-1}) > 0 \\ 0, & \text{caso contrário} \end{cases}, \quad (37)$$

em que \hat{y}_t é o valor previsto por um dado modelo, y_t o valor observado correspondente no instante de tempo t , e \bar{y} é a média dos valores observados.

O MSE e o RMSE avaliam a média dos erros quadráticos, enquanto o MAE calcula a média dos erros absolutos. O RMSE é calculado simplesmente como a raiz quadrada do MSE. O RMSE e o MAE são dados na mesma unidade de medida dos valores da série temporal em estudo. Por sua vez, o MAPE fornece uma medida da precisão do modelo em termos relativos. De uma maneira geral, quanto mais próximos de zero forem os valores de MSE, RMSE, MAE e MAPE, mais precisas serão as estimativas do modelo.

O coeficiente U de Theil mede o quanto o modelo se aproxima de um passeio aleatório (*random walk*). Caso U de Theil = 1 (U de Theil < 1), o modelo tem desempenho igual a (melhor que) um *random walk*. O R^2 mensura o quanto da variação dos dados é explicada pelo modelo. Assumindo valores no intervalo [0; 1], quanto mais próximo de 1 (0), melhor (pior) será o modelo. O POCID mede o percentual de acerto quanto às mudanças de direções na série. Quanto mais próximo de 100 (0) for o valor do POCID, melhor (pior) será o modelo.

As Equações 38, 39 e 40 mostram respectivamente o cálculo do Critério de Informação de Akaike (AIC), Critério de Informação de Akaike corrigido (AICc) e Critério de Informação Bayesiano (BIC), usados na construção dos modelos individuais ARIMA e ETS e do combinador cCP.

$$AIC = 2k - 2 \ln L^*, \quad (38)$$

$$AICc = AIC + \frac{2k^2 + 2k}{n - k - 1}, \quad (39)$$

$$BIC = k \ln n - 2 \ln L^*, \quad (40)$$

em que k é o número de parâmetros do modelo, L^* é o máximo valor observado da função de verossimilhança para o modelo estimado e n é o tamanho da amostra.

Em resumo, as métricas MSE, RMSE, MAE, MAPE, U de Theil, além dos critérios AIC, AICc e BIC seguem a regra de "quanto menor, melhor". Os valores de POCID e R^2 são analisados segundo a regra de "quanto maior, melhor".

A partir das medidas de desempenho (MD), foi possível classificar os modelos do melhor para o pior, em cada série temporal. Por exemplo, para a MD RMSE, o modelo que obteve o melhor (pior) valor recebeu classificação 1 (12). Realizada a classificação para cada MD, adotou-se a mediana como medida de tendência central para resumir a classificação considerando todas as MD. Por exemplo, para o modelo ARIMA, foram tomadas as suas classificações segundo cada MD e calculou-se a mediana das 6 classificações.

3.10 Implementação computacional

O experimento foi realizado em um Notebook com processador AMD Ryzen 5 3500U, RAM instalada 8,00 GB (utilizável: 5,88 GB), com sistema operacional de 64 bits, Windows 10 *Home Single Language*, e processador baseado em x64. O código-fonte implementado em linguagem R e utilizado para a realização dos experimentos do presente trabalho está disponível com livre acesso em repositórios *online* do autor, no endereço: <<https://github.com/rubensocj/dissertacao>>.

4 RESULTADOS E DISCUSSÃO

4.1 Resultados gerais

Os dados de níveis de água subterrânea (NAS) dos poços da bacia Sedimentar do Araripe (BSA) foram obtidos do portal RIMAS *Web* (RIMAS, 2021). Então, séries temporais mensais foram calculadas a partir das médias mensais dos valores de níveis d'água em cada poço. Todas as séries apresentaram valores ausentes, cujas porcentagens variaram de 4,0% (na série P11) a 35,9% (na série P17).

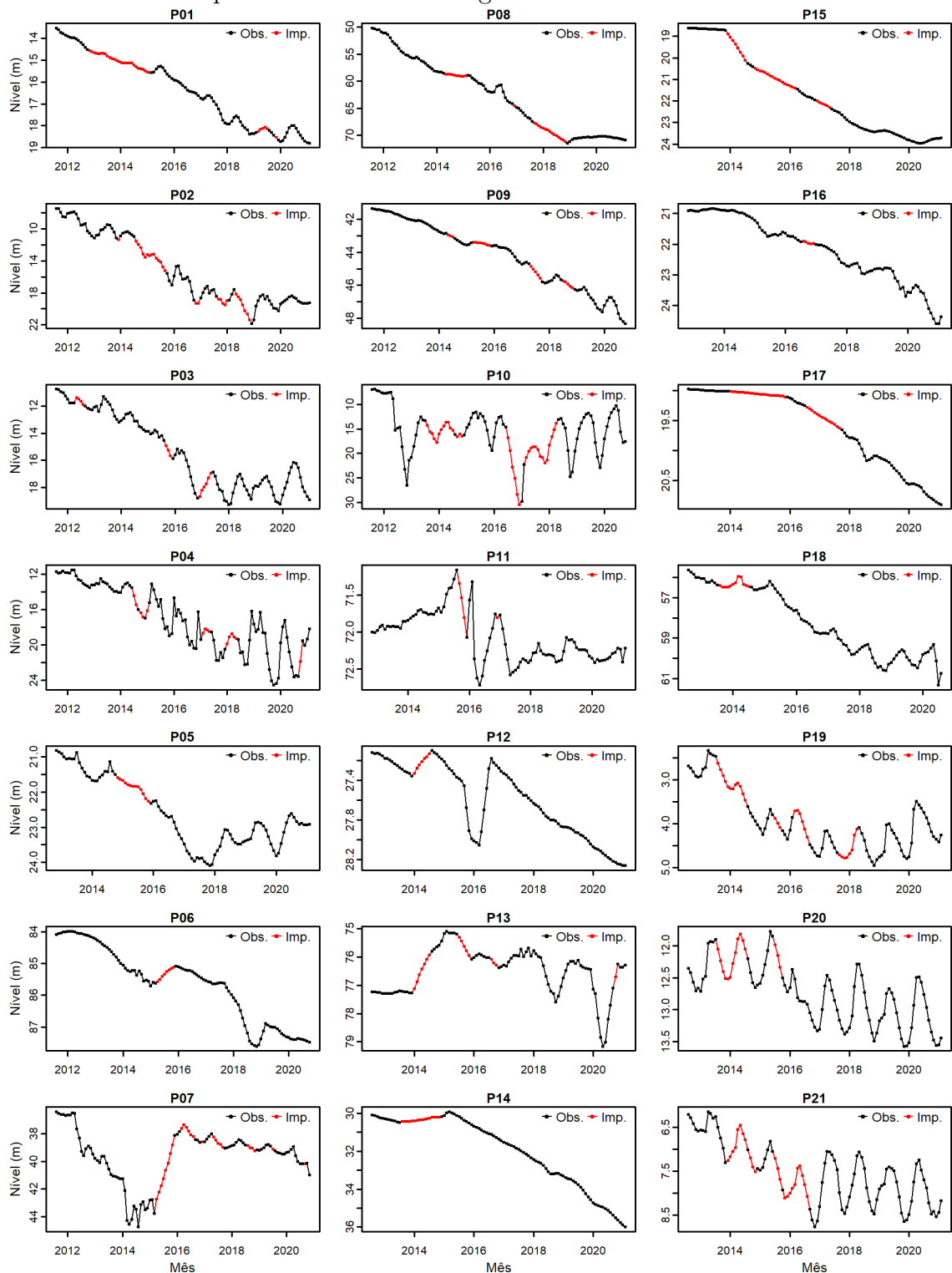
Segundo autores como Oikonomou *et al.* (2018), Asgharinia e Petroselli (2020) e He *et al.* (2020), o problema das falhas em dados de água subterrânea é comum até mesmo em locais onde há redes de monitoramento instaladas e em funcionamento. Os resultados obtidos neste trabalho mostram que isto também ocorre na BSA. As causas destas falhas na região de estudo devem ser investigadas.

Para a solução do problema das falhas nos dados, recorreu-se a um método de imputação. O preenchimento das falhas nas séries foi realizado pelo método univariado de imputação com decomposição sazonal. A Figura 8 mostra as séries mensais de NAS nos poços selecionados, obtidas após a etapa de preenchimento de falhas, com destaque para os valores imputados. As séries mostradas na Figura 8 são descritas a seguir, e foram utilizadas nas etapas de modelagem e previsão.

A Tabela 2 mostra as medidas descritivas dos dados. O número de observações (N) das séries variou de 97 a 115. O período de observação variou entre os poços, mas considerando todo o conjunto de dados, o estudo abrangeu o período de Agosto de 2011 a Fevereiro de 2021. Segundo as séries mensais analisadas, o nível estático (h) da água subterrânea na BSA variou de 2,3 m (na série P19) a 87,6 m (na série P06) nos diferentes poços analisados. O nível estático médio (\bar{h}) oscilou entre 3,9 m (P19) e 85,6 m (P06). Considerando o início e o final de cada série, a variação total do nível (Δ_h) mostrou que na maior parte dos poços houve diminuição do nível, que variou de 20,8 m (P08) a 0,2 m (P11). Apenas a série P13 apresentou uma elevação no nível total, correspondente a 0,9 m. Considerando as 21 séries, a diminuição média dos níveis foi de 5,1 m.

A região do Cariri, no Estado do Ceará, é abastecida em quase sua totalidade por água subterrânea. Por conta disso, o crescimento populacional e econômico na região nas últimas décadas têm levado a um aumento da exploração deste recurso (VERISSIMO,

Figura 8 – Gráficos das séries mensais de NAS na BSA, com destaque para os valores observados (cor preta) e imputados (cor vermelha). Diferentes períodos indicados nos gráficos



Fonte: Elaborado pelo autor (2022).

1999; SILVA *et al.*, 2021). Apesar da sua importância para a região, os sistemas de água subterrânea da BSA encontram-se em situação de vulnerabilidade (COGERH, 2009). Segundo o Boletim de Poços Monitorados no Cariri da Companhia de Gestão de Recursos

Tabela 2 – Medidas descritivas das séries mensais de NAS na BSA

Série	N	Início	Fim	$h_{m\acute{a}x}$	$h_{m\acute{i}n}$	Δ_h	\bar{h}	σ_h
P01	115	8/2011	2/2021	18,8	13,5	5,3	16,3	1,6
P02	115	8/2011	2/2021	21,8	7,5	11,8	15,1	4,2
P03	114	8/2011	1/2021	19,2	10,8	8,1	15,4	2,7
P04	115	8/2011	2/2021	24,5	11,5	6,5	17,1	3,6
P05	100	11/2012	2/2021	24,1	20,8	2,1	22,6	1,0
P06	111	8/2011	10/2020	87,6	84,0	3,4	85,6	1,1
P07	112	8/2011	11/2020	44,7	36,4	4,6	39,6	1,9
P08	115	8/2011	2/2021	71,5	50,1	20,8	62,8	6,7
P09	111	8/2011	10/2020	48,3	41,4	7,0	44,2	1,9
P10	111	8/2011	10/2020	30,4	6,9	10,7	15,9	4,8
P11	100	11/2012	2/2021	72,7	71,2	0,2	72,1	0,3
P12	103	8/2012	2/2021	28,3	27,1	1,1	27,6	0,4
P13	103	8/2012	2/2021	79,2	75,1	-0,9	76,5	0,8
P14	103	8/2012	2/2021	36,0	29,9	5,9	32,0	1,8
P15	103	8/2012	2/2021	24,0	18,6	5,1	21,7	1,8
P16	100	11/2012	2/2021	24,6	20,8	3,5	22,2	1,0
P17	103	8/2012	2/2021	20,9	19,0	1,9	19,6	0,6
P18	97	8/2012	8/2020	61,3	55,6	5,1	58,2	1,7
P19	103	8/2012	2/2021	5,0	2,3	1,6	3,9	0,7
P20	103	8/2012	2/2021	13,6	11,8	1,1	12,7	0,5
P21	103	8/2012	2/2021	8,8	6,1	2,0	7,5	0,7

Nota: N – número de observações, $h_{m\acute{a}x}$ – nível máximo (em metros), $h_{m\acute{i}n}$ – nível mínimo (em metros), Δ_h – variação total do nível (em metros), \bar{h} – nível médio (em metros), σ_h – desvio padrão (em metros).

Fonte: Elaborado pelo autor (2022).

Hídricos do Ceará (COGERH, 2019), observando-se o comparativo de 10 anos de monitoramento, de 2009 a 2019, a maior parte dos poços monitorados pelo órgão apresentaram diminuição, por vezes expressivas, em relação aos valores iniciais.

Um outro fator que pode estar associado à diminuição dos níveis identificada trata-se da seca extrema que ocorreu na região semiárida do Brasil no período de 2010 a 2017 (BURITI *et al.*, 2020). Causada por anomalias nas temperaturas superficiais nos oceanos Pacífico e Atlântico, a seca gerou impactos sociais, ambientais e econômicos para a região (SÃO JOSÉ *et al.*, 2021). Por outro lado, autores como Ashraf *et al.* (2021) apontam que as secas hidrológicas e meteorológicas podem impulsionar e intensificar o processo de diminuição dos níveis de água subterrânea.

Conforme mostrado na Figura 8, as séries apresentaram diferentes e complexos padrões, tais como tendência e variações sazonais. A tendência nas séries se manifesta em decorrência da diminuição total dos níveis observados nos poços, enquanto as variações

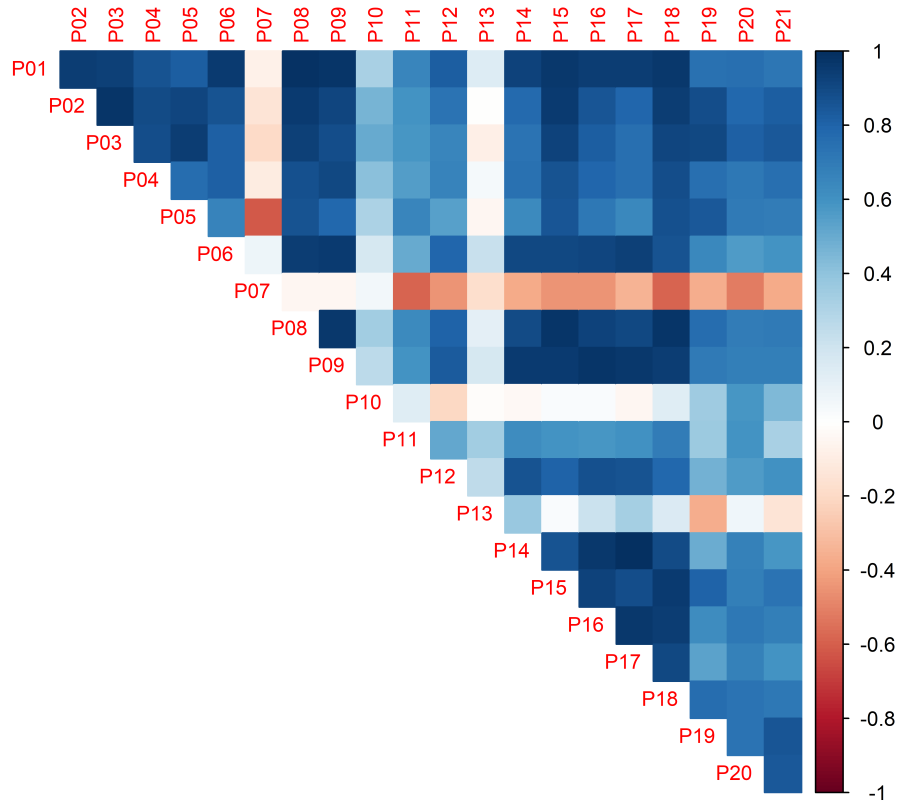
sazonais são devidas à recarga do lençol freático, sobretudo na estação chuvosa na região de estudo. Em geral, os diferentes padrões observados nas séries podem estar relacionados à distribuição espacial da recarga na área de estudo. Este complexo fenômeno sofre influência de diversos fatores, tais como precipitação, uso e ocupação do solo, taxas de infiltração, demanda pelo uso da água, contexto litoestratigráfico, entre outros (MANZIONE, 2015). Particularmente, Silva *et al.* (2021) destacam a distribuição espacial dos diferentes tipos de uso e ocupação do solo na região do Cariri cearense e a diversidade geomorfológica da BSA. Tais fatores podem estar relacionados às diferentes recargas e, por consequência, às diferentes variações sazonais observadas nas séries estudadas.

A Figura B.1 do Apêndice B mostra os gráficos da função de autocorrelação (FAC) e da função de autocorrelação parcial (FACP) das séries. A autocorrelação mede a relação linear entre valores defasados (*lags*) de uma série temporal (HYNDMAN; ATHANASOPOULOS, 2018). Os resultados obtidos mostram uma forte autocorrelação nas séries usadas neste estudo. De fato, é comum que as séries temporais de NAS apresentem elevada autocorrelação. Por conta disso, os valores passados da própria série tornam-se importantes variáveis preditoras para a modelagem (WUNSCH *et al.*, 2021). A análise dos gráficos da FAC e FACP fornece um indicativo dos *lags* mais relevantes para a modelagem da série (SAKIZADEH *et al.*, 2019). Contudo, diante da complexidade das flutuações dos NAS, é comum a adoção de técnicas de seleção de entradas dos modelos que considerem a natureza não linear das séries temporais, tais como a otimização por meio de algoritmos de busca estocástica (SAHU *et al.*, 2020; MÜLLER *et al.*, 2021).

Uma análise de correlação foi realizada para identificar a forma como as séries do conjunto de dados estão associadas entre si. Foi utilizado o coeficiente de correlação linear de Pearson (R) como métrica para a análise. A Figura 9 mostra graficamente a matriz de correlações entre as séries na forma de um correlograma. Identificou-se associação linear entre as séries analisadas. Os valores de R variaram de $-0,619$ a $0,996$, a mediana foi igual a $0,716$, e o 1º e 3º quartis foram respectivamente $0,365$ e $0,883$. A maior parte das correlações entre pares de séries (87,1%) foi positiva. A série P07 apresentou correlação linear negativa com a maioria das séries, enquanto que as séries P10 e P13 mostraram uma menor correlação linear com as demais séries.

A relação entre a distribuição espacial dos poços na área de estudo e correlação observada nos NAS também foi investigada. Para isto, foi construída a matriz de distâncias

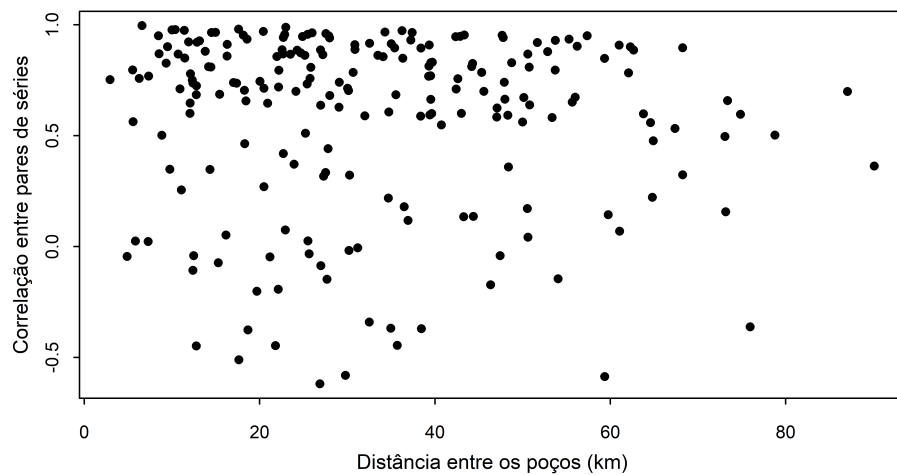
Figura 9 – Correlograma das séries mensais de NAS na BSA



Fonte: Elaborado pelo autor (2022).

dos poços analisados, mostrada na Tabela B.1 do Apêndice B. A Figura 10 mostra o gráfico de dispersão do coeficiente de Pearson R entre pares de séries (210 pares) e as respectivas distâncias entre os poços.

Figura 10 – Gráfico de dispersão dos coeficientes de correlação linear de Pearson R entre pares de séries mensais de NAS e as distâncias (em km) entre os poços na BSA



Fonte: Elaborado pelo autor (2022).

Calculou-se o coeficiente de correlação de postos de Spearman entre os pontos dispostos no gráfico da Figura 10, obtendo-se o valor igual a $\rho = -0,099$. A fim de verificar se a correlação observada entre os pontos do gráfico é estatisticamente significativa ou não, realizou-se o teste de associação entre amostras pareadas usando o coeficiente de correlação de Spearman, através da função `cor.test` em linguagem R. A hipótese nula é de que a correlação entre as variáveis não é significativa. Adotando um nível de significância igual a 5%, tem-se que para $p\text{-value} > 5\%$, não há motivos para se rejeitar a hipótese nula. Como resultados, obteve-se $p\text{-value} = 0,1505$. Portanto, a correlação entre os pares de séries e a distância entre os poços não estão significativamente correlacionadas, uma vez que $p\text{-value} > 5\%$. Isto sugere que as variações temporais dos NAS na BSA não estão associadas à distribuição espacial dos poços. Resultados semelhantes foram obtidos por Roshni *et al.* (2020). Utilizando uma abordagem de modelagem multivariada, os autores realizaram uma análise de agrupamento nos dados, a fim de identificar grupos de séries temporais que possuam características mais similares entre si, em vez de técnicas baseadas em vizinhos mais próximos.

4.2 Etapa de modelagem

As séries temporais de níveis de água subterrânea (NAS) foram usadas para a construção dos modelos individuais e combinados, e para a realização das previsões. Para a etapa de modelagem, realizou-se uma partição em conjuntos de treinamento e teste, de modo que o conjunto de teste correspondeu ao último ano de cada série. A Tabela 3 mostra a descrição dos conjuntos de treinamento, validação e teste, com destaque para o número de observações e o período a que se referem. O conjunto de validação foi usado para a otimização dos parâmetros dos modelos ANN, SVR, ELM, cANN e cSVR. Considerando todo o conjunto de dados, o número de observações disponíveis para a etapa de modelagem variou de 73 a 91. O tamanho do conjunto de treinamento variou de 75% a 79%, e o tamanho dos conjuntos de validação e teste variou entre 10% e 12%.

Foi gerado um total de 252 modelos, sendo 12 (6 individuais e 6 combinados) para cada uma das 21 séries. A seguir, são apresentados e descritos os modelos construídos. Os modelos combinados cSA e cSM não necessitam de etapa de modelagem, portanto, não são considerados nos resultados que seguem. As Tabelas C.1 a C.21, no Apêndice C, mostram a descrição dos modelos individuais e combinados construídos para cada

Tabela 3 – Etapas de modelagem e previsão para as séries temporais (ST)

ST	N_{Tr}	$\%_{Tr}$	Tr	N_V	$\%_V$	V	N_{Ts}	$\%_{Ts}$	Ts
P01	91	79	08/2012-02/2019	12	10	03/2019-02/2020	12	10	03/2020-02/2021
P02	91	79	08/2012-02/2019	12	10	03/2019-02/2020	12	10	03/2020-02/2021
P03	90	79	08/2012-01/2019	12	11	02/2019-01/2020	12	11	02/2020-01/2021
P04	91	79	08/2012-02/2019	12	10	03/2019-02/2020	12	10	03/2020-02/2021
P05	76	76	11/2013-02/2019	12	12	03/2019-02/2020	12	12	03/2020-02/2021
P06	87	78	08/2012-10/2019	12	11	11/2019-10/2020	12	11	11/2020-10/2021
P07	88	79	08/2012-11/2019	12	11	12/2019-11/2020	12	11	12/2020-11/2021
P08	91	79	08/2012-02/2019	12	10	03/2019-02/2020	12	10	03/2020-02/2021
P09	87	78	08/2012-10/2019	12	11	11/2019-10/2020	12	11	11/2020-10/2021
P10	87	78	08/2012-10/2019	12	11	11/2019-10/2020	12	11	11/2020-10/2021
P11	76	76	11/2013-02/2019	12	12	03/2019-02/2020	12	12	03/2020-02/2021
P12	79	77	08/2013-02/2019	12	12	03/2019-02/2020	12	12	03/2020-02/2021
P13	79	77	08/2013-02/2019	12	12	03/2019-02/2020	12	12	03/2020-02/2021
P14	79	77	08/2013-02/2019	12	12	03/2019-02/2020	12	12	03/2020-02/2021
P15	79	77	08/2013-02/2019	12	12	03/2019-02/2020	12	12	03/2020-02/2021
P16	76	76	11/2013-02/2019	12	12	03/2019-02/2020	12	12	03/2020-02/2021
P17	79	77	08/2013-02/2019	12	12	03/2019-02/2020	12	12	03/2020-02/2021
P18	73	75	08/2013-08/2019	12	12	09/2019-08/2020	12	12	09/2020-08/2021
P19	79	77	08/2013-02/2019	12	12	03/2019-02/2020	12	12	03/2020-02/2021
P20	79	77	08/2013-02/2019	12	12	03/2019-02/2020	12	12	03/2020-02/2021
P21	79	77	08/2013-02/2019	12	12	03/2019-02/2020	12	12	03/2020-02/2021

Nota: N_{Tr} , N_V e N_{Ts} – número de observações nos conjuntos de treinamento (Tr), validação (V) e teste (Ts), respectivamente. $\%_{Tr}$, $\%_V$ e $\%_{Ts}$ – porcentagem dedicada aos conjuntos Tr , V e Ts , respectivamente.

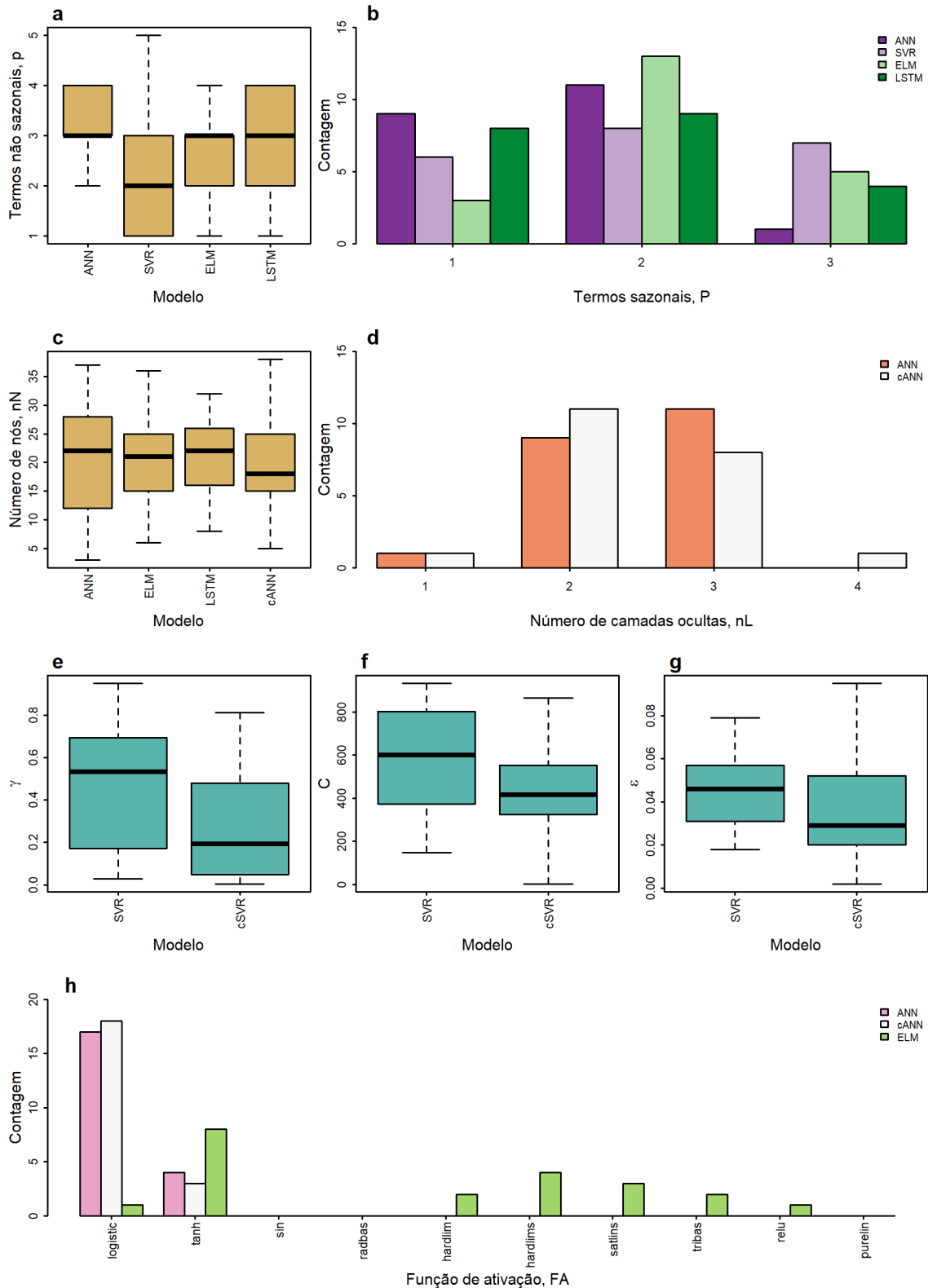
Fonte: Elaborado pelo autor (2022).

série temporal. A Figura 11 mostra graficamente o resumo dos parâmetros dos modelos individuais e combinados baseados em ML.

Considerando todo o conjunto de dados, os modelos ANN individuais foram gerados usando de 2 a 4 termos autorregressivos não sazonais, p . Quanto aos termos sazonais, P , a maior parte dos modelos ANN usou 1 (em 9 séries) ou 2 (em 11 séries) termos, e apenas em uma série foram selecionados 3 termos. O número de nós nas camadas ocultas, nN , da ANN variou entre 3 e 37. Na maioria das séries, o modelo ANN foi gerado usando 2 (em 9 séries) ou 3 (em 11 séries) camadas ocultas, nL . Apenas em uma série, usou-se uma única camada oculta na ANN. Em 17 séries, a função de ativação (FA), selecionada foi a logística, enquanto em 4 séries a função tangente hiperbólica foi usada.

Quanto ao modelo combinado cANN, o parâmetro nN variou de 5 a 38 nós. Os valores de nL variaram de 1 a 4, mas concentraram-se entre 2 ou 3 camadas ocultas. No tocante à FA, prevaleceu a função logística (em 18 séries), frente a \tanh (em 3 séries).

Figura 11 – Parâmetros dos modelos baseados em ML considerando todas as séries. a. termos autorregressivos não sazonais (p), b. termos autorregressivos sazonais (P), c. número de nós nas camadas ocultas (nN), d. número de camadas ocultas (nL), e. f. g. parâmetros γ , C e ε da SVR, respectivamente, h. função de ativação (FA)



Fonte: Elaborado pelo autor (2022).

Comparativamente, os parâmetros nN , nL e FA dos modelos ANN e cANN apresentaram-se distribuídos de forma semelhante.

Em relação ao modelo individual SVR, considerando todas as séries analisadas, o parâmetro p variou entre 1 e 5, enquanto o valor de P apresentou-se variando de 1 a 3. Os valores mínimo e máximo selecionados para o parâmetro γ foram 0,028 e 0,95, respectivamente. Quanto ao custo, C , este apresentou-se em uma faixa de valores de 147,71 a 932,74. Por fim, o parâmetro ε variou de 0,018 a 0,079.

A distribuição dos parâmetros γ , C e ε do modelo combinado cSVR no conjunto de dados ocorreu de maneira semelhante à do modelo individual SVR. Para o modelo cSVR, o parâmetro γ variou de 0,005 a 0,812, o parâmetro C apresentou-se variando entre 2,665 e 865,62, e os valores mínimo e máximo de ε foram respectivamente 0,002 e 0,095.

Para o modelo individual ELM, considerando todas as séries avaliadas, o parâmetro p variou de 1 a 4, de modo que 10 séries foram modeladas usando 3 termos não sazonais. O parâmetro P variou de 1 a 3, concentrando-se em 2 termos sazonais (em 13 séries). Os valores mínimo e máximo de nN foram, respectivamente, 6 e 36 nós na camada oculta única dos modelos ELM. Quanto à FA do modelo ELM, a função tangente hiperbólica foi selecionada em 8 séries e a logística em apenas 1. Por outro lado, em 12 séries, a ELM foi gerada usando funções de ativação lineares, tais como `hardlim`, `hardlims`, `relu`, `satlins` e `tribas`.

Em relação ao modelo individual LSTM, levando em consideração todas as séries temporais, o parâmetro p variou de 1 a 4 termos não sazonais, os valores de P observados estiveram entre 1 e 3 termos sazonais, enquanto o parâmetro nN variou de 8 a 32 nós. Ao se comparar os modelos LSTM candidatos, não houve diferença significativa entre os modelos LSTM construídos com os parâmetros p , P e nN fornecidos por ANN ou ELM. A LSTM foi construída em 11 séries com os parâmetros da ANN e em 10 séries com os parâmetros da ELM.

No geral, o modelo ARIMA sazonal foi selecionado em 11 séries, enquanto o modelo ARIMA não sazonal foi usado em 10 séries. A ordem p variou de 0 (em 8 séries) a 5 (em 1 série), e a ordem P apresentou-se variando de 0 (em 15 séries) a 2 (em 4 séries). Quanto ao modelo ETS, a componente de erro (E) aditivo foi usada em todas as séries. A tendência (T) selecionada foi do tipo aditiva (em 6 séries), *damped* aditiva (em 8 séries), e em 7 séries o modelo ETS não adotou a componente de tendência. A sazonalidade (S) foi

incorporada em 7 séries, sendo do tipo aditiva, enquanto que em 14 séries, a componente de sazonalidade não foi usada.

Considerando todos os modelos construídos, os pesos atribuídos pelo combinador cMV ao modelo individual ARIMA variaram entre -0,260 a 6,667. Os pesos do preditor ETS assumiram valores entre -5,889 e 0,329. Para o modelo ANN, os valores mínimo e máximo assumidos pelos pesos foram -0,260 e 0,972, respectivamente. Os pesos do preditor SVR variaram entre -0,222 e 0,965. Para a ELM, os pesos mínimo e máximo foram, respectivamente, -0,346 e 0,699. Por fim, os pesos atribuídos pelo combinador cMV ao preditor individual LSTM variaram entre -0,319 e 0,806.

Quanto ao modelo cCP, as cópulas selecionadas foram Normal (em 18 séries) e Clayton (em 3 séries). Em relação às distribuições marginais dos erros, foi modelado um total de 126 séries de erros, sendo 6 (referentes aos 6 modelos individuais) para cada uma das 21 séries. Destes, 51% foi modelado segundo a distribuição Laplace (64 casos), 24% segundo a distribuição Normal (30 casos) e 13% (16 casos) segundo as distribuições Cauchy e Normal assimétrica, respectivamente.

A duração da etapa de modelagem também foi avaliada. A Tabela 4 mostra os tempos decorridos para a modelagem, medidos em segundos, para as séries temporais. Os melhores resultados estão destacados em negrito. Os tempos na Tabela 4 consideram os diferentes tipos de modelagem. Para ARIMA e ETS, mediu-se o tempo necessário para a seleção do modelo usando o algoritmo de Hyndman e Khandakar (2008). As medidas para ANN, SVR, ELM, cANN e cSVR são o tempo decorrido para a otimização dos parâmetros com GA. Quanto a LSTM, cMV e cCP, os valores avaliados são o tempo necessário para a construção de um único modelo. Os tempos para os modelos combinados não consideram o tempo necessário para a construção dos modelos individuais, porque medem apenas a duração da etapa de modelagem combinada, que ocorre após a etapa de modelagem individual e usa as previsões e resíduos dos modelos individuais.

Entre os modelos individuais, o modelo ETS obteve o menor tempo de modelagem na maioria das séries (17 séries), seguido pelo modelo ARIMA (4 séries). Quanto aos modelos combinados, o modelo cMV obteve o menor tempo de modelagem em todas as séries. Considerando todos os modelos construídos, o menor tempo de modelagem observado foi para a construção do modelo cMV nas séries P04, P09, P15 e P18 (0 s), e o maior tempo foi para a otimização da ANN com GA na série P04 (359,6711 s).

Tabela 4 – Tempos de modelagem (em segundos) dos modelos individuais e combinados para as séries temporais (ST). Os melhores resultados em cada caso estão destacados em negrito

ST	ARIMA	ETS	ANN	SVR	ELM	LSTM	cANN	cSVR	cMV	cCP
P01	2,104	0,616	83,929	17,769	29,475	177,078	32,138	7,284	0,004	6,794
P02	6,005	0,504	155,361	15,348	14,345	210,602	45,270	10,626	0,002	6,090
P03	2,979	0,647	124,218	24,303	14,660	194,295	40,828	6,038	0,004	4,528
P04	0,461	0,610	359,671	19,099	11,539	157,621	49,055	4,367	0,000	2,461
P05	0,273	0,326	20,418	8,741	6,177	91,675	25,331	2,867	0,001	1,631
P06	2,810	0,482	52,447	17,189	12,872	145,833	51,585	6,897	0,004	5,675
P07	0,936	0,481	175,470	18,274	12,596	157,230	45,102	5,856	0,004	5,799
P08	0,840	0,564	75,979	16,378	13,208	151,750	26,620	5,497	0,002	5,595
P09	0,220	0,484	64,033	18,122	13,614	178,153	23,126	6,132	0,000	5,167
P10	7,913	0,362	96,176	10,157	7,383	101,023	34,370	3,085	0,001	1,809
P11	0,501	0,297	88,311	10,366	7,696	97,627	62,660	6,103	0,001	1,611
P12	1,620	0,318	68,208	9,393	8,603	85,488	26,581	4,445	0,001	2,924
P13	0,479	0,283	84,793	9,342	7,802	85,979	38,581	4,006	0,004	3,815
P14	7,359	0,520	43,922	12,814	13,103	156,263	19,681	5,571	0,004	4,982
P15	0,202	0,561	62,745	14,315	15,444	174,989	39,654	8,081	0,000	2,483
P16	0,688	0,335	62,538	8,967	6,281	96,626	16,381	2,934	0,007	3,948
P17	0,395	0,314	21,145	7,363	7,339	85,903	20,667	3,265	0,001	1,750
P18	4,285	0,502	54,574	16,814	14,149	115,600	19,799	3,843	0,000	3,570
P19	1,920	0,477	91,376	27,134	9,048	124,232	42,440	6,330	0,004	4,141
P20	4,951	0,319	153,087	9,588	13,106	154,596	28,472	5,724	0,004	4,443
P21	1,838	0,446	93,439	15,601	11,900	118,644	92,801	7,716	0,001	3,121
Mín.	0,202	0,283	20,418	7,363	6,176	85,488	16,381	2,867	0,000	1,611
μ	2,323	0,450	96,754	14,623	11,921	136,248	37,197	5,556	0,002	3,921
Máx.	7,913	0,647	359,671	27,134	29,475	210,601	92,801	10,627	0,007	6,795

Nota: O tempo medido considera os diferentes critérios para construção dos modelos: ANN, SVR, ELM, cANN e cSVR – otimização de parâmetros com GA. ARIMA e ETS – seleção do modelo usando o algoritmo de Hyndman e Khandakar (2008). LSTM, cMV e cCP – construção de um único modelo. Mín – mínimo, μ – média, Máx. – máximo.

Fonte: Elaborado pelo autor (2022).

Os modelos LSTM, ANN e cANN apresentaram, em média, os maiores tempos de modelagem, cujos valores foram respectivamente 136,248 s, 96,754 s e 37,197 s. Destaque-se que o tempo da LSTM se refere à construção de um único modelo, enquanto os valores de ANN e cANN são o tempo decorrido para a otimização com GA. Autores como Müller *et al.* (2021) ressaltam o custo computacional para treinamento de redes LSTM. Além disso, modelos baseados em ML, como a LSTM, são sensíveis à escolha dos parâmetros (SAHU *et al.*, 2020). De fato, no presente trabalho, a LSTM foi testada com diferentes configurações. Para o conjunto de dados usado, o valor de 400 *epochs* mostrou-se satisfatório para a construção de redes LSTM com um bom desempenho nas etapas de modelagem e previsão.

Contudo, este valor elevou o tempo de treinamento e tornou inviável computacionalmente a otimização dos parâmetros do modelo LSTM por meio de GA.

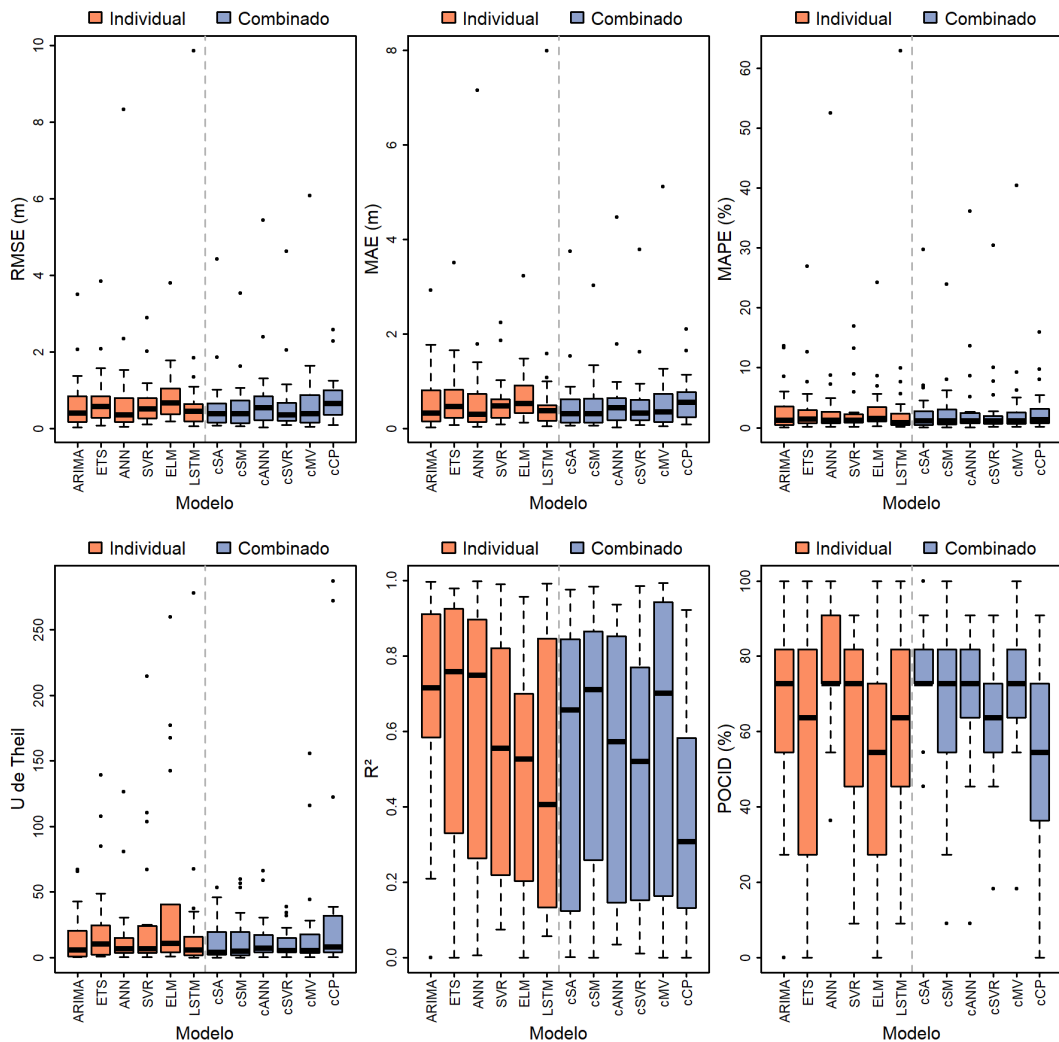
4.3 Etapa de previsão

Os modelos construídos foram usados para realizar previsões a longo prazo dos níveis de água subterrânea (NAS) na bacia Sedimentar do Araripe (BSA). As previsões foram realizadas considerando um horizonte de tempo que não foi usado na etapa de modelagem, para avaliar a capacidade de generalização dos modelos construídos. O conjunto de testes consistiu no último ano de cada série (12 observações). As Figuras D.1 a D.21, no Apêndice D, mostram os gráficos das previsões realizadas para todas as séries de NAS. Os modelos foram capazes de prever o comportamento do nível da água subterrânea na BSA. As medidas de desempenho (MD) RMSE, MAE, MAPE, U de Theil, R^2 e POCID foram usadas para avaliar os modelos. Os resultados das previsões para cada série temporal estão mostrados nas Tabelas E.1 a E.21, no Apêndice E.

A Figura 12 mostra graficamente a distribuição das medidas de desempenho dos modelos individuais e combinados no conjunto de testes considerando todas as séries. Os valores de RMSE variaram de 0,0246 m (ARIMA na série P17) a 9,8635 m (LSTM na série P10), com média de 0,7767 m. Quanto ao MAE, os valores mínimo e máximo observados foram respectivamente 0,0194 m (cANN na série P12) e 7,9835 m (LSTM na série P10), com média igual a 0,6466 m. Em relação ao MAPE, os valores variaram de 0,0301% (ARIMA na série P06) a 62,8512% (LSTM na série P10), sendo a média igual a 3,5908%. Os valores de U de Theil mínimo e máximo observados foram respectivamente 0,1031 (LSTM na série P20) e 287,2939 (cCP na série P06), enquanto a média foi 23,846. Os valores de R^2 se apresentaram variando entre 0 e 0,9984 (ANN na série P01), e a média foi igual a 0,542. Quanto ao POCID, os valores mínimo e máximo observados foram 0% e 100%, com um valor médio igual a 63,9623%.

Após a avaliação do desempenho, os modelos foram classificados segundo medida de desempenho. As Tabelas E.22 a E.42 do Apêndice E mostram a classificação dos modelos no conjunto de teste segundo medida de desempenho para todas as séries. Então, calculou-se a mediana das classificações considerando todas as medidas de desempenho. A Tabela 5 mostra a classificação mediana dos modelos, com destaque para o melhor modelo em cada série e o número de séries em que um dado modelo foi o melhor.

Figura 12 – Distribuição das medidas de desempenho dos modelos individuais (cor laranja) e combinados (cor azul) no conjunto de testes considerando todas as séries



Fonte: Elaborado pelo autor (2022).

Os modelos combinados foram superiores em relação aos individuais em 8 séries temporais, correspondendo a 38% do conjunto de dados. Por outro lado, os modelos individuais superaram os combinados em 13 séries temporais, equivalendo a 62% dos casos avaliados. Os modelos LSTM e ARIMA obtiveram as melhores classificações entre os individuais, enquanto os modelos cSA, cANN e cMV foram os melhores classificados entre os combinados. Considerando todas as técnicas, os modelos LSTM e ARIMA obtiveram a melhor classificação, seguido pelos modelos cSA, cANN e cMV. O modelo ELM foi o pior classificado entre os individuais, e o modelo cCP foi o pior entre os combinados. No geral, os modelos ELM e cCP mostraram as piores classificações.

O modelo ETS apresentou desempenho relativamente ruim em algumas séries. Por outro lado, apesar de ter sido o melhor classificado, o modelo ARIMA também mostrou

Tabela 5 – Classificação dos modelos para as séries temporais (ST) de NAS na BSA. Os melhores resultados estão destacados em negrito

ST	ARIMA	ETS	ANN	SVR	ELM	LSTM	cSA	cSM	cANN	cSVR	cMV	cCP
P01	11,0	12,0	2,0	5,5	9,0	1,0	4,5	7,0	3,0	4,0	8,0	10,0
P02	11,0	4,5	1,0	3,0	10,0	10,5	7,0	6,0	4,0	2,0	8,0	8,0
P03	9,5	8,0	3,0	9,5	10,5	1,0	2,0	5,0	5,5	8,0	3,0	7,0
P04	12,0	6,5	10,0	8,5	3,0	4,5	4,0	1,0	9,5	7,0	2,0	8,0
P05	11,0	12,0	8,0	1,0	9,0	2,0	5,5	9,5	3,5	7,0	4,5	4,5
P06	1,0	7,0	2,0	10,0	9,0	11,0	4,0	5,0	6,0	3,0	7,5	12,0
P07	2,0	5,5	7,0	8,5	8,0	1,0	3,0	5,0	10,0	3,0	12,0	11,0
P08	3,0	7,0	9,0	8,0	11,0	4,0	2,0	6,0	5,0	1,0	10,0	12,0
P09	3,5	11,0	7,0	9,0	12,0	1,0	5,0	2,0	6,0	7,0	3,5	10,0
P10	1,0	6,0	11,0	3,0	5,0	12,0	7,0	4,0	9,0	8,0	10,0	2,0
P11	1,0	4,0	9,0	5,0	11,0	6,0	2,0	3,0	12,0	10,0	7,0	8,0
P12	8,0	5,0	2,0	10,0	12,0	4,0	9,0	7,0	1,0	6,0	3,0	11,0
P13	10,0	11,5	11,0	6,0	4,0	5,0	1,0	3,0	4,5	3,5	8,0	8,5
P14	3,0	2,0	7,0	11,0	10,0	6,0	4,5	4,5	8,5	8,0	1,0	12,0
P15	10,0	11,0	7,0	5,0	12,0	8,0	4,0	6,0	1,0	2,0	9,0	3,0
P16	8,0	1,0	2,0	9,0	11,0	5,5	5,5	4,5	10,0	3,5	7,0	12,0
P17	1,0	7,0	11,0	10,0	12,0	3,0	8,0	9,0	4,0	6,0	2,0	5,0
P18	1,0	4,0	8,5	8,0	6,0	5,0	2,0	4,0	12,0	7,0	9,0	10,5
P19	3,5	10,0	3,0	10,5	5,0	8,5	1,0	3,0	10,5	7,0	5,5	7,5
P20	9,5	11,5	9,0	5,0	10,0	1,0	3,0	2,0	8,0	4,0	7,0	6,0
P21	8,0	9,5	8,0	3,5	11,0	6,5	3,5	2,0	5,5	6,0	1,0	11,0
N1	5	1	1	1	0	5	2	1	2	1	2	0

Nota: N1 – contagem do número de vezes em que um dado formalismo foi superior aos outros, considerando todas as séries temporais do conjunto de dados.

Fonte: Elaborado pelo autor (2022).

resultados ruins em alguns casos. Isto pode ser devido às limitações de tais modelos lineares em lidar com padrões não lineares comumente apresentados por séries temporais de NAS, conforme apontam Zanotti *et al.* (2019).

O bom desempenho do modelo ANN em algumas séries pode estar relacionado ao uso de redes com múltiplas camadas ocultas. Conforme mostra a Figura 11.d, em muitos casos o modelo ANN foi construído usando 2 ou 3 camadas ocultas. Estratégias semelhantes têm obtido resultados atraentes em estudos de Sahu *et al.* (2020).

O modelo SVR com *kernel Radial Basis Function* (RBF) usado neste trabalho mostrou-se satisfatório em muitas séries temporais. Esta configuração de SVR também têm sido empregada por outros hidrogeologistas, a exemplo de Bozorg-Haddad *et al.* (2019) e Mozaffari *et al.* (2022). Os autores destacam ainda que o *kernel* RBF é o mais usado na literatura de modelagem e previsão de NAS.

O desempenho do modelo ELM pode ter sido influenciado pela função de ativação (FA) usada. Particularmente, os modelos ELM que usaram as funções `hardlim` e `hardlims` merecem destaque. Nestes casos, percebeu-se um comportamento insatisfatório das previsões nos conjuntos de treinamento e teste. O modelo ELM com tais funções lineares não foi capaz de modelar os padrões não lineares mostrados nas séries P04, P08, P09, P11, P16 e P17. Embora a FA tenha sido selecionada pelo GA, a ELM construída pode não ter adquirido boa capacidade de generalização. Isto é, apesar de mostrar o melhor desempenho no conjunto de validação, o modelo selecionado não performou de maneira satisfatória no conjunto de teste. Em estudos semelhantes, Malekzadeh *et al.* (2019) usaram uma abordagem baseada em tentativa-e-erro para definir a FA da ELM, entre alternativas lineares e não lineares, sendo a função logística a que obteve melhores resultados. Em alguns trabalhos, a exemplo de Wu *et al.* (2021) e Azizpour *et al.* (2021), não há detalhes sobre qual FA foi usada.

Em trabalhos recentes sobre a previsão de séries temporais de NAS, os pesquisadores têm adotado diferentes estratégias para a seleção da arquitetura das redes LSTM. Por exemplo, Müller *et al.* (2021) usaram algoritmos de otimização, enquanto Solgi *et al.* (2021) seguiram uma abordagem baseada em tentativa-e-erro. Diante dos resultados obtidos no presente trabalho, a estratégia adotada para a seleção do modelo LSTM mostrou-se satisfatória para o conjunto de dados usado. De fato, embora seja relativamente mais custosa computacionalmente na etapa de modelagem, a LSTM superou os demais modelos em muitos casos.

Apesar de ser o combinador relativamente mais simples, o modelo cSA foi capaz de reduzir a variação das estimativas dos modelos individuais na série P19. Conforme destacam Hajirahimi e Khashei (2019), este modelo combinado simples pode apresentar bom desempenho e superar combinadores mais complexos em alguns casos. Por outro lado, em estudos semelhantes de Najafabadipour *et al.* (2022), os combinadores cSA e cSM foram superados por outros modelos combinados lineares.

Os modelos combinados cANN e cSVR foram superiores particularmente em séries que mostraram apenas padrão de tendência no conjunto de teste. Apesar disso, tais modelos não foram capazes de atribuir pesos aos preditores individuais de forma satisfatória em outras séries. Por sua vez, o modelo cMV foi capaz de melhorar o desempenho dos modelos em séries que apresentaram tendência ou sazonalidade. Quanto ao combinador

cCP, o seu desempenho pode estar relacionado ao método adotado para inferência dos parâmetros da cópula e ao número limitado de cópulas investigado.

De uma maneira geral, percebeu-se que a precisão das previsões decaiu conforme aumentou-se o horizonte de previsões. Isto é, nos últimos meses do conjunto de teste, as previsões dos modelos distanciaram-se dos valores observados na série. Outro aspecto que pode ter influenciado o desempenho dos modelos é o fato de algumas séries mostrarem comportamentos diferentes no conjunto de teste, em comparação com o conjunto de treinamento. Além disso, o número de observações das séries temporais destinadas à etapa de modelagem pode ter afetado o treinamento dos modelos. Por fim, o resultado ruim de alguns dos modelos individuais pode ter afetado o desempenho dos modelos combinados em algumas séries temporais.

5 CONSIDERAÇÕES FINAIS

O presente trabalho desenvolveu modelos individuais e combinados de previsão de séries temporais de níveis de água subterrânea (NAS) na bacia Sedimentar do Araripe (BSA). Os dados de água subterrânea coletados apresentaram falhas, que foram preenchidas por meio de métodos de imputação de valores ausentes, para a obtenção de um conjunto de dados completo. As séries temporais estudadas exibiram diferentes e complexos padrões. A maior parte dos poços considerados apresentou diminuição expressiva dos níveis de água subterrânea. Isto pode estar associado à seca que atingiu a região no período de estudo e à vulnerabilidade, em termos de quantidade, à que estão submetidos os recursos hídricos subterrâneos da região do Cariri, Ceará.

Os modelos individuais e combinados construídos foram capazes de prever o comportamento das flutuações freáticas na área de estudo. A avaliação do desempenho dos modelos mostrou que as alternativas de modelagem combinada superaram as estratégias individuais em muitos casos (8 séries). Por outro lado, os modelos individuais obtiveram melhores classificações em 13 séries, com destaque para o modelo autorregressivo integrado de médias móveis (ARIMA) e rede neural *long short-term memory* (LSTM). Diante disso, os modelos combinados de previsão de séries temporais de níveis de água subterrânea propostos no presente trabalho podem ser aprimorados.

5.1 Limitações

São limitações do presente estudo o número de observações das séries temporais estudadas, bem como o tamanho do conjunto de validação dedicado para a otimização dos parâmetros dos modelos baseados em aprendizado de máquina. O elevado custo computacional da etapa de modelagem da rede LSTM em comparação aos demais modelos consistiu em uma dificuldade para a execução do trabalho.

5.2 Trabalhos futuros

Como trabalhos futuros, deve-se investigar o uso de outros modelos individuais do estado-da-arte da modelagem e previsão de séries temporais, a exemplo das redes neurais dos tipos *transformer* e *N-beats*. Para a combinação, surge o interesse em aprimorar os

modelos combinados, por meio da seleção dos modelos individuais que resultem no melhor desempenho dos combinadores.

É interessante investigar o desenvolvimento de modelos combinados de previsão de séries de níveis de água subterrânea que incorporem variáveis exógenas que influenciem as variações dos níveis freáticos, tais como precipitação, temperatura, evapotranspiração, entre outras. Ainda, a adoção de uma abordagem de previsão múltiplos passos adiante e a aplicação de modelos combinados para a previsão de níveis diários devem ser avaliadas.

REFERÊNCIAS

- ABDELKARIM, B.; TELAHIGUE, F.; AGOUBI, B. Assessing and delineation of groundwater recharge areas in coastal arid area southern tunisia. **Groundwater for Sustainable Development**, p. 100760, 2022. ISSN 2352-801X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2352801X22000376>>.
- ADIMALLA, N.; WU, J. Groundwater quality and associated health risks in a semi-arid region of south india: Implication to sustainable groundwater management. **Human and Ecological Risk Assessment: An International Journal**, Taylor Francis, v. 25, n. 1-2, p. 191–216, 2019. Disponível em: <<https://doi.org/10.1080/10807039.2018.1546550>>.
- AFAN, H. A.; OSMAN, A. I. A.; ESSAM, Y.; AHMED, A. N.; HUANG, Y. F.; KISI, O.; SHERIF, M.; SEFELNASR, A.; CHAU, K. W.; EL-SHAFIE, A. Modeling the fluctuations of groundwater level by employing ensemble deep learning techniques. **Engineering Applications of Computational Fluid Mechanics**, Taylor Francis, v. 15, n. 1, p. 1420–1439, 2021. Disponível em: <<https://doi.org/10.1080/19942060.2021.1974093>>.
- AHMADI, A.; OLYAEI, M.; HEYDARI, Z.; EMAMI, M.; ZEYNOLABEDIN, A.; GHOMLAGHI, A.; DACCACHE, A.; FOGG, G. E.; SADEGH, M. Groundwater level modeling with machine learning: A systematic review and meta-analysis. **Water**, v. 14, n. 6, 2022. ISSN 2073-4441. Disponível em: <<https://www.mdpi.com/2073-4441/14/6/949>>.
- ALLAIRE, J.; CHOLLET, F. **keras: R Interface to 'Keras'**. [S.l.], 2022. R package version 2.8.0. Disponível em: <<https://CRAN.R-project.org/package=keras>>.
- ALLAIRE, J.; TANG, Y. **tensorflow: R Interface to 'TensorFlow'**. [S.l.], 2022. R package version 2.8.0. Disponível em: <<https://CRAN.R-project.org/package=tensorflow>>.
- ASGHARINIA, S.; PETROSELLI, A. A comparison of statistical methods for evaluating missing data of monitoring wells in the kazeroun plain, fars province, iran. **Groundwater for Sustainable Development**, v. 10, p. 100294, 2020. ISSN 2352-801X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2352801X18302753>>.
- ASHRAF, S.; NAZEMI, A.; AGHAKOUCHAK, A. Anthropogenic drought dominates groundwater depletion in iran. **Scientific reports**, Nature Publishing Group, v. 11, n. 1, p. 1–10, 2021.
- ASSIS, T. F. O. **Cópias para combinação de modelos de séries temporais**. 105 p. Doutorado em Biometria e Estatística Aplicada — Universidade Federal Rural de Pernambuco, Recife, 2017.
- AWAD, M.; KHANNA, R. **Efficient Learning Machines: Theories, concepts, and applications for engineers and system designers**. [S.l.]: Apress open, 2015.
- AYDILEK, I. B.; ARSLAN, A. A hybrid method for imputation of missing values using optimized fuzzy c-means with support vector regression and a genetic algorithm. **Information Sciences**, Elsevier, v. 233, p. 25–35, 2013.
- AZIZPOUR, A.; IZADBAKHS, M. A.; SHABANLOU, S.; YOSEFVAND, F.; RAJABI, A. Estimation of water level fluctuations in groundwater through a hybrid learning

- machine. **Groundwater for Sustainable Development**, v. 15, p. 100687, 2021. ISSN 2352-801X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2352801X21001442>>.
- BATES, J. M.; GRANGER, C. W. The combination of forecasts. **Journal of the Operational Research Society**, Taylor & Francis, v. 20, n. 4, p. 451–468, 1969.
- BERGMEIR, C.; HYNDMAN, R. J.; BENITEZ, J. M. Bagging exponential smoothing methods using stl decomposition and box–cox transformation. **International Journal of Forecasting**, v. 32, n. 2, p. 303–312, 2016. ISSN 0169-2070. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0169207015001120>>.
- BEZERRA, N. F. Água no semi-árido nordestino: experiências e desafios. In: **Água e desenvolvimento sustentável no Semi-Árido**. Fortaleza: Fundação Konrad Adenauer, 2002. v. 24, p. 35 – 52.
- BOKDE, N.; BECK, M. W.; ÁLVAREZ, F. M.; KULAT, K. A novel imputation methodology for time series based on pattern sequence forecasting. **Pattern Recognition Letters**, v. 116, p. 88–96, 2018. ISSN 0167-8655. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167865518306500>>.
- BOX, G. E.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. **Time series analysis: forecasting and control**. [S.l.]: John Wiley & Sons, 2015.
- BOZORG-HADDAD, O.; DELPASAND, M.; LOÁICIGA, H. A. Self-optimizer data-mining method for aquifer level prediction. **Water Supply**, v. 20, n. 2, p. 724–736, 12 2019. ISSN 1606-9749. Disponível em: <<https://doi.org/10.2166/ws.2019.204>>.
- BRASIL. Lei nº 9.433, de 8 de janeiro de 1997. Lei das águas. **Diário Oficial da União**, Poder Executivo, Brasília, DF, 8 jan. 1997.
- BURITI, C. O.; BARBOSA, H. A.; PAREDES-TREJO, F. J.; KUMAR, T. V.; THAKUR, M. K.; RAO, K. K. Un siglo de sequías: ¿Por qué las políticas de agua no desarrollaron la región semiárida brasileña? **Revista Brasileira de Meteorologia**, SciELO Brasil, v. 35, p. 683–688, 2020.
- CAMPOS, J. N. B. Secas e políticas públicas no semiárido: ideias, pensadores e períodos. **Estudos Avançados**, scielo, v. 28, p. 65 – 88, 12 2014. ISSN 0103-4014. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-40142014000300005&nrm=iso>.
- CHAI, T.; DRAXLER, R. R. Root mean square error (rmse) or mean absolute error (mae)? – arguments against avoiding rmse in the literature. **Geoscientific Model Development**, v. 7, n. 3, p. 1247–1250, 2014. Disponível em: <<https://gmd.copernicus.org/articles/7/1247/2014/>>.
- CHAN, K.; CRYER, J. **Time series analysis with applications in r**. [S.l.: s.n.], 2008.
- CHAPRA, S.; CANALE, R. **Metodos Numericos para Engenharia**. McGraw Hill Brasil, 2016. ISBN 9788580555691. Disponível em: <<https://books.google.com.br/books?id=8pMyDAAAQBAJ>>.

- CIRILO, J. A. Políticas públicas de recursos hídricos para o semi-árido. **Estudos Avançados**, scielo, v. 22, p. 61 – 82, 00 2008. ISSN 0103-4014. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-40142008000200005&nrm=iso>.
- CLEVELAND, R. B.; CLEVELAND, W. S.; MCRAE, J. E.; TERPENNING, I. STL: A seasonal-trend decomposition procedure based on loess. **Journal of Official Statistics**, v. 6, n. 1, p. 3–33, 1990.
- CLEVELAND, W. S.; DEVLIN, S. J. Locally weighted regression: an approach to regression analysis by local fitting. **Journal of the American statistical association**, Taylor & Francis, v. 83, n. 403, p. 596–610, 1988.
- COGERH. **Plano de Monitoramento e Gestão dos Aquíferos da Bacia do Araripe**: Estado do Ceará. Fortaleza, CE, 2009.
- COGERH. **Boletim de Monitoramento dos Poços com Datalogger no Cariri - CE (Ano de 2019)**: Estado do Ceará. Fortaleza, CE, 2019.
- COUNCIL, N. R. **Our Common Journey**: A transition toward sustainability. Washington, DC: The National Academies Press, 1999.
- CPRM. **Implantação de Rede Integrada de Monitoramento das Águas Subterrâneas**. Brasil, 2009. Disponível em: <http://rimasweb.cprm.gov.br/layout/pdf/proposta_monitoramento_CPRM_2009.pdf>. Acesso em: 28 mar. 2020.
- DA CUNHA JÚNIOR, R. O.; COSTA, C. T. F.; FIRMINO, P. R. A. Preenchimento de falhas em séries temporais de níveis de águas subterrâneas usando métodos univariados. In: **Anais do XIV ENREHSE - Encontro de Recursos Hídricos em Sergipe**. Aracaju, Sergipe: ABRHidro - Associação Brasileira de Recursos Hídricos, 2022. p. 1–6. Disponível em: <<https://anais.abrhidro.org.br/job.php?Job=13793>>.
- DA CUNHA JÚNIOR, R. O.; FIRMINO, P. R. A. Simulação de valores ausentes em séries temporais de precipitação para avaliação de métodos de imputação. **Revista Brasileira de Climatologia**, v. 30, n. 18, p. 691–714, 2022. Disponível em: <<https://ojs.ufgd.edu.br/index.php/rbclima/article/view/15243>>.
- DE MATTOS NETO, P. S. G.; FIRMINO, P. R. A.; SIQUEIRA, H.; TADANO, Y. D. S.; ALVES, T. A.; OLIVEIRA, J. F. L. D.; MARINHO, M. H. D. N.; MADEIRO, F. Neural-based ensembles for particulate matter forecasting. **IEEE Access**, v. 9, p. 14470–14490, 2021.
- DE OLIVEIRA, J. F. L.; PACÍFICO, L. D. S.; DE MATTOS NETO, P. S. G.; BARREIROS, E. F. S.; RODRIGUES, C. M. O.; FILHO, A. T. A. A hybrid optimized error correction system for time series forecasting. **Applied Soft Computing**, v. 87, p. 105970, Feb 2020. ISSN 1568-4946. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1568494619307513>>.
- DE OLIVEIRA, R. T.; DE ASSIS, T. F. O.; FIRMINO, P. R. A.; FERREIRA, T. A. Copulas-based time series combined forecasters. **Information Sciences**, v. 376, p. 110–124, 2017. ISSN 0020-0255. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0020025516312397>>.

DIRECTIVE, W. F. **Common Implementation Strategy for the Water Framework Directive (2000/60/EC)**: Guidance Document No7. Monitoring under the Water Framework Directive. Technical Report – 002 – 2003. Luxemburgo, 2003.

ENDERS, C. **Applied Missing Data Analysis**. [S.l.]: Guilford Publications, 2010. (Methodology in the social sciences). ISBN 9781606236390.

ENEMARK, T.; PEETERS, L. J.; MALLANTS, D.; BATELAAN, O. Hydrogeological conceptual model building and testing: A review. **Journal of Hydrology**, v. 569, p. 310–329, 2019. ISSN 0022-1694. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0022169418309387>>.

FAMBRINI, G. L.; SILVESTRE, D. C.; BARRETO JUNIOR, A. M.; SILVA-FILHO, W. F. Estratigrafia da bacia do araripe: estado da arte, revisão crítica e resultados novos. **Geologia USP. Série Científica**, v. 20, n. 4, p. 169–212, dez. 2020. Disponível em: <<https://www.revistas.usp.br/guspsc/article/view/180295>>.

FEITOSA, F. A. C. **Hidrogeologia**: conceitos e aplicações. Rio de Janeiro: CPRM, 2008.

FENG, L.; NOWAK, G.; O'NEILL, T.; WELSH, A. Cutoff: A spatio-temporal imputation method. **Journal of Hydrology**, v. 519, p. 3591–3605, 2014. ISSN 0022-1694. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S002216941400897X>>.

FIRMINO, P. R. A.; DE MATTOS NETO, P. S.; FERREIRA, T. A. Correcting and combining time series forecasters. **Neural Networks**, Elsevier, v. 50, p. 1–11, 2014.

FRITSCH, S.; GUENTHER, F.; WRIGHT, M. N. **neuralnet: Training of Neural Networks**. [S.l.], 2019. R package version 1.44.2. Disponível em: <<https://CRAN.R-project.org/package=neuralnet>>.

GARJULLI, R. A. Os recursos hídricos no semi-Árido. **Ciência e Cultura**, scielocec, v. 55, p. 38 – 39, 12 2003. ISSN 0009-6725. Disponível em: <http://cienciaecultura.bvs.br/scielo.php?script=sci_arttext&pid=S0009-67252003000400021&nrm=iso>.

GUPTA, A.; LAM, M. S. Estimating missing values using neural networks. **Journal of the Operational Research Society**, Springer, v. 47, n. 2, p. 229–238, 1996.

HAJIRAHIMI, Z.; KHASHEI, M. Hybrid structures in time series modeling and forecasting: A review. **Engineering Applications of Artificial Intelligence**, Elsevier, v. 86, p. 83–106, 2019.

HAN, K.; ZUO, R.; NI, P.; XUE, Z.; XU, D.; WANG, J.; ZHANG, D. Application of a genetic algorithm to groundwater pollution source identification. **Journal of Hydrology**, v. 589, p. 125343, 2020. ISSN 0022-1694. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0022169420308039>>.

HAN, Y.; WANG, L.; GAO, J.; XING, Z.; TAO, T. Combination forecasting based on svm and neural network for urban rail vehicle spare parts demand. In: **2017 36th Chinese Control Conference (CCC)**. [S.l.: s.n.], 2017. p. 4660–4665.

HARMAN, B. I.; KOSEOGLU, H.; YIGIT, C. O. Performance evaluation of idw, kriging and multiquadric interpolation methods in producing noise mapping: A case study at the city of isparta, turkey. **Applied Acoustics**, Elsevier, v. 112, p. 147–157, 2016.

- HAYKIN, S. **Neural Networks: A comprehensive foundation**. [S.l.]: Prentice Hall, 1999.
- HE, L.; CHEN, S.; LIANG, Y.; HOU, M.; CHEN, J. Infilling the missing values of groundwater level using time and space series: case of nantong city, east coast of china. **Earth Science Informatics**, Springer, v. 13, n. 4, p. 1445–1459, 2020.
- HOCHREITER, S.; SCHMIDHUBER, J. Long Short-Term Memory. **Neural Computation**, v. 9, n. 8, p. 1735–1780, 11 1997. ISSN 0899-7667. Disponível em: <<https://doi.org/10.1162/neco.1997.9.8.1735>>.
- HOFERT, M.; KOJADINOVIC, I.; MAECHLER, M.; YAN, J. **copula: Multivariate Dependence with Copulas**. [S.l.], 2020. R package version 1.0-1. Disponível em: <<https://CRAN.R-project.org/package=copula>>.
- HONGHAI, F.; GUOSHUN, C.; CHENG, Y.; BINGRU, Y.; YUMEI, C. A svm regression based approach to filling in missing values. In: SPRINGER. **International Conference on Knowledge-Based and Intelligent Information and Engineering Systems**. [S.l.], 2005. p. 581–587.
- HUANG, G. B.; ZHU, Q. Y.; SIEW, C. K. Extreme learning machine: Theory and applications. **Neurocomputing**, v. 70, n. 1, p. 489–501, 2006. ISSN 0925-2312. Neural Networks. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0925231206000385>>.
- HUNTER, J. S. The exponentially weighted moving average. **Journal of quality technology**, Taylor & Francis, v. 18, n. 4, p. 203–210, 1986.
- HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: principles and practice**. Melbourne, Australia: OTexts, 2018. Disponível em: <<https://www.OTexts.com/fpp2>>. Acesso em: 28 mar. 2020.
- HYNDMAN, R. J.; KHANDAKAR, Y. Automatic time series forecasting: the forecast package for R. **Journal of Statistical Software**, v. 26, n. 3, p. 1–22, 2008. Disponível em: <<http://www.jstatsoft.org/article/view/v027i03>>.
- HYNDMAN, R. J.; KOEHLER, A. B.; ORD, J. K.; SNYDER, R. D. **Forecasting with Exponential Smoothing: The state space approach**. [S.l.]: Springer, 2008.
- IPECE. **Perfil das regiões de planejamento - Cariri 2017**. Fortaleza, CE, 2017. Disponível em: <http://www2.ipece.ce.gov.br/estatistica/perfil_regional/2017/pr_cariri_2017.pdf>.
- IQBAL, N.; KHAN, A. N.; RIZWAN, A.; AHMAD, R.; KIM, B. W.; KIM, K.; KIM, D. H. Groundwater level prediction model using correlation and difference mechanisms based on boreholes data for sustainable hydraulic resource management. **IEEE Access**, v. 9, p. 96092–96113, 2021.
- JIANG, Z.; YANG, S.; LIU, Z.; XU, Y.; SHEN, T.; QI, S.; PANG, Q.; XU, J.; LIU, F.; XU, T. Can ensemble machine learning be used to predict the groundwater level dynamics of farmland under future climate: a 10-year study on huaibei plain. **Environmental Science and Pollution Research**, Feb 2022. ISSN 1614-7499. Disponível em: <<https://doi.org/10.1007/s11356-022-18809-8>>.

- JOE, H.; XU, J. J. The estimation method of inference functions for margins for multivariate models. 1996. Disponível em: <<https://open.library.ubc.ca/collections/52383/items/1.0225985>>.
- JUNNINEN, H.; NISKA, H.; TUPPURAINEN, K.; RUUSKANEN, J.; KOLEHMAINEN, M. Methods for imputation of missing values in air quality data sets. **Atmospheric Environment**, Elsevier, v. 38, n. 18, p. 2895–2907, 2004.
- KARAMOUZ, M.; NAZIF, S.; FALAHI, M. **Hydrology and hydroclimatology: principles and applications**. [S.l.]: CRC Press, 2012.
- KASHANI, M. H.; DINPASHOH, Y. Evaluation of efficiency of different estimation methods for missing climatological data. **Stochastic Environmental Research and Risk Assessment**, Springer, v. 26, n. 1, p. 59–71, 2012.
- KATOCH, S.; CHAUHAN, S. S.; KUMAR, V. A review on genetic algorithm: past, present, and future. **Multimedia Tools and Applications**, v. 80, n. 5, p. 8091–8126, Feb 2021. ISSN 1573-7721. Disponível em: <<https://doi.org/10.1007/s11042-020-10139-6>>.
- KIM, I.; LEE, J. Performance analysis of ann prediction for groundwater level considering regional-specific influence components. **Groundwater**, v. 60, n. 3, p. 344–361, 2022. Disponível em: <<https://ngwa.onlinelibrary.wiley.com/doi/abs/10.1111/gwat.13156>>.
- LEE, C. A review of applications of genetic algorithms in operations management. **Engineering Applications of Artificial Intelligence**, v. 76, p. 1–12, 2018. ISSN 0952-1976. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S095219761830174X>>.
- LEE, S.; LEE, K. K.; YOON, H. Using artificial neural network models for groundwater level forecasting and assessment of the relative impacts of influencing factors. **Hydrogeology Journal**, Springer, v. 27, n. 2, p. 567–579, 2019.
- LI, Q.; FU, Y.; ZHOU, X.; XU, Y. The investigation and application of svc and svr in handling missing values. In: IEEE. **2009 First International Conference on Information Science and Engineering**. [S.l.], 2009. p. 1002–1005.
- LITTLE, R.; RUBIN, D. **Statistical Analysis with Missing Data**. [S.l.]: Wiley, 2019. (Wiley Series in Probability and Statistics). ISBN 9780470526798.
- LIU, Y.; DILLON, T.; YU, W.; RAHAYU, W.; MOSTAFA, F. Missing value imputation for industrial iot sensor data with large gaps. **IEEE Internet of Things Journal**, IEEE, v. 7, n. 8, p. 6855–6867, 2020.
- LORENA, A. C.; CARVALHO, A. de. Uma introdução às support vector machines. **Revista de Informática Teórica e Aplicada**, v. 14, n. 2, p. 43–67, 2007. ISSN 21752745. Disponível em: <https://seer.ufrgs.br/rita/article/view/rita_v14_n2_p43-67>.
- MAITY, R. **Statistical methods in hydrology and hydroclimatology**. [S.l.]: Springer, 2018. v. 585.
- MALEKZADEH, M.; KARDAR, S.; SHABANLOU, S. Simulation of groundwater level using modflow, extreme learning machine and wavelet-extreme learning machine models. **Groundwater for Sustainable Development**, v. 9, p. 100279, 2019.

ISSN 2352-801X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2352801X19302127>>.

MALIK, A.; BHAGWAT, A. Modelling groundwater level fluctuations in urban areas using artificial neural network. **Groundwater for Sustainable Development**, v. 12, p. 100484, 2021. ISSN 2352-801X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2352801X20303829>>.

MANDIC, D.; CHAMBERS, J. **Recurrent neural networks for prediction: learning algorithms, architectures and stability**. [S.l.]: Wiley, 2001.

MANZIONE, R. L. **Águas Subterrânea: conceitos e aplicações sob uma visão multidisciplinar**. Jundiaí: Paco editorial, 2015. 388 p. ISBN 978-85-8148-786-1.

MEKANIK, F.; IMTEAZ, M.; GATO-TRINIDAD, S.; ELMAHDI, A. Multiple regression and artificial neural network for long-term rainfall forecasting using large scale climate modes. **Journal of Hydrology**, Elsevier, v. 503, p. 11–21, 2013.

MEYER, D.; DIMITRIADOU, E.; HORNIK, K.; WEINGESSEL, A.; LEISCH, F. **e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien**. [S.l.], 2019. R package version 1.7-3. Disponível em: <<https://CRAN.R-project.org/package=e1071>>.

MIRARABI, A.; NASSERY, H.; NAKHAEI, M.; ADAMOWSKI, J.; AKBARZADEH, A.; ALIJANI, F. Evaluation of data-driven models (svr and ann) for groundwater-level prediction in confined and unconfined systems. **Environmental Earth Sciences**, Springer, v. 78, n. 15, p. 1–15, 2019.

MIRJALILI, S. Genetic algorithm. In: _____. **Evolutionary Algorithms and Neural Networks: Theory and Applications**. Cham: Springer International Publishing, 2019. p. 43–55. ISBN 978-3-319-93025-1. Disponível em: <https://doi.org/10.1007/978-3-319-93025-1_4>.

MORETTIN, P. A.; TOLOI, C. M. C. **Modelos para previsão de séries temporais**. Rio de Janeiro: IMPA, 1981.

MORITZ, S.; BARTZ-BEIELSTEIN, T. imputets: time series missing value imputation in r. **R J.**, v. 9, n. 1, p. 207, 2017.

MORITZ, S.; BARTZ-BEIELSTEIN, T. imputeTS: Time Series Missing Value Imputation in R. **The R Journal**, v. 9, n. 1, p. 207–218, 2017.

MORITZ, S.; SARDÁ, A.; BARTZ-BEIELSTEIN, T.; ZAEFFERER, M.; STORK, J. Comparison of different methods for univariate time series imputation in r. **arXiv preprint arXiv:1510.03924**, 2015.

MOUSELIMIS, L. **elmNNRcpp: The Extreme Learning Machine Algorithm**. [S.l.], 2022. R package version 1.0.4. Disponível em: <<https://CRAN.R-project.org/package=elmNNRcpp>>.

MOZAFFARI, S.; JAVADI, S.; MOGHADDAM, H. K.; RANDHIR, T. O. Forecasting groundwater levels using a hybrid of support vector regression and particle swarm optimization. **Water Resources Management**, v. 36, n. 6, p. 1955–1972, Apr 2022. ISSN 1573-1650. Disponível em: <<https://doi.org/10.1007/s11269-022-03118-z>>.

MÜLLER, J.; PARK, J.; SAHU, R.; VARADHARAJAN, C.; ARORA, B.; FAYBISHENKO, B.; AGARWAL, D. Surrogate optimization of deep neural networks for groundwater predictions. **Journal of Global Optimization**, v. 81, n. 1, p. 203–231, Sep 2021. ISSN 1573-2916. Disponível em: <<https://doi.org/10.1007/s10898-020-00912-0>>.

NAJAFABADIPOUR, A.; KAMALI, G.; NEZAMABADI-POUR, H. The innovative combination of time series analysis methods for the forecasting of groundwater fluctuations. **Water Resources**, v. 49, n. 2, p. 283–291, Apr 2022. ISSN 1608-344X. Disponível em: <<https://doi.org/10.1134/S0097807822020026>>.

NELSEN, R. B. **An introduction to copulas**. [S.l.]: Springer Science & Business Media, 2013.

NOORI, A. R.; SINGH, S. K. Spatial and temporal trend analysis of groundwater levels and regional groundwater drought assessment of kabul, afghanistan. **Environmental Earth Sciences**, Springer, v. 80, n. 20, p. 1–16, 2021.

OIKONOMOU, P. D.; ALZRAIEE, A. H.; KARAVITIS, C. A.; WASKOM, R. M. A novel framework for filling data gaps in groundwater level observations. **Advances in Water Resources**, v. 119, p. 111–124, 2018. ISSN 0309-1708. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S030917081731179X>>.

OLIVEIRA, A. L. I. Estimation of software project effort with support vector regression. **Neurocomputing**, Elsevier, v. 69, p. 1749 – 1753, 2006.

OLIVEIRA, R. T. A. Mestrado em Informática Aplicada, **Modelagem e simulação computacional da combinação de preditores de séries temporais por meio de cópulas**. Recife: [s.n.], 2014. 105 p.

ONU. **Transforming our world: the 2030 Agenda for Sustainable Development**. 2015. Acessado em: 28 mai 2021. Disponível em: <<https://sdgs.un.org/2030agenda>>.

PALIT, A. K.; POPOVIC, D. **Computational Intelligence in Time Series Forecasting**: Theory and engineering applications. [S.l.]: Springer, 2005.

PATRO, S.; SAHU, K. K. Normalization: A preprocessing stage. **arXiv preprint arXiv:1503.06462**, 2015.

PAULHUS, J. L. H.; KOHLER, M. A. Interpolation of missing precipitation records. **Monthly Weather Review**, American Meteorological Society, Boston MA, USA, v. 80, n. 8, p. 129 – 133, 1952. Disponível em: <https://journals.ametsoc.org/view/journals/mwre/80/8/1520-0493_1952_080_0129_iompr_2_0_co_2.xml>.

PHAN, T. T. H.; CAILLAULT, P.; LEFEBVRE, A.; BIGAND, A. Dynamic time warping-based imputation for univariate time series data. **Pattern Recognition Letters**, v. 139, p. 139–147, 2020. ISSN 0167-8655. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167865517302751>>.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2022. Disponível em: <<https://www.R-project.org/>>.

RAJAEI, T.; EBRAHIMI, H.; NOURANI, V. A review of the artificial intelligence methods in groundwater level modeling. **Journal of hydrology**, Elsevier, v. 572, p. 336–351, 2019.

REBOUÇAS, A. C. Água na região nordeste: desperdício e escassez. **Estudos Avançados**, scielo, v. 11, p. 127 – 154, 04 1997. ISSN 0103-4014. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-40141997000100007&nrm=iso>.

RIMAS. **Rede Integrada de Monitoramento das Águas Subterrâneas**. CPRM - Serviço Geológico do Brasil, 2021. Acesso em: 28 nov. 2021. Disponível em: <<http://rimasweb.cprm.gov.br/layout/>>.

ROSHNI, T.; JHA, M. K.; DRISYA, J. Neural network modeling for groundwater-level forecasting in coastal aquifers. **Neural Computing and Applications**, v. 32, n. 16, p. 12737–12754, Aug 2020. ISSN 1433-3058. Disponível em: <<https://doi.org/10.1007/s00521-020-04722-z>>.

ROY, D. K.; BISWAS, S. K.; MATTAR, M. A.; EL-SHAFAEI, A. A.; MURAD, K. F. I.; SAHA, K. K.; DATTA, B.; DEWIDAR, A. Z. Groundwater level prediction using a multiple objective genetic algorithm-grey relational analysis based weighted ensemble of anfis models. **Water**, v. 13, n. 21, 2021. ISSN 2073-4441. Disponível em: <<https://www.mdpi.com/2073-4441/13/21/3130>>.

SÃO JOSÉ, R. V.; COLTRI, P. P.; GRECO, R.; MELO, H. L. S.; SANTOS, K. A.; SOUZA, I. S. Seca extrema de 2012 no semiárido baiano e seus impactos: Informações climáticas difundidas pela mídia. **Revista Brasileira de Climatologia**, v. 29, p. 414–439, 2021.

SACHS, I. **Caminhos para o desenvolvimento sustentável**. [S.l.]: Editora Garamond, 2000.

SAHU, R. K.; MÜLLER, J.; PARK, J.; VARADHARAJAN, C.; ARORA, B.; FAYBISHENKO, B.; AGARWAL, D. Impact of input feature selection on groundwater level prediction from a multi-layer perceptron neural network. **Frontiers in Water**, Frontiers, v. 2, p. 46, 2020.

SAKIZADEH, M.; MOHAMED, M. M.; KLAMMLER, H. Trend analysis and spatial prediction of groundwater levels using time series forecasting and a novel spatio-temporal method. **Water Resources Management**, Springer, v. 33, n. 4, p. 1425–1437, 2019.

SALES, J. P. **Modelling and forecasting sustainable development time series via single and combined approaches**. Dissertação (Mestrado em Desenvolvimento Regional Sustentável) — Universidade Federal do Cariri, Crato, 2019.

SAVARIMUTHU, N.; KARESIDDIAIAH, S. An unsupervised neural network approach for imputation of missing values in univariate time series data. **Concurrency and Computation: Practice and Experience**, v. 33, n. 9, p. e6156, 2021. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.6156>>.

SCRUCCA, L. GA: A package for genetic algorithms in R. **Journal of Statistical Software**, v. 53, n. 4, p. 1–37, 2013. Disponível em: <<http://www.jstatsoft.org/v53/i04/>>.

SEARCY, J. K.; HARDISON, C. H. **Double-mass curves**. [S.l.]: US Government Printing Office, 1960.

SEIFI, A.; EHTERAM, M.; SINGH, V. P.; MOSAVI, A. Modeling and uncertainty analysis of groundwater level using six evolutionary optimization algorithms hybridized with anfis, svm, and ann. **Sustainability**, Multidisciplinary Digital Publishing Institute, v. 12, n. 10, p. 4023, 2020.

SIAMI-NAMINI, S.; TAVAKOLI, N.; NAMIN, A. S. A comparative analysis of forecasting financial time series using arima, lstm, and bilstm. **arXiv preprint arXiv:1911.09512**, 2019.

SILVA, M.; GONÇALVES, A.; LOPES, W.; LIMA, M.; COSTA, C.; PARIS, M.; FIRMINO, P.; DE PAULA FILHO, F. Assessment of groundwater quality in a brazilian semiarid basin using an integration of gis, water quality index and multivariate statistical techniques. **Journal of Hydrology**, v. 598, p. 126346, 2021. ISSN 0022-1694. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0022169421003930>>.

SILVA, T. A. **Estudo do desempenho da combinação de preditores baseados em cópulas e support vector machines para séries temporais úteis ao desenvolvimento sustentável**. 96 p. Dissertação (Mestrado em Biometria e Estatística Aplicada) — Universidade Federal Rural de Pernambuco, Recife, 2020.

SMOLA, A. J.; SCHOLKOPF, B. A tutorial on support vector regression. **Statistics and Computing**, v. 14, p. 199–222, 2003.

SOLGI, R.; LOÁICIGA, H. A.; KRAM, M. Long short-term memory neural network (lstm-nn) for aquifer level time series forecasting using in-situ piezometric observations. **Journal of Hydrology**, v. 601, p. 126800, 2021. ISSN 0022-1694. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0022169421008507>>.

SOUZA FILHO, F. A. A política nacional de recursos hídricos: desafios para sua implantação no semiárido brasileiro. In: **Recursos hídricos em regiões áridas e semiáridas**. Campina Grande: Instituto Nacional do Semiárido, 2011. p. 01 – 25.

STEINWART, I.; CHRISTMANN, A. **Support Vector Machines**. [S.l.]: Springer, 2008.

SUDENE. Resolução nº 107, de 27 de julho de 2017. Estabelece critérios técnicos e científicos para delimitação do semiárido brasileiro e procedimentos para revisão de sua abrangência. Recife, PE, 27 jul. 2017.

TAO, H.; HAMEED, M. M.; MARHOON, H. A.; ZOUNEMAT-KERMANI, M.; HEDDAM, S.; KIM, S.; SULAIMAN, S. O.; TAN, M. L.; SA'ADI, Z.; MEHR, A. D.; ALLAWI, M. F.; ABBA, S.; ZAIN, J. M.; FALAH, M. W.; JAMEI, M.; BOKDE, N. D.; BAYATVARKESHI, M.; AL-MUKHTAR, M.; BHAGAT, S. K.; TIYASHA, T.; KHEDHER, K. M.; AL-ANSARI, N.; SHAHID, S.; YASEEN, Z. M. Groundwater level prediction using machine learning models: A comprehensive review. **Neurocomputing**, v. 489, p. 271–308, 2022. ISSN 0925-2312. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S092523122200282X>>.

TEEGAVARAPU, R. S. Precipitation imputation with probability space-based weighting methods. **Journal of Hydrology**, v. 581, p. 124447, 2020. ISSN 0022-1694. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0022169419311825>>.

TEGEGNE, G.; KIM, Y.-O. Representing inflow uncertainty for the development of monthly reservoir operations using genetic algorithms. **Journal of Hydrology**, v. 586, p. 124876, 2020. ISSN 0022-1694. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S002216942030336X>>.

TEIXEIRA, F. J. C. **Modelos de gerenciamento de recursos hídricos: análises e proposta de aperfeiçoamento do sistema do ceará**. Dissertação (Mestrado em Recursos Hídricos) — Universidade Federal do Ceará, Fortaleza, 2003.

THEODOSIOU, M. Forecasting monthly and quarterly time series using stl decomposition. **International Journal of Forecasting**, Elsevier, v. 27, n. 4, p. 1178–1195, 2011.

VERISSIMO, L. S. **A importância das Águas subterrâneas para o Desenvolvimento socioeconômico do Eixo CRAJUBAR, Cariri ocidental - Estado do Ceará**. Dissertação (Mestrado em Hidrogeologia) — Universidade Federal do Ceará, Fortaleza, 1999.

WHEATER, H. S.; MATHIAS, S. A.; LI, X. **Groundwater Modelling in Arid and Semi-Arid Areas**. New York: Cambridge University Press, 2010. ISBN 978-0-521-11129-4.

WHITLEY, D. A genetic algorithm tutorial. **Statistics and computing**, Springer, v. 4, n. 2, p. 65–85, 1994.

WU, M.; FENG, Q.; WEN, X.; YIN, Z.; YANG, L.; SHENG, D. Deterministic analysis and uncertainty analysis of ensemble forecasting model based on variational mode decomposition for estimation of monthly groundwater level. **Water**, v. 13, n. 2, 2021. ISSN 2073-4441. Disponível em: <<https://www.mdpi.com/2073-4441/13/2/139>>.

WUNSCH, A.; LIESCH, T.; BRODA, S. Groundwater level forecasting with artificial neural networks: a comparison of long short-term memory (lstm), convolutional neural networks (cnns), and non-linear autoregressive networks with exogenous input (narx). **Hydrology and Earth System Sciences**, v. 25, n. 3, p. 1671–1687, 2021. Disponível em: <<https://hess.copernicus.org/articles/25/1671/2021/>>.

XING, Z.; ZHANG, H.; JI, Y.; XINGLONG, G.; FU, Q.; LI, H. Markov chain Monte Carlo based on adaptive Metropolis algorithm applied in combined model to deal with the uncertainty of weights of single models. **Water Supply**, v. 19, n. 4, p. 1129–1136, 10 2018. ISSN 1606-9749. Disponível em: <<https://doi.org/10.2166/ws.2018.164>>.

XIONG, T.; LI, C.; BAO, Y. Seasonal forecasting of agricultural commodity price using a hybrid stl and elm method: Evidence from the vegetable market in china. **Neurocomputing**, v. 275, p. 2831–2844, 2018. ISSN 0925-2312. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S092523121731812X>>.

YADAV, B.; CH, S.; MATHUR, S.; ADAMOWSKI, J. Assessing the suitability of extreme learning machines (elm) for groundwater level prediction. **Journal of Water and Land Development**, v. 32, p. 103–112, 2017.

YANG, L.; SHAMI, A. On hyperparameter optimization of machine learning algorithms: Theory and practice. **Neurocomputing**, v. 415, p. 295–316, 2020. ISSN 0925-2312. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0925231220311693>>.

YU, L.; WANG, S.; LAI, K. K. A novel nonlinear ensemble forecasting model incorporating glar and ann for foreign exchange rates. **Computers & Operations Research**, Elsevier, v. 32, n. 10, p. 2523–2541, 2005.

YU, P.; CHEN, S.; CHANG, I. Support vector regression for real-time flood stage forecasting. **Journal of Hydrology**, Elsevier, v. 328, p. 704–716, 2006.

ZANOTTI, C.; ROTIROTI, M.; STERLACCHINI, S.; CAPPELLINI, G.; FUMAGALLI, L.; STEFANIA, G. A.; NANNUCCI, M. S.; LEONI, B.; BONOMI, T. Choosing between linear and nonlinear models and avoiding overfitting for short and long term groundwater level forecasting in a linear system. **Journal of Hydrology**, v. 578, p. 124015, 2019. ISSN 0022-1694. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0022169419307425>>.

ZHANG, S. P. An investigation of neural networks for linear time-series forecasting. **Computers Operations Research**, Elsevier, n. 28, p. 1183–1202, 2001.

ZHANG, Z.; YANG, X.; LI, H.; LI, W.; YAN, H.; SHI, F. Application of a novel hybrid method for spatiotemporal data imputation: A case study of the minqin county groundwater level. **Journal of Hydrology**, v. 553, p. 384–397, 2017. ISSN 0022-1694. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0022169417305188>>.

APÊNDICE A – INFORMAÇÕES SUPLEMENTARES

Tabela A.1 – Informações gerais e geológicas dos poços na BSA

ID RIMAS	Série	Cota (m)	Latitude	Longitude	Município	Nome	Formação	Aquífero
2300020827	P01	402,89	-7,319722	-39,257777	Barbalha	Sítio Santana III	Missão Velha	Médio
2300022129	P02	399,67	-7,496388	-38,911944	Brejo santo	Jenipapeiro	Missão Velha	Médio
2300022135	P03	362,34	-7,408055	-38,932222	Milagres	Água Vermelha	Missão Velha	Médio
2300022154	P04	360,40	-7,269166	-39,175000	Missão velha	Sítio Jerimum	Missão Velha	Médio
2300022506	P05	458,29	-7,591388	-39,025000	Brejo santo	Sítio Boqueirão	Missão Velha	Médio
2300022592	P06	466,02	-7,257778	-39,303056	Barbalha	Sede UFCA (Campus Cariri)	Missão Velha	Médio
2300022593	P07	455,15	-7,371666	-39,129444	Missão velha	Sítio Carrancudo	Missão Velha	Médio
2300022598	P08	451,91	-7,289722	-39,051666	Milagres	Sítio Melo	Missão Velha	Médio
2300022599	P09	450,71	-7,355000	-39,170555	Missão velha	Sítio Canta Galo	Missão Velha	Médio
2300022600	P10	373,13	-7,347777	-38,985000	Abaiaara	Sítio Cajueiro	Missão Velha	Médio
2300022889	P11	586,20	-7,331111	-38,593333	Mauriti	Sítio Bananeiras	Mauriti	Inferior
2300022891	P12	417,17	-7,286666	-38,817500	Milagres	Sítio São Sebastião	Mauriti	Inferior
2300022892	P13	505,62	-7,291944	-38,717222	Mauriti	Sítio Deserto dos Pebas	Mauriti	Inferior
2300022893	P14	395,04	-7,491111	-38,802222	Mauriti	Sítio Jequi	Mauriti	Inferior
2300022894	P15	378,59	-7,292777	-38,948333	Milagres	Sítio Serrote	Missão Velha	Médio
2300022895	P16	398,00	-7,391111	-39,015277	Abaiaara	Sítio Queimadas	Missão Velha	Médio
2300022896	P17	383,40	-7,494722	-38,861944	Brejo santo	Sítio Lagoa da Vaca	Missão Velha	Médio
2300022906	P18	461,84	-7,270000	-39,379444	Crato	Baixio do Muquem (Escola)	Missão Velha	Médio
2300022907	P19	405,87	-7,206666	-39,399722	Crato	Sítio São Bento	Missão Velha	Médio
2300022908	P20	394,10	-7,296111	-39,270277	Barbalha	Sítio Buriti (EMBRAPA)	Missão Velha	Médio
2300022909	P21	348,70	-7,218611	-39,201111	Missão velha	Sítio Arraial de Cima	Missão Velha	Médio

Fonte: (RIMAS, 2021).

Tabela A.2 – Funções de ativação disponíveis por padrão nos pacotes `neuralnet` e `elmNNRcpp` para ANN e ELM, respectivamente

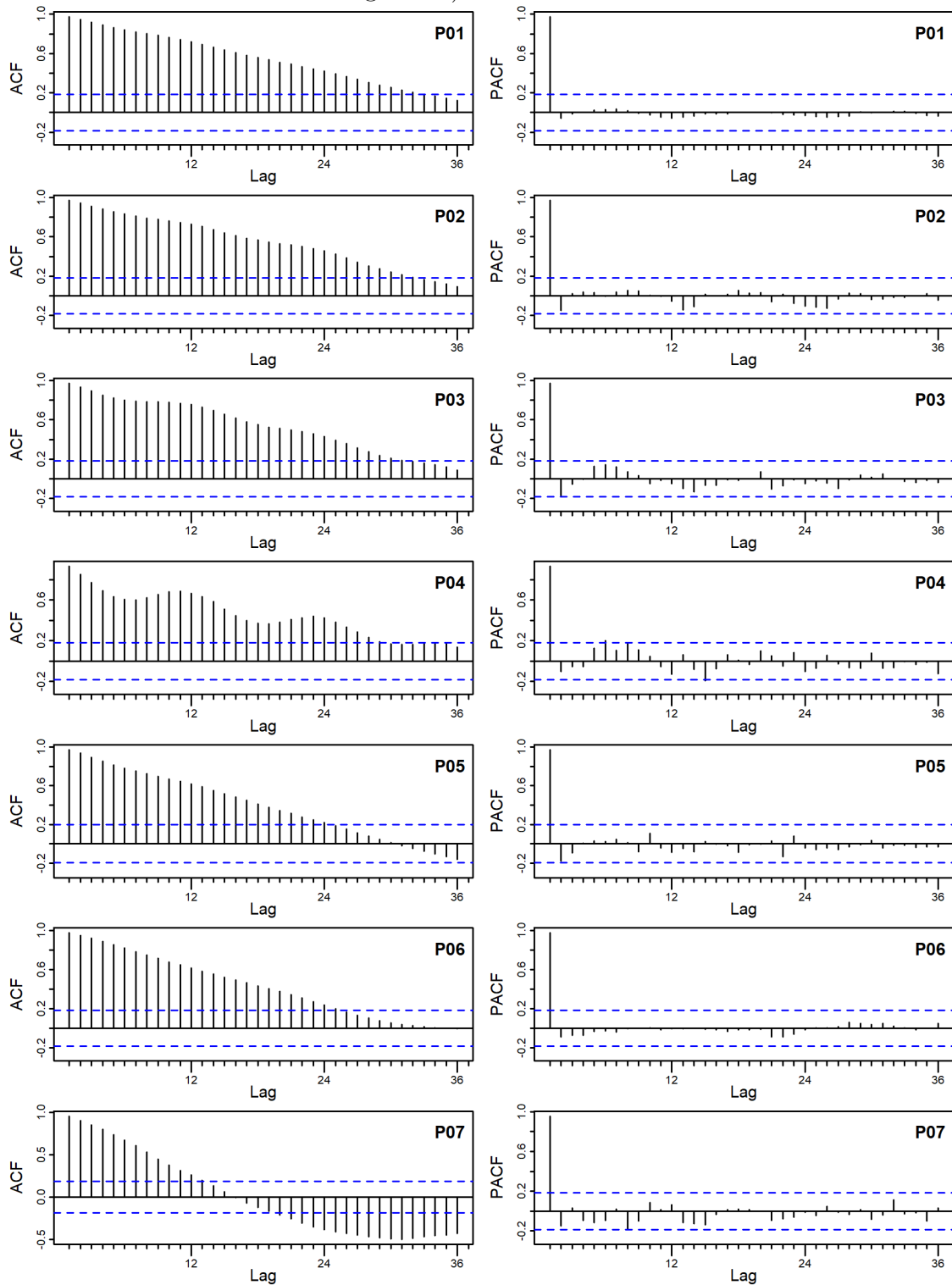
Função	neuralnet	elmNNRcpp	Equação
Logística	<code>logistic</code>	<code>sig</code>	$f(x) = 1/(1 + \exp(-x))$
Tangente hiperbólica	<code>tanh</code>	<code>tansig</code>	$f(x) = \sinh(x)/\cosh(x)$
Seno	N/A	<code>sin</code>	$f(x) = \sin(x)$
<i>Radial Basis</i>	N/A	<code>radbas</code>	$f(x) = \exp(-x^2)$
<i>Hard-limit</i>	N/A	<code>hardlim</code>	$f(x) = \begin{cases} 1, & \text{se } x \geq 0 \\ 0, & \text{caso contrário} \end{cases}$
<i>Hard-limit</i> simétrica	N/A	<code>hardlims</code>	$f(x) = \begin{cases} 1, & \text{se } x \geq 0 \\ -1, & \text{caso contrário} \end{cases}$
<i>Symmetric saturating linear</i>	N/A	<code>satlins</code>	$f(x) = \begin{cases} -1, & \text{se } x \leq -1 \\ x, & \text{se } -1 \leq x \leq 1 \\ 1, & \text{se } x \geq 1 \end{cases}$
<i>Triangular basis</i>	N/A	<code>tribas</code>	$f(x) = \begin{cases} 1 - x , & \text{se } -1 \leq x \leq 1 \\ 0, & \text{caso contrário} \end{cases}$
<i>Rectified linear unit</i>	N/A	<code>relu</code>	$f(x) = \begin{cases} x, & \text{se } x > 0 \\ 0, & \text{caso contrário} \end{cases}$
Linear	N/A	<code>purelin</code>	$f(x) = x$

Nota: N/A indica que um pacote não possui dada função.

Fonte: Elaborado pelo autor (2022). Disponível em: <https://www.mathworks.com/help/index.html> e <https://keras.io/api/layers/activations/>.

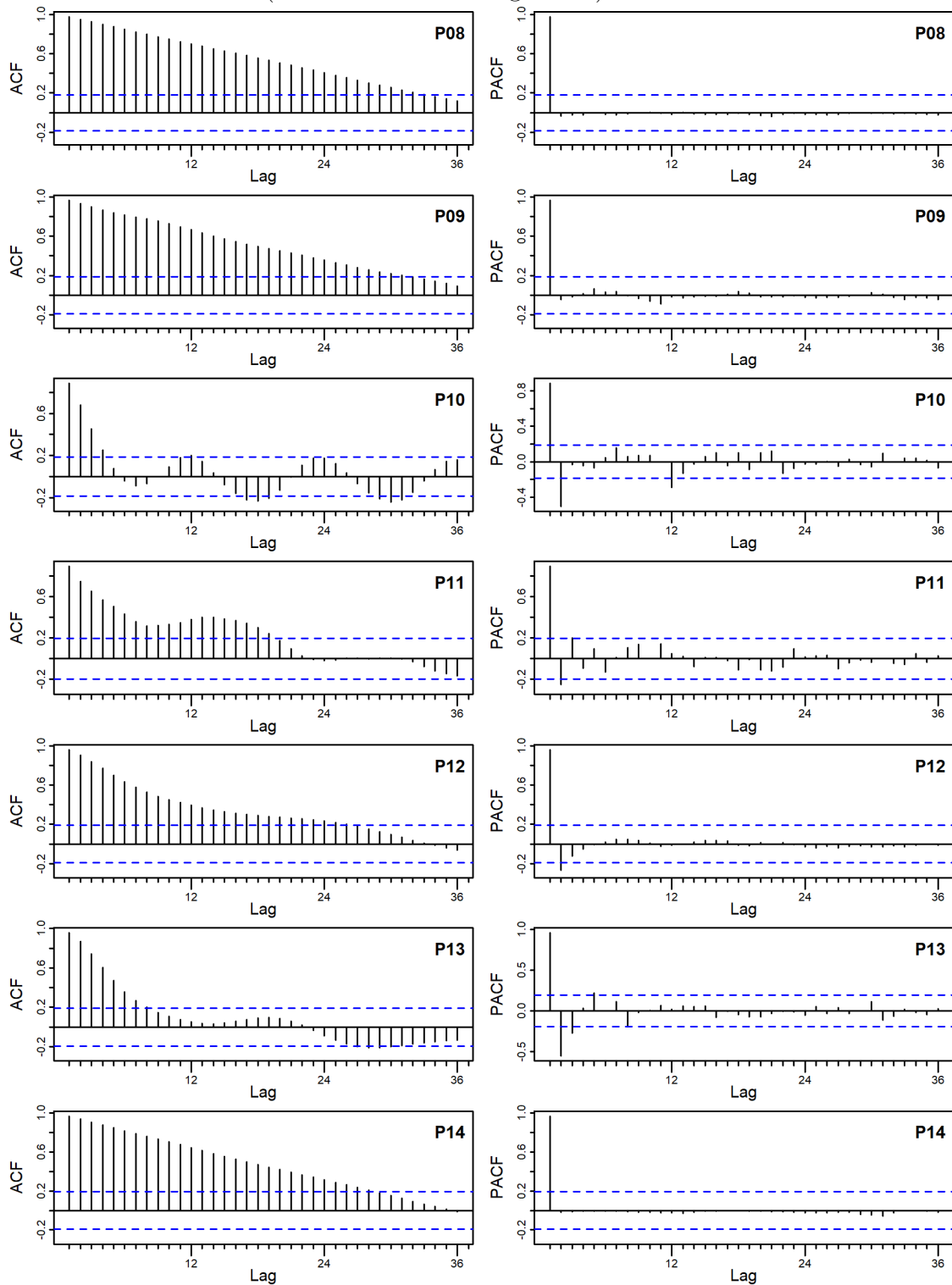
APÊNDICE B – ANÁLISE EXPLORATÓRIA DOS DADOS

Figura B.1 – Gráficos das Funções de Autocorrelação (FAC) e Autocorrelação Parcial (FACP) até o lag 36 das séries mensais de NAS (séries indicadas nos gráficos)



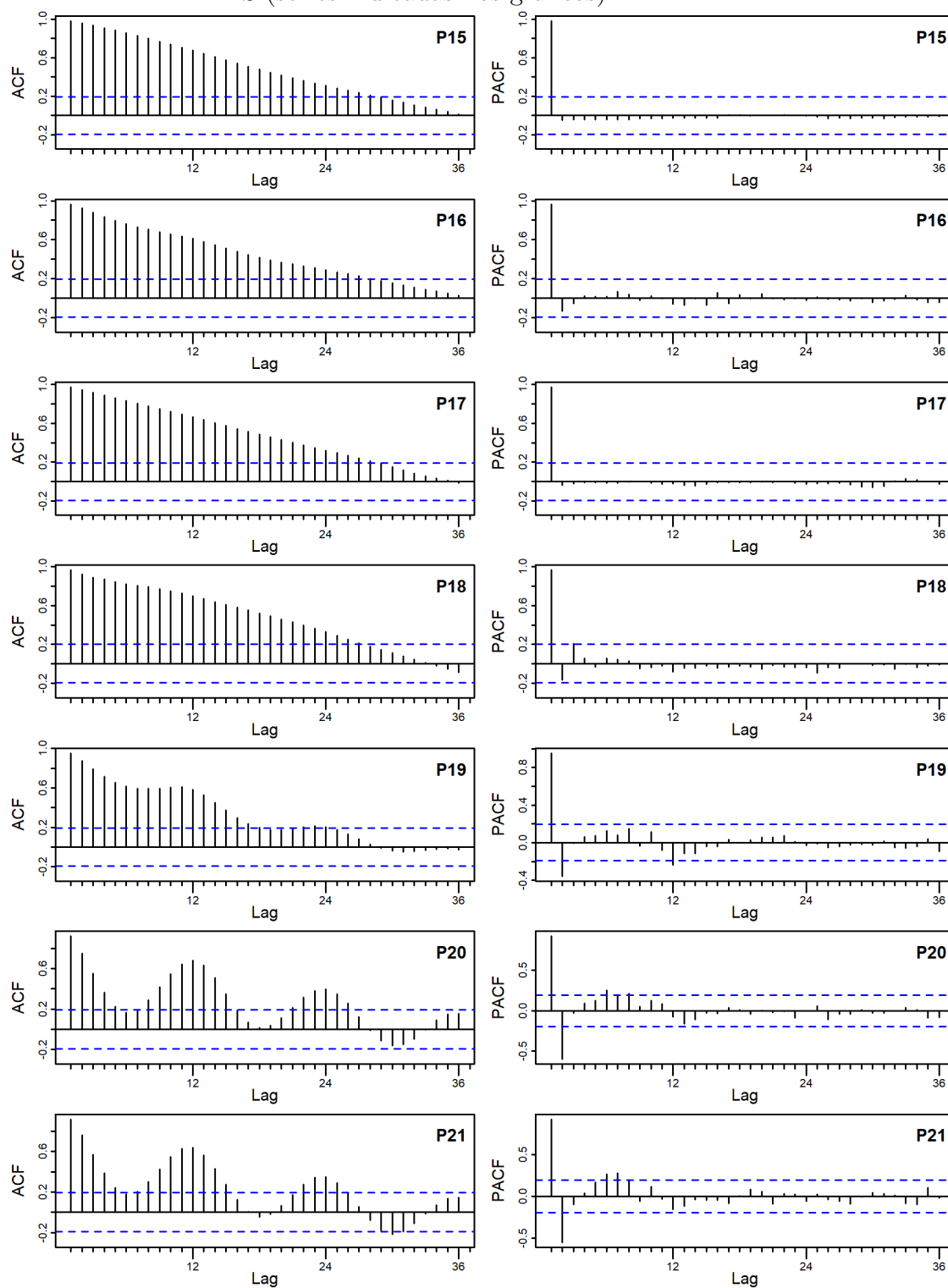
Nota: Continua na próxima página.

Figura B.1 – (Continuação) Gráficos das Funções de Autocorrelação (FAC) e Autocorrelação Parcial (FACP) até o lag 36 das séries mensais de NAS (séries indicadas nos gráficos)



Nota: Continua na próxima página.

Figura B.1 – (Continuação) Gráficos das Funções de Autocorrelação (FAC) e Autocorrelação Parcial (FACP) até o lag 36 das séries mensais de NAS (séries indicadas nos gráficos)



Fonte: Elaborado pelo autor (2022).

Tabela B.1 – Matriz de distâncias (em km) entre os poços da BSA

Poço	P02	P03	P04	P05	P06	P07	P08	P08	P09	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21
P01	42,9	37,2	10,7	39,5	8,5	15,3	23,0	10,4	30,3	73,4	48,8	59,8	53,7	34,3	27,9	47,8	14,5	20,1	3,0	12,8	
P02		10,0	38,4	16,3	50,6	27,7	27,6	32,6	18,3	39,6	25,4	31,2	12,1	22,9	16,3	5,5	57,4	62,7	45,3	44,3	
P03			30,9	22,7	44,2	22,1	18,6	27,0	8,9	38,4	18,5	27,0	17,0	12,9	9,4	12,3	51,7	56,2	39,3	36,3	
P04				39,3	14,2	12,4	13,8	9,5	22,7	64,6	39,5	50,6	47,9	25,2	22,2	42,6	22,6	25,8	10,9	6,3	
P05					48,0	26,9	33,5	30,7	27,3	55,7	40,7	47,4	27,0	34,1	22,2	20,9	52,9	59,3	42,4	45,6	
P06						22,9	28,0	18,2	36,5	78,8	53,7	64,8	61,0	39,4	35,0	55,3	8,5	12,1	5,6	12,1	
P07							12,5	4,9	16,2	59,4	35,7	46,4	38,5	21,8	12,8	32,5	29,8	35,0	17,7	18,7	
P08								15,0	9,8	50,8	25,9	36,9	35,4	11,4	11,9	30,9	36,3	39,5	24,2	18,3	
P09									20,5	63,8	39,7	50,5	43,4	25,5	17,6	37,4	24,9	30,2	12,8	15,5	
P10										43,3	19,7	30,2	25,7	7,3	5,8	21,2	44,4	48,4	32,0	27,8	
P11											25,2	14,3	29,1	39,4	47,1	34,7	87,1	90,1	74,8	68,3	
P12												11,1	22,7	14,5	24,7	23,5	62,1	64,9	50,0	43,0	
P13														23,9	25,5	34,7	27,5	73,2	76,0	61,1	54,0
P14															27,2	26,0	6,6	68,3	73,1	56,0	53,4
P15																13,1	24,3	47,7	50,8	35,6	29,1
P16																	20,4	42,4	47,1	30,1	28,0
P17																		62,3	67,4	50,1	48,3
P18																			7,4	12,4	20,5
P19																				17,4	22,0
P20																					11,5
P21																					

Fonte: Elaborado pelo autor (2022).

APÊNDICE C – DESCRIÇÃO DOS MODELOS

Tabela C.1 – Descrição dos modelos construídos para a série P01

Formalismo	Descrição
ARIMA	ARIMA(0,0,0)(2,1,0)[12]
ETS	ETS(A,Ad,A) $\alpha = 0,933$ $\beta = 0,071$ $\gamma = 0,067$
ANN	p=4, P=1, nN=21, nL=3, FA=logistic
SVR	p=1, P=1, $\gamma=0,191$, C=746,06, $\varepsilon=0,019$, nSV=33
ELM	p=4, P=2, nN=13, FA=satlins
LSTM	p=4, P=1, nN=21
cANN	nN=15, nL=2, FA=logistic
cSVR	$\gamma=0,193$, C=90,905, $\varepsilon=0,025$, nSV=15
cMV	$\omega_{ARIMA} = -1e-03$, $\omega_{ETS} = 0,146$, $\omega_{ANN} = -6e-02$, $\omega_{SVR} = 0,021$, $\omega_{ELM} = 0,699$, $\omega_{LSTM} = 0,195$
cCP	Normal($\rho_1=0,094$, $\rho_2=0,286$, $\rho_3=0,194$, $\rho_4=0,105$, $\rho_5=0,151$, $\rho_6=0,533$, $\rho_7=0,55$, $\rho_8=0,447$, $\rho_9=0,274$, $\rho_{10}=0,767$, $\rho_{11}=0,414$, $\rho_{12}=0,666$, $\rho_{13}=0,345$, $\rho_{14}=0,49$, $\rho_{15}=0,298$) $E_{ARIMA} \sim snorm(0,021$, $0,138, 1,683)$, $E_{ETS} \sim norm(4e-03, 0,027)$, $E_{ANN} \sim laplace(-1e-03$, $2e-02)$, $E_{SVR} \sim laplace(-1e-03, 0,025)$, $E_{ELM} \sim laplace(1e-03, 0,013)$, $E_{LSTM} \sim norm(0,014, 0,029)$

Nota: p, P – termos autorregressivos não sazonais e sazonais, respectivamente, nL – número de camadas ocultas, nN – número de nós nas camadas ocultas, FA – função de ativação, nSV – número de *Support Vectors*, ω_{modelo} – peso atribuído ao preditor, ρ_i – parâmetro da Cópula, E_{modelo} – resíduo do preditor.

Fonte: Elaborado pelo autor (2022).

Tabela C.2 – Descrição dos modelos construídos para a série P02

Formalismo	Descrição
ARIMA	ARIMA(0,1,2)(0,1,1)[12]
ETS	ETS(A,A,A) $\alpha = 0,952$ $\beta = 0,134$ $\gamma = 0,041$
ANN	p=4, P=1, nN=26, nL=3, FA=logistic
SVR	p=2, P=1, $\gamma=0,138$, C=201,626, $\varepsilon=0,049$, nSV=38
ELM	p=3, P=2, nN=25, FA=tansig
LSTM	p=4, P=1, nN=26
cANN	nN=13, nL=3, FA=logistic
cSVR	$\gamma=0,812$, C=602,442, $\varepsilon=0,043$, nSV=32
cMV	$\omega_{ARIMA} = 0,063$, $\omega_{ETS} = 0,296$, $\omega_{ANN} = 0,23$, $\omega_{SVR} = -0,222$, ω_{ELM} $= 0,504$, $\omega_{LSTM} = 0,129$
cCP	Normal($\rho_1=0,85$, $\rho_2=0,593$, $\rho_3=0,65$, $\rho_4=0,49$, $\rho_5=0,485$, $\rho_6=0,459$, $\rho_7=0,473$, $\rho_8=0,424$, $\rho_9=0,37$, $\rho_{10}=0,802$, $\rho_{11}=0,685$, $\rho_{12}=0,648$, $\rho_{13}=0,702$, $\rho_{14}=0,657$, $\rho_{15}=0,635$) $E_{ARIMA} \sim norm(-1e-03$, $0,052)$, $E_{ETS} \sim norm(0, 0,052)$, $E_{ANN} \sim norm(2e-03, 0,051)$, $E_{SVR} \sim norm(2e-03, 0,059)$, $E_{ELM} \sim norm(6e-03, 0,045)$, $E_{LSTM} \sim norm(-0,025, 0,055)$

Nota: p, P – termos autorregressivos não sazonais e sazonais, respectivamente, nL – número de camadas ocultas, nN – número de nós nas camadas ocultas, FA – função de ativação, nSV – número de *Support Vectors*, ω_{modelo} – peso atribuído ao preditor, ρ_i – parâmetro da Cópula, E_{modelo} – resíduo do preditor.

Fonte: Elaborado pelo autor (2022).

Tabela C.3 – Descrição dos modelos construídos para a série P03

Formalismo	Descrição
ARIMA	ARIMA(1,1,0)(0,1,1)[12]
ETS	ETS(A,N,A) $\alpha = 0,915$ $\gamma = 0$
ANN	p=3, P=3, nN=37, nL=2, FA=logistic
SVR	p=1, P=2, $\gamma=0,637$, C=932,147, $\varepsilon=0,056$, nSV=35
ELM	p=3, P=1, nN=17, FA=tansig
LSTM	p=3, P=1, nN=17
cANN	nN=15, nL=2, FA=logistic
cSVR	$\gamma=0,059$, C=323,686, $\varepsilon=3e-02$, nSV=34
cMV	$\omega_{ARIMA} = -0,076$, $\omega_{ETS} = 0,289$, $\omega_{ANN} = 0,656$, $\omega_{SVR} = 0,113$, $\omega_{ELM} = 0,106$, $\omega_{LSTM} = -0,089$
cCP	Normal($\rho_1=0,854$, $\rho_2=0,484$, $\rho_3=0,335$, $\rho_4=0,69$, $\rho_5=0,536$, $\rho_6=0,427$, $\rho_7=0,326$, $\rho_8=0,589$, $\rho_9=0,479$, $\rho_{10}=0,65$, $\rho_{11}=0,618$, $\rho_{12}=0,734$, $\rho_{13}=0,483$, $\rho_{14}=0,62$, $\rho_{15}=0,7$) $E_{ARIMA} \sim snorm(-3e-03$, $0,066, 0,668)$, $E_{ETS} \sim snorm(0,021, 0,064, 0,455)$, $E_{ANN} \sim norm(4e-$ $03, 0,048)$, $E_{SVR} \sim norm(-8e-03, 0,064)$, $E_{ELM} \sim norm(0,015, 0,061)$, $E_{LSTM} \sim norm(0,024, 0,063)$

Nota: p, P – termos autorregressivos não sazonais e sazonais, respectivamente, nL – número de camadas ocultas, nN – número de nós nas camadas ocultas, FA – função de ativação, nSV – número de *Support Vectors*, ω_{modelo} – peso atribuído ao preditor, ρ_i – parâmetro da Cópula, E_{modelo} – resíduo do preditor.

Fonte: Elaborado pelo autor (2022).

Tabela C.4 – Descrição dos modelos construídos para a série P04

Formalismo	Descrição
ARIMA	ARIMA(0,1,0)
ETS	ETS(A,N,A) $\alpha = 0,631$ $\gamma = 0$
ANN	p=3, P=2, nN=12, nL=3, FA=logistic
SVR	p=2, P=2, $\gamma=0,111$, C=268,47, $\varepsilon=0,077$, nSV=34
ELM	p=3, P=3, nN=8, FA=hardlims
LSTM	p=3, P=3, nN=8
cANN	nN=25, nL=3, FA=logistic
cSVR	$\gamma=0,182$, C=364,754, $\varepsilon=0,052$, nSV=12
cMV	$\omega_{ARIMA} = -0,26$, $\omega_{ETS} = 0,329$, $\omega_{ANN} = 0,481$, $\omega_{SVR} = 0,088$, ω_{ELM} $= -0,346$, $\omega_{LSTM} = 0,708$
cCP	Normal($\rho_1=0,848$, $\rho_2=0,353$, $\rho_3=0,587$, $\rho_4=0,515$, $\rho_5=0,586$, $\rho_6=0,42$, $\rho_7=0,633$, $\rho_8=0,588$, $\rho_9=0,579$, $\rho_{10}=0,335$, $\rho_{11}=0,328$, $\rho_{12}=0,178$, $\rho_{13}=0,783$, $\rho_{14}=0,68$, $\rho_{15}=0,695$) $E_{ARIMA} \sim laplace(0,023$, $0,162)$, $E_{ETS} \sim laplace(6e-02, 0,133)$, $E_{ANN} \sim cauchy(0, 0,012)$, $E_{SVR} \sim laplace(0,038, 0,144)$, $E_{ELM} \sim laplace(0,032, 0,173)$, $E_{LSTM} \sim laplace(3e-03, 0,115)$

Nota: p, P – termos autorregressivos não sazonais e sazonais, respectivamente, nL – número de camadas ocultas, nN – número de nós nas camadas ocultas, FA – função de ativação, nSV – número de *Support Vectors*, ω_{modelo} – peso atribuído ao preditor, ρ_i – parâmetro da Cópula, E_{modelo} – resíduo do preditor.

Fonte: Elaborado pelo autor (2022).

Tabela C.5 – Descrição dos modelos construídos para a série P05

Formalismo	Descrição
ARIMA	ARIMA(2,1,0)
ETS	ETS(A,Ad,N) $\alpha = 0,847$ $\beta = 0,706$
ANN	p=2, P=2, nN=10, nL=2, FA=logistic
SVR	p=1, P=2, $\gamma=0,695$, C=241,162, $\varepsilon=0,031$, nSV=30
ELM	p=3, P=1, nN=21, FA=tansig
LSTM	p=3, P=1, nN=21
cANN	nN=18, nL=3, FA=logistic
cSVR	$\gamma=0,049$, C=217,304, $\varepsilon=0,082$, nSV=3
cMV	$\omega_{ARIMA} = -0,199$, $\omega_{ETS} = 0,242$, $\omega_{ANN} = -0,103$, $\omega_{SVR} = 0,143$, $\omega_{ELM} = 0,645$, $\omega_{LSTM} = 0,272$
cCP	Normal($\rho_1=0,971$, $\rho_2=0,485$, $\rho_3=0,181$, $\rho_4=0,64$, $\rho_5=0,556$, $\rho_6=0,395$, $\rho_7=0,084$, $\rho_8=0,637$, $\rho_9=0,577$, $\rho_{10}=0,785$, $\rho_{11}=0,361$, $\rho_{12}=0,631$, $\rho_{13}=0,083$, $\rho_{14}=0,454$, $\rho_{15}=0,513$) $E_{ARIMA} \sim \text{laplace}(8e-03, 4e-02)$, $E_{ETS} \sim \text{laplace}(2e-03, 0,041)$, $E_{ANN} \sim \text{norm}(-4e-03, 0,056)$, $E_{SVR} \sim \text{cauchy}(2e-03, 0,031)$, $E_{ELM} \sim \text{snorm}(2e-03, 0,046, 0,464)$, $E_{LSTM} \sim \text{norm}(1e-03, 0,048)$

Nota: p, P – termos autorregressivos não sazonais e sazonais, respectivamente, nL – número de camadas ocultas, nN – número de nós nas camadas ocultas, FA – função de ativação, nSV – número de *Support Vectors*, ω_{modelo} – peso atribuído ao preditor, ρ_i – parâmetro da Cópula, E_{modelo} – resíduo do preditor.

Fonte: Elaborado pelo autor (2022).

Tabela C.6 – Descrição dos modelos construídos para a série P06

Formalismo	Descrição
ARIMA	ARIMA(5,1,1)
ETS	ETS(A,A,N) $\alpha = 0,995$ $\beta = 0,59$
ANN	p=4, P=2, nN=11, nL=3, FA=logistic
SVR	p=5, P=3, $\gamma=0,281$, C=659,969, $\varepsilon=0,057$, nSV=10
ELM	p=2, P=3, nN=25, FA=satlins
LSTM	p=2, P=3, nN=25
cANN	nN=32, nL=3, FA=logistic
cSVR	$\gamma=0,599$, C=474,329, $\varepsilon=0,028$, nSV=22
cMV	$\omega_{ARIMA} = 0,463$, $\omega_{ETS} = -0,098$, $\omega_{ANN} = -0,118$, $\omega_{SVR} = 0,24$, $\omega_{ELM} = 0,377$, $\omega_{LSTM} = 0,137$
cCP	Normal($\rho_1=0,877$, $\rho_2=0,698$, $\rho_3=0,118$, $\rho_4=0,504$, $\rho_5=0,201$, $\rho_6=0,667$, $\rho_7=-0,076$, $\rho_8=0,514$, $\rho_9=0,365$, $\rho_{10}=0,387$, $\rho_{11}=0,67$, $\rho_{12}=0,35$, $\rho_{13}=0,27$, $\rho_{14}=0,042$, $\rho_{15}=0,401$) $E_{ARIMA} \sim \text{laplace}(4e-03, 0,028)$, $E_{ETS} \sim \text{laplace}(4e-03, 0,034)$, $E_{ANN} \sim \text{laplace}(-1e-03, 0,026)$, $E_{SVR} \sim \text{norm}(2e-03, 5e-02)$, $E_{ELM} \sim \text{laplace}(-4e-03, 0,024)$, $E_{LSTM} \sim \text{laplace}(0,019, 0,039)$

Nota: p, P – termos autorregressivos não sazonais e sazonais, respectivamente, nL – número de camadas ocultas, nN – número de nós nas camadas ocultas, FA – função de ativação, nSV – número de *Support Vectors*, ω_{modelo} – peso atribuído ao preditor, ρ_i – parâmetro da Cópula, E_{modelo} – resíduo do preditor.

Fonte: Elaborado pelo autor (2022).

Tabela C.7 – Descrição dos modelos construídos para a série P07

Formalismo	Descrição
ARIMA	ARIMA(0,1,1)(1,0,0)[12]
ETS	ETS(A,N,N) $\alpha = 1$
ANN	p=3, P=2, nN=26, nL=3, FA=logistic
SVR	p=2, P=2, $\gamma=0,848$, C=932,749, $\varepsilon=0,018$, nSV=48
ELM	p=1, P=2, nN=27, FA=relu
LSTM	p=1, P=2, nN=27
cANN	nN=11, nL=3, FA=logistic
cSVR	$\gamma=0,565$, C=416,969, $\varepsilon=2e-02$, nSV=30
cmV	$\omega_{ARIMA} = -0,064$, $\omega_{ETS} = 0,079$, $\omega_{ANN} = 0,43$, $\omega_{SVR} = 0,725$, $\omega_{ELM} = 7e-03$, $\omega_{LSTM} = -0,175$
cCP	Normal($\rho_1=0,862$, $\rho_2=0,46$, $\rho_3=0,392$, $\rho_4=0,579$, $\rho_5=0,552$, $\rho_6=0,344$, $\rho_7=0,307$, $\rho_8=0,634$, $\rho_9=0,64$, $\rho_{10}=0,675$, $\rho_{11}=0,335$, $\rho_{12}=0,507$, $\rho_{13}=0,469$, $\rho_{14}=0,57$, $\rho_{15}=0,664$) $E_{ARIMA} \sim cauchy(7e-03, 0,029)$, $E_{ETS} \sim cauchy(8e-03, 0,039)$, $E_{ANN} \sim laplace(-1e-03, 2e-02)$, $E_{SVR} \sim laplace(-3e-03, 0,019)$, $E_{ELM} \sim cauchy(4e-03, 0,018)$, $E_{LSTM} \sim laplace(0,012, 0,046)$

Nota: p, P – termos autorregressivos não sazonais e sazonais, respectivamente, nL – número de camadas ocultas, nN – número de nós nas camadas ocultas, FA – função de ativação, nSV – número de *Support Vectors*, ω_{modelo} – peso atribuído ao preditor, ρ_i – parâmetro da Cópula, E_{modelo} – resíduo do preditor.

Fonte: Elaborado pelo autor (2022).

Tabela C.8 – Descrição dos modelos construídos para a série P08

Formalismo	Descrição
ARIMA	ARIMA(1,1,0)
ETS	ETS(A,A,N) $\alpha = 1$ $\beta = 0,841$
ANN	p=2, P=2, nN=30, nL=2, FA=logistic
SVR	p=2, P=2, $\gamma=0,466$, C=281,463, $\varepsilon=0,034$, nSV=15
ELM	p=2, P=3, nN=21, FA=hardlims
LSTM	p=2, P=3, nN=21
cANN	nN=5, nL=3, FA=logistic
cSVR	$\gamma=0,479$, C=719,376, $\varepsilon=0,066$, nSV=6
cmV	$\omega_{ARIMA} = 0,324$, $\omega_{ETS} = 0,127$, $\omega_{ANN} = -0,26$, $\omega_{SVR} = 0,419$, $\omega_{ELM} = 1e-02$, $\omega_{LSTM} = 0,38$
cCP	Normal($\rho_1=0,903$, $\rho_2=0,501$, $\rho_3=0,256$, $\rho_4=0,114$, $\rho_5=0,027$, $\rho_6=0,421$, $\rho_7=0,258$, $\rho_8=-0,021$, $\rho_9=-0,071$, $\rho_{10}=0,695$, $\rho_{11}=0,245$, $\rho_{12}=0,515$, $\rho_{13}=0,167$, $\rho_{14}=0,396$, $\rho_{15}=0,374$) $E_{ARIMA} \sim cauchy(3e-03, 7e-03)$, $E_{ETS} \sim cauchy(0, 6e-03)$, $E_{ANN} \sim laplace(-2e-03, 0,017)$, $E_{SVR} \sim norm(-5e-03, 0,023)$, $E_{ELM} \sim snorm(2e-03, 0,058, 1,546)$, $E_{LSTM} \sim laplace(1e-03, 0,018)$

Nota: p, P – termos autorregressivos não sazonais e sazonais, respectivamente, nL – número de camadas ocultas, nN – número de nós nas camadas ocultas, FA – função de ativação, nSV – número de *Support Vectors*, ω_{modelo} – peso atribuído ao preditor, ρ_i – parâmetro da Cópula, E_{modelo} – resíduo do preditor.

Fonte: Elaborado pelo autor (2022).

Tabela C.9 – Descrição dos modelos construídos para a série P09

Formalismo	Descrição
ARIMA	ARIMA(0,0,0)(1,1,0)[12]
ETS	ETS(A,Ad,N) $\alpha = 1$ $\beta = 0$, 958
ANN	p=4, P=2, nN=12, nL=3, FA=tanh
SVR	p=2, P=3, $\gamma=0,642$, C=818,991, $\varepsilon=0,021$, nSV=19
ELM	p=2, P=2, nN=35, FA=hardlim
LSTM	p=4, P=2, nN=12
cANN	nN=15, nL=2, FA=logistic
cSVR	$\gamma=0,138$, C=339,591, $\varepsilon=0,014$, nSV=16
cMV	$\omega_{ARIMA} = 6e-03$, $\omega_{ETS} = 0,134$, $\omega_{ANN} = 0,266$, $\omega_{SVR} = 0,363$, $\omega_{ELM} = -0,026$, $\omega_{LSTM} = 0,257$
cCP	Clayton($\rho_1=0,117$) $E_{ARIMA} \sim norm(0,075, 0,183)$, $E_{ETS} \sim laplace(2e-03, 0,019)$, $E_{ANN} \sim cauchy(2e-03, 7e-03)$, $E_{SVR} \sim norm(-2e-03, 0,023)$, $E_{ELM} \sim laplace(0, 5e-02)$, $E_{LSTM} \sim laplace(0,014, 0,018)$

Nota: p, P – termos autorregressivos não sazonais e sazonais, respectivamente, nL – número de camadas ocultas, nN – número de nós nas camadas ocultas, FA – função de ativação, nSV – número de *Support Vectors*, ω_{modelo} – peso atribuído ao preditor, ρ_i – parâmetro da Cópula, E_{modelo} – resíduo do preditor.

Fonte: Elaborado pelo autor (2022).

Tabela C.10 – Descrição dos modelos construídos para a série P10

Formalismo	Descrição
ARIMA	ARIMA(1,0,1)(0,1,1)[12]
ETS	ETS(A,N,A) $\alpha = 0,998$ $\gamma = 0$
ANN	p=4, P=2, nN=16, nL=1, FA=logistic
SVR	p=3, P=3, $\gamma=0,841$, C=813,85, $\varepsilon=0,045$, nSV=33
ELM	p=3, P=2, nN=6, FA=tribas
LSTM	p=4, P=2, nN=16
cANN	nN=21, nL=4, FA=logistic
cSVR	$\gamma=0,600$, C=285,896, $\varepsilon=0,074$, nSV=14
cMV	$\omega_{ARIMA} = -0,137$, $\omega_{ETS} = 0,134$, $\omega_{ANN} = 0,21$, $\omega_{SVR} = 0,416$, $\omega_{ELM} = -7e-03$, $\omega_{LSTM} = 0,384$
cCP	Normal($\rho_1=0,854$, $\rho_2=0,45$, $\rho_3=0,241$, $\rho_4=0,256$, $\rho_5=0,261$, $\rho_6=0,333$, $\rho_7=0,149$, $\rho_8=0,214$, $\rho_9=0,285$, $\rho_{10}=0,159$, $\rho_{11}=0,076$, $\rho_{12}=0,374$, $\rho_{13}=0,159$, $\rho_{14}=0,125$, $\rho_{15}=0,233$) $E_{ARIMA} \sim laplace(3e-03, 0,091)$, $E_{ETS} \sim laplace(0,012, 0,093)$, $E_{ANN} \sim laplace(4e-03, 0,063)$, $E_{SVR} \sim snorm(-4e-02, 0,065, 1e-03)$, $E_{ELM} \sim snorm(0,038, 0,283, 1,508)$, $E_{LSTM} \sim laplace(-0,042, 0,057)$

Nota: p, P – termos autorregressivos não sazonais e sazonais, respectivamente, nL – número de camadas ocultas, nN – número de nós nas camadas ocultas, FA – função de ativação, nSV – número de *Support Vectors*, ω_{modelo} – peso atribuído ao preditor, ρ_i – parâmetro da Cópula, E_{modelo} – resíduo do preditor.

Fonte: Elaborado pelo autor (2022).

Tabela C.11 – Descrição dos modelos construídos para a série P11

Formalismo	Descrição
ARIMA	ARIMA(1,1,2)
ETS	ETS(A,N,N) $\alpha = 1$
ANN	p=4, P=1, nN=26, nL=3, FA=logistic
SVR	p=3, P=1, $\gamma=0,721$, C=438,352, $\varepsilon=0,079$, nSV=40
ELM	p=4, P=2, nN=14, FA=hardlims
LSTM	p=4, P=1, nN=26
cANN	nN=22, nL=2, FA=tanh
cSVR	$\gamma=0,226$, C=865,622, $\varepsilon=5e-03$, nSV=50
cMV	$\omega_{ARIMA} = 0,021$, $\omega_{ETS} = -0,029$, $\omega_{ANN} = 0,885$, $\omega_{SVR} = 0,026$, $\omega_{ELM} = 0,012$, $\omega_{LSTM} = 0,085$
cCP	Normal($\rho_1=0,844$, $\rho_2=0,445$, $\rho_3=0,584$, $\rho_4=0,394$, $\rho_5=0,315$, $\rho_6=0,34$, $\rho_7=0,561$, $\rho_8=0,474$, $\rho_9=0,272$, $\rho_{10}=0,645$, $\rho_{11}=0,19$, $\rho_{12}=0,673$, $\rho_{13}=0,189$, $\rho_{14}=0,576$, $\rho_{15}=0,302$) $E_{ARIMA} \sim cauchy(0,013, 7e-02)$, $E_{ETS} \sim laplace(-3e-03, 0,171)$, $E_{ANN} \sim laplace(2e-03, 0,035)$, $E_{SVR} \sim snorm(1e-03, 0,075, 0,466)$, $E_{ELM} \sim norm(0,039, 0,223)$, $E_{LSTM} \sim laplace(9e-03, 0,046)$

Nota: p, P – termos autorregressivos não sazonais e sazonais, respectivamente, nL – número de camadas ocultas, nN – número de nós nas camadas ocultas, FA – função de ativação, nSV – número de *Support Vectors*, ω_{modelo} – peso atribuído ao preditor, ρ_i – parâmetro da Cópula, E_{modelo} – resíduo do preditor.

Fonte: Elaborado pelo autor (2022).

Tabela C.12 – Descrição dos modelos construídos para a série P12

Formalismo	Descrição
ARIMA	ARIMA(1,1,1)
ETS	ETS(A,Ad,N) $\alpha = 1$ $\beta = 0,34$
ANN	p=2, P=2, nN=28, nL=2, FA=logistic
SVR	p=3, P=2, $\gamma=0,172$, C=393,936, $\varepsilon=0,046$, nSV=23
ELM	p=4, P=1, nN=19, FA=tansig
LSTM	p=2, P=2, nN=28
cANN	nN=17, nL=2, FA=logistic
cSVR	$\gamma=0,024$, C=452,039, $\varepsilon=0,029$, nSV=23
cMV	$\omega_{ARIMA} = -0,156$, $\omega_{ETS} = -7e-03$, $\omega_{ANN} = 0,972$, $\omega_{SVR} = 0,308$, $\omega_{ELM} = -0,094$, $\omega_{LSTM} = -0,023$
cCP	Normal($\rho_1=0,636$, $\rho_2=0,609$, $\rho_3=0,651$, $\rho_4=0,677$, $\rho_5=0,282$, $\rho_6=0,397$, $\rho_7=0,561$, $\rho_8=0,426$, $\rho_9=0,078$, $\rho_{10}=0,639$, $\rho_{11}=0,698$, $\rho_{12}=0,583$, $\rho_{13}=0,693$, $\rho_{14}=0,372$, $\rho_{15}=0,519$) $E_{ARIMA} \sim cauchy(0,011, 0,022)$, $E_{ETS} \sim cauchy(0,025, 2e-02)$, $E_{ANN} \sim cauchy(-3e-03, 0,014)$, $E_{SVR} \sim snorm(0,011, 0,047, 3,882)$, $E_{ELM} \sim laplace(3e-03, 4e-02)$, $E_{LSTM} \sim laplace(-0,025, 0,047)$

Nota: p, P – termos autorregressivos não sazonais e sazonais, respectivamente, nL – número de camadas ocultas, nN – número de nós nas camadas ocultas, FA – função de ativação, nSV – número de *Support Vectors*, ω_{modelo} – peso atribuído ao preditor, ρ_i – parâmetro da Cópula, E_{modelo} – resíduo do preditor.

Fonte: Elaborado pelo autor (2022).

Tabela C.13 – Descrição dos modelos construídos para a série P13

Formalismo	Descrição
ARIMA	ARIMA(3,1,1)
ETS	ETS(A,Ad,N) $\alpha = 0,711$ $\beta = 0,711$
ANN	p=3, P=2, nN=32, nL=3, FA=logistic
SVR	p=5, P=3, $\gamma=0,950$, C=608,511, $\varepsilon=0,056$, nSV=19
ELM	p=3, P=2, nN=11, FA=tansig
LSTM	p=3, P=2, nN=32
cANN	nN=28, nL=2, FA=logistic
cSVR	$\gamma=0,180$, C=2,665, $\varepsilon=0,095$, nSV=7
cMV	$\omega_{ARIMA} = 0,435$, $\omega_{ETS} = -0,524$, $\omega_{ANN} = 0,649$, $\omega_{SVR} = 0$, $\omega_{ELM} = -0,259$, $\omega_{LSTM} = 0,699$
cCP	Normal($\rho_1=0,951$, $\rho_2=0,457$, $\rho_3=0,192$, $\rho_4=0,741$, $\rho_5=0,689$, $\rho_6=0,508$, $\rho_7=0,162$, $\rho_8=0,766$, $\rho_9=0,747$, $\rho_{10}=0,616$, $\rho_{11}=0,686$, $\rho_{12}=0,607$, $\rho_{13}=0,57$, $\rho_{14}=0,305$, $\rho_{15}=0,757$) $E_{ARIMA} \sim \text{laplace}(2e-02, 0,092)$, $E_{ETS} \sim \text{laplace}(0,013, 0,095)$, $E_{ANN} \sim \text{laplace}(0, 0,059)$, $E_{SVR} \sim \text{cauchy}(9e-03, 0,055)$, $E_{ELM} \sim \text{laplace}(1e-03, 0,083)$, $E_{LSTM} \sim \text{norm}(-3e-03, 0,085)$

Nota: p, P – termos autorregressivos não sazonais e sazonais, respectivamente, nL – número de camadas ocultas, nN – número de nós nas camadas ocultas, FA – função de ativação, nSV – número de *Support Vectors*, ω_{modelo} – peso atribuído ao preditor, ρ_i – parâmetro da Cópula, E_{modelo} – resíduo do preditor.

Fonte: Elaborado pelo autor (2022).

Tabela C.14 – Descrição dos modelos construídos para a série P14

Formalismo	Descrição
ARIMA	ARIMA(2,2,1)(0,0,1)[12]
ETS	ETS(A,A,N) $\alpha = 1$ $\beta = 1$
ANN	p=3, P=2, nN=22, nL=2, FA=logistic
SVR	p=1, P=3, $\gamma=0,028$, C=930,892, $\varepsilon=7e-02$, nSV=2
ELM	p=3, P=2, nN=15, FA=sig
LSTM	p=3, P=2, nN=22
cANN	nN=18, nL=2, FA=logistic
cSVR	$\gamma=0,022$, C=826,358, $\varepsilon=6e-03$, nSV=21
cMV	$\omega_{ARIMA} = 0,452$, $\omega_{ETS} = 0,166$, $\omega_{ANN} = 0,207$, $\omega_{SVR} = 0,073$, $\omega_{ELM} = 0,114$, $\omega_{LSTM} = -0,011$
cCP	Normal($\rho_1=0,815$, $\rho_2=0,133$, $\rho_3=-0,217$, $\rho_4=0,491$, $\rho_5=-0,044$, $\rho_6=0,035$, $\rho_7=-0,211$, $\rho_8=0,298$, $\rho_9=-0,069$, $\rho_{10}=-0,137$, $\rho_{11}=0,126$, $\rho_{12}=0,172$, $\rho_{13}=0,302$, $\rho_{14}=0,338$, $\rho_{15}=0,059$) $E_{ARIMA} \sim \text{laplace}(2e-03, 7e-03)$, $E_{ETS} \sim \text{laplace}(-1e-03, 9e-03)$, $E_{ANN} \sim \text{norm}(2e-03, 0,019)$, $E_{SVR} \sim \text{snorm}(2e-02, 0,042, 0,367)$, $E_{ELM} \sim \text{snorm}(5e-03, 0,013, 1,705)$, $E_{LSTM} \sim \text{snorm}(0,026, 0,042, 1e-03)$

Nota: p, P – termos autorregressivos não sazonais e sazonais, respectivamente, nL – número de camadas ocultas, nN – número de nós nas camadas ocultas, FA – função de ativação, nSV – número de *Support Vectors*, ω_{modelo} – peso atribuído ao preditor, ρ_i – parâmetro da Cópula, E_{modelo} – resíduo do preditor.

Fonte: Elaborado pelo autor (2022).

Tabela C.15 – Descrição dos modelos construídos para a série P15

Formalismo	Descrição
ARIMA	ARIMA(1,1,0)
ETS	ETS(A,A,N) $\alpha = 1 \beta = 1$
ANN	p=4, P=1, nN=12, nL=3, FA=tanh
SVR	p=1, P=3, $\gamma=0,077$, C=373,596, $\varepsilon=0,055$, nSV=2
ELM	p=4, P=3, nN=36, FA=tansig
LSTM	p=4, P=1, nN=12
cANN	nN=38, nL=2, FA=logistic
cSVR	$\gamma=0,556$, C=408,463, $\varepsilon=4e-03$, nSV=19
cMV	$\omega_{ARIMA} = 6,667$, $\omega_{ETS} = -5,889$, $\omega_{ANN} = 0,244$, $\omega_{SVR} = 0,21$, $\omega_{ELM} = 0,087$, $\omega_{LSTM} = -0,319$
cCP	Clayton($\rho_1=0,433$) $E_{ARIMA} \sim \text{laplace}(1e-03, 4e-03)$, $E_{ETS} \sim \text{laplace}(0, 5e-03)$, $E_{ANN} \sim \text{laplace}(-1e-03, 8e-03)$, $E_{SVR} \sim \text{snorm}(1e-03, 3e-02, 1,575)$, $E_{ELM} \sim \text{cauchy}(1e-03, 3e-03)$, $E_{LSTM} \sim \text{norm}(0,023, 0,024)$

Nota: p, P – termos autorregressivos não sazonais e sazonais, respectivamente, nL – número de camadas ocultas, nN – número de nós nas camadas ocultas, FA – função de ativação, nSV – número de *Support Vectors*, ω_{modelo} – peso atribuído ao preditor, ρ_i – parâmetro da Cópula, E_{modelo} – resíduo do preditor.

Fonte: Elaborado pelo autor (2022).

Tabela C.16 – Descrição dos modelos construídos para a série P16

Formalismo	Descrição
ARIMA	ARIMA(2,1,1)
ETS	ETS(A,A,N) $\alpha = 1 \beta = 0$
ANN	p=3, P=1, nN=21, nL=2, FA=tanh
SVR	p=1, P=3, $\gamma=0,355$, C=714,911, $\varepsilon=0,033$, nSV=13
ELM	p=1, P=2, nN=25, FA=hardlims
LSTM	p=1, P=2, nN=25
cANN	nN=24, nL=2, FA=logistic
cSVR	$\gamma=5e-03$, C=778,921, $\varepsilon=0,034$, nSV=6
cMV	$\omega_{ARIMA} = 0,351$, $\omega_{ETS} = -0,468$, $\omega_{ANN} = 0,439$, $\omega_{SVR} = 0,155$, $\omega_{ELM} = 0,034$, $\omega_{LSTM} = 0,49$
cCP	Normal($\rho_1=0,866$, $\rho_2=0,226$, $\rho_3=0,375$, $\rho_4=0,228$, $\rho_5=0,701$, $\rho_6=0,292$, $\rho_7=0,403$, $\rho_8=0,441$, $\rho_9=0,763$, $\rho_{10}=0,224$, $\rho_{11}=0,156$, $\rho_{12}=0,318$, $\rho_{13}=0,227$, $\rho_{14}=0,658$, $\rho_{15}=0,246$) $E_{ARIMA} \sim \text{laplace}(1e-02, 0,037)$, $E_{ETS} \sim \text{laplace}(-2e-03, 0,039)$, $E_{ANN} \sim \text{cauchy}(1e-03, 8e-03)$, $E_{SVR} \sim \text{snorm}(-7e-03, 0,039, 0,548)$, $E_{ELM} \sim \text{laplace}(0, 0,069)$, $E_{LSTM} \sim \text{laplace}(-2e-03, 0,026)$

Nota: p, P – termos autorregressivos não sazonais e sazonais, respectivamente, nL – número de camadas ocultas, nN – número de nós nas camadas ocultas, FA – função de ativação, nSV – número de *Support Vectors*, ω_{modelo} – peso atribuído ao preditor, ρ_i – parâmetro da Cópula, E_{modelo} – resíduo do preditor.

Fonte: Elaborado pelo autor (2022).

Tabela C.17 – Descrição dos modelos construídos para a série P17

Formalismo	Descrição
ARIMA	ARIMA(2,2,1)
ETS	ETS(A,Ad,N) $\alpha = 1 \beta = 1$
ANN	p=2, P=1, nN=12, nL=3, FA=logistic
SVR	p=2, P=1, $\gamma=0,033$, C=147,714, $\varepsilon=0,066$, nSV=4
ELM	p=3, P=2, nN=21, FA=hardlim
LSTM	p=2, P=1, nN=12
cANN	nN=30, nL=1, FA=logistic
cSVR	$\gamma=1e-02$, C=552,469, $\varepsilon=2e-03$, nSV=51
cMV	$\omega_{ARIMA} = 0,813$, $\omega_{ETS} = 0$, $\omega_{ANN} = 0,149$, $\omega_{SVR} = -0,105$, $\omega_{ELM} = 0,065$, $\omega_{LSTM} = 0,078$
cCP	Normal($\rho_1=0,827$, $\rho_2=0,139$, $\rho_3=0,179$, $\rho_4=0,053$, $\rho_5=0,241$, $\rho_6=0,027$, $\rho_7=0,134$, $\rho_8=0,054$, $\rho_9=0,358$, $\rho_{10}=0,555$, $\rho_{11}=0,374$, $\rho_{12}=0,389$, $\rho_{13}=0,266$, $\rho_{14}=0,619$, $\rho_{15}=0,115$) $E_{ARIMA} \sim \text{laplace}(1e-03, 0,013)$, $E_{ETS} \sim \text{laplace}(4e-03, 0,016)$, $E_{ANN} \sim \text{laplace}(-2e-03, 0,029)$, $E_{SVR} \sim \text{snorm}(-0,019, 0,036, 684,375)$, $E_{ELM} \sim \text{laplace}(0, 0,043)$, $E_{LSTM} \sim \text{snorm}(-0,012, 0,034, 2,945)$

Nota: p, P – termos autorregressivos não sazonais e sazonais, respectivamente, nL – número de camadas ocultas, nN – número de nós nas camadas ocultas, FA – função de ativação, nSV – número de *Support Vectors*, ω_{modelo} – peso atribuído ao preditor, ρ_i – parâmetro da Cópula, E_{modelo} – resíduo do preditor.

Fonte: Elaborado pelo autor (2022).

Tabela C.18 – Descrição dos modelos construídos para a série P18

Formalismo	Descrição
ARIMA	ARIMA(0,1,1)(2,1,0)[12]
ETS	ETS(A,Ad,N) $\alpha = 1 \beta = 0,228$
ANN	p=3, P=2, nN=28, nL=2, FA=logistic
SVR	p=3, P=2, $\gamma=0,646$, C=470,937, $\varepsilon=0,029$, nSV=26
ELM	p=2, P=2, nN=29, FA=tansig
LSTM	p=2, P=2, nN=29
cANN	nN=18, nL=3, FA=tanh
cSVR	$\gamma=0,015$, C=434,187, $\varepsilon=0,027$, nSV=23
cMV	$\omega_{ARIMA} = -0,181$, $\omega_{ETS} = 0,047$, $\omega_{ANN} = 0,045$, $\omega_{SVR} = 0,965$, $\omega_{ELM} = 0,123$, $\omega_{LSTM} = 2e-03$
cCP	Normal($\rho_1=0,783$, $\rho_2=0,491$, $\rho_3=0,637$, $\rho_4=0,657$, $\rho_5=0,533$, $\rho_6=0,391$, $\rho_7=0,541$, $\rho_8=0,373$, $\rho_9=0,696$, $\rho_{10}=0,594$, $\rho_{11}=0,581$, $\rho_{12}=0,336$, $\rho_{13}=0,528$, $\rho_{14}=0,417$, $\rho_{15}=0,419$) $E_{ARIMA} \sim \text{norm}(-1e-03, 5e-02)$, $E_{ETS} \sim \text{norm}(0,013, 0,058)$, $E_{ANN} \sim \text{norm}(3e-03, 0,041)$, $E_{SVR} \sim \text{norm}(0, 0,028)$, $E_{ELM} \sim \text{laplace}(-4e-03, 0,036)$, $E_{LSTM} \sim \text{norm}(-0,011, 0,065)$

Nota: p, P – termos autorregressivos não sazonais e sazonais, respectivamente, nL – número de camadas ocultas, nN – número de nós nas camadas ocultas, FA – função de ativação, nSV – número de *Support Vectors*, ω_{modelo} – peso atribuído ao preditor, ρ_i – parâmetro da Cópula, E_{modelo} – resíduo do preditor.

Fonte: Elaborado pelo autor (2022).

Tabela C.19 – Descrição dos modelos construídos para a série P19

Formalismo	Descrição
ARIMA	ARIMA(0,1,0)(2,1,0)[12]
ETS	ETS(A,Ad,N) $\alpha = 0,999$ $\beta = 0,792$
ANN	p=3, P=1, nN=3, nL=3, FA=tanh
SVR	p=3, P=1, $\gamma=0,754$, C=801,038, $\varepsilon=3e-02$, nSV=47
ELM	p=3, P=3, nN=26, FA=tribas
LSTM	p=3, P=3, nN=26
cANN	nN=15, nL=3, FA=tanh
cSVR	$\gamma=0,327$, C=231,822, $\varepsilon=0,029$, nSV=17
cMV	$\omega_{ARIMA} = -0,186$, $\omega_{ETS} = 3e-02$, $\omega_{ANN} = 0,119$, $\omega_{SVR} = 0,092$, $\omega_{ELM} = 0,139$, $\omega_{LSTM} = 0,806$
cCP	Clayton($\rho_1=0,53$) $E_{ARIMA} \sim \text{laplace}(-0,011, 0,068)$, $E_{ETS} \sim \text{laplace}(0,015, 0,081)$, $E_{ANN} \sim \text{laplace}(-0,012, 0,055)$, $E_{SVR} \sim \text{laplace}(0, 0,049)$, $E_{ELM} \sim \text{laplace}(6e-03, 0,055)$, $E_{LSTM} \sim \text{laplace}(-2e-03, 0,035)$

Nota: p, P – termos autorregressivos não sazonais e sazonais, respectivamente, nL – número de camadas ocultas, nN – número de nós nas camadas ocultas, FA – função de ativação, nSV – número de *Support Vectors*, ω_{modelo} – peso atribuído ao preditor, ρ_i – parâmetro da Cópula, E_{modelo} – resíduo do preditor.

Fonte: Elaborado pelo autor (2022).

Tabela C.20 – Descrição dos modelos construídos para a série P20

Formalismo	Descrição
ARIMA	ARIMA(3,0,0)(2,1,0)[12]
ETS	ETS(A,N,A) $\alpha = 1$ $\gamma = 0$
ANN	p=3, P=1, nN=26, nL=2, FA=logistic
SVR	p=3, P=2, $\gamma=0,534$, C=601,184, $\varepsilon=0,072$, nSV=33
ELM	p=3, P=2, nN=22, FA=satlins
LSTM	p=3, P=1, nN=26
cANN	nN=11, nL=2, FA=logistic
cSVR	$\gamma=0,409$, C=457,428, $\varepsilon=0,089$, nSV=10
cMV	$\omega_{ARIMA} = -0,025$, $\omega_{ETS} = 0,197$, $\omega_{ANN} = 0,797$, $\omega_{SVR} = -0,028$, $\omega_{ELM} = -0,151$, $\omega_{LSTM} = 0,21$
cCP	Normal($\rho_1=0,734$, $\rho_2=0,282$, $\rho_3=0,335$, $\rho_4=0,462$, $\rho_5=0,598$, $\rho_6=0,2$, $\rho_7=0,228$, $\rho_8=0,317$, $\rho_9=0,42$, $\rho_{10}=0,595$, $\rho_{11}=0,554$, $\rho_{12}=0,545$, $\rho_{13}=0,574$, $\rho_{14}=0,614$, $\rho_{15}=0,621$) $E_{ARIMA} \sim \text{laplace}(0,021, 0,089)$, $E_{ETS} \sim \text{laplace}(0, 0,092)$, $E_{ANN} \sim \text{cauchy}(0, 0,025)$, $E_{SVR} \sim \text{laplace}(-9e-03, 7e-02)$, $E_{ELM} \sim \text{laplace}(0,027, 0,079)$, $E_{LSTM} \sim \text{laplace}(-5e-03, 0,065)$

Nota: p, P – termos autorregressivos não sazonais e sazonais, respectivamente, nL – número de camadas ocultas, nN – número de nós nas camadas ocultas, FA – função de ativação, nSV – número de *Support Vectors*, ω_{modelo} – peso atribuído ao preditor, ρ_i – parâmetro da Cópula, E_{modelo} – resíduo do preditor.

Fonte: Elaborado pelo autor (2022).

Tabela C.21 – Descrição dos modelos construídos para a série P21

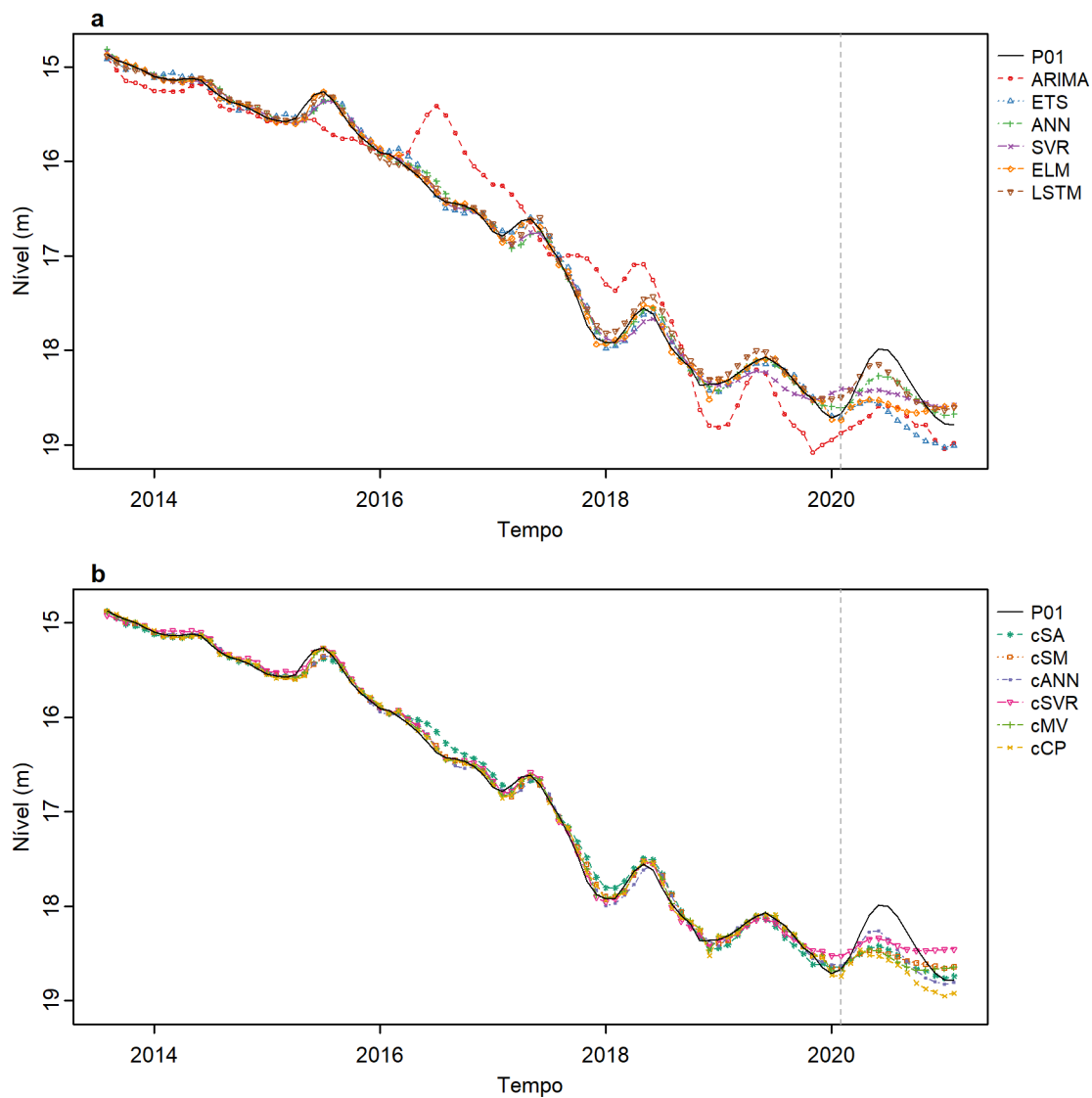
Formalismo	Descrição
ARIMA	ARIMA(0,1,1)(0,1,1)[12]
ETS	ETS(A,N,A) $\alpha = 0,975$ $\gamma = 0$
ANN	p=2, P=1, nN=31, nL=2, FA=logistic
SVR	p=2, P=1, $\gamma=0,557$, C=466,461, $\varepsilon=0,031$, nSV=50
ELM	p=2, P=2, nN=15, FA=tansig
LSTM	p=2, P=2, nN=15
cANN	nN=25, nL=2, FA=logistic
cSVR	$\gamma=0,403$, C=353,331, $\varepsilon=2e-02$, nSV=43
cMV	$\omega_{ARIMA} = 0,125$, $\omega_{ETS} = 0,171$, $\omega_{ANN} = -0,116$, $\omega_{SVR} = 0,836$, $\omega_{ELM} = 0,058$, $\omega_{LSTM} = -0,074$
cCP	Normal($\rho_1=0,761$, $\rho_2=0,471$, $\rho_3=0,463$, $\rho_4=0,365$, $\rho_5=0,375$, $\rho_6=0,333$, $\rho_7=0,409$, $\rho_8=0,293$, $\rho_9=0,496$, $\rho_{10}=0,878$, $\rho_{11}=0,812$, $\rho_{12}=0,536$, $\rho_{13}=0,671$, $\rho_{14}=0,514$, $\rho_{15}=0,444$) $E_{ARIMA} \sim \text{laplace}(-3e-03, 0,062)$, $E_{ETS} \sim \text{laplace}(0,025, 0,064)$, $E_{ANN} \sim \text{norm}(0, 0,073)$, $E_{SVR} \sim \text{laplace}(-3e-03, 0,044)$, $E_{ELM} \sim \text{norm}(3e-02, 0,092)$, $E_{LSTM} \sim \text{norm}(0,042, 0,126)$

Nota: p, P – termos autorregressivos não sazonais e sazonais, respectivamente, nL – número de camadas ocultas, nN – número de nós nas camadas ocultas, FA – função de ativação, nSV – número de *Support Vectors*, ω_{modelo} – peso atribuído ao preditor, ρ_i – parâmetro da Cópula, E_{modelo} – resíduo do preditor.

Fonte: Elaborado pelo autor (2022).

APÊNDICE D – GRÁFICOS DAS PREVISÕES

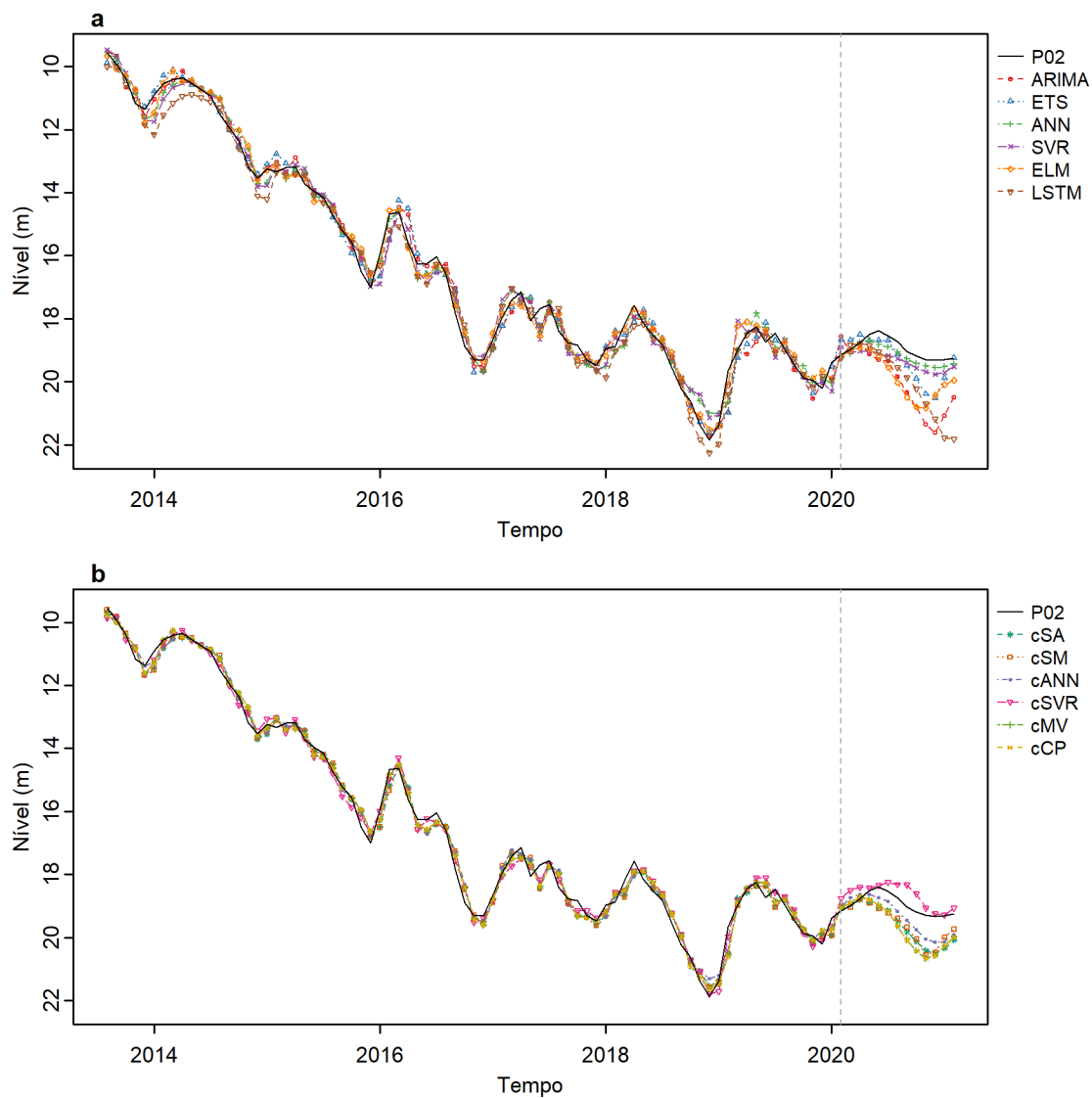
Figura D.1 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P01



Nota: A linha vertical em cor cinza indica a separação em conjuntos de treinamento e teste.

Fonte: Elaborado pelo autor (2022).

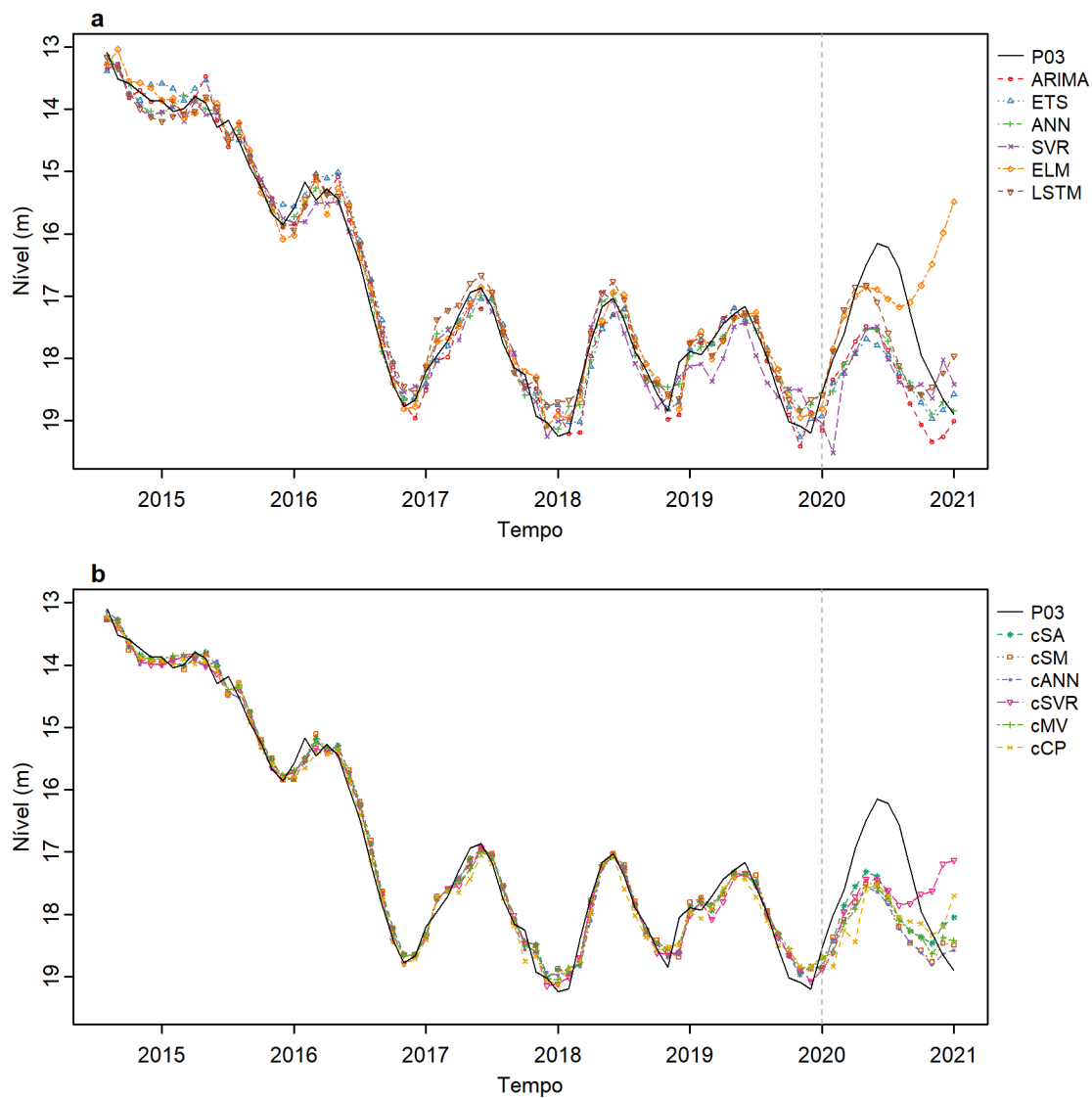
Figura D.2 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P02



Nota: A linha vertical em cor cinza indica a separação em conjuntos de treinamento e teste.

Fonte: Elaborado pelo autor (2022).

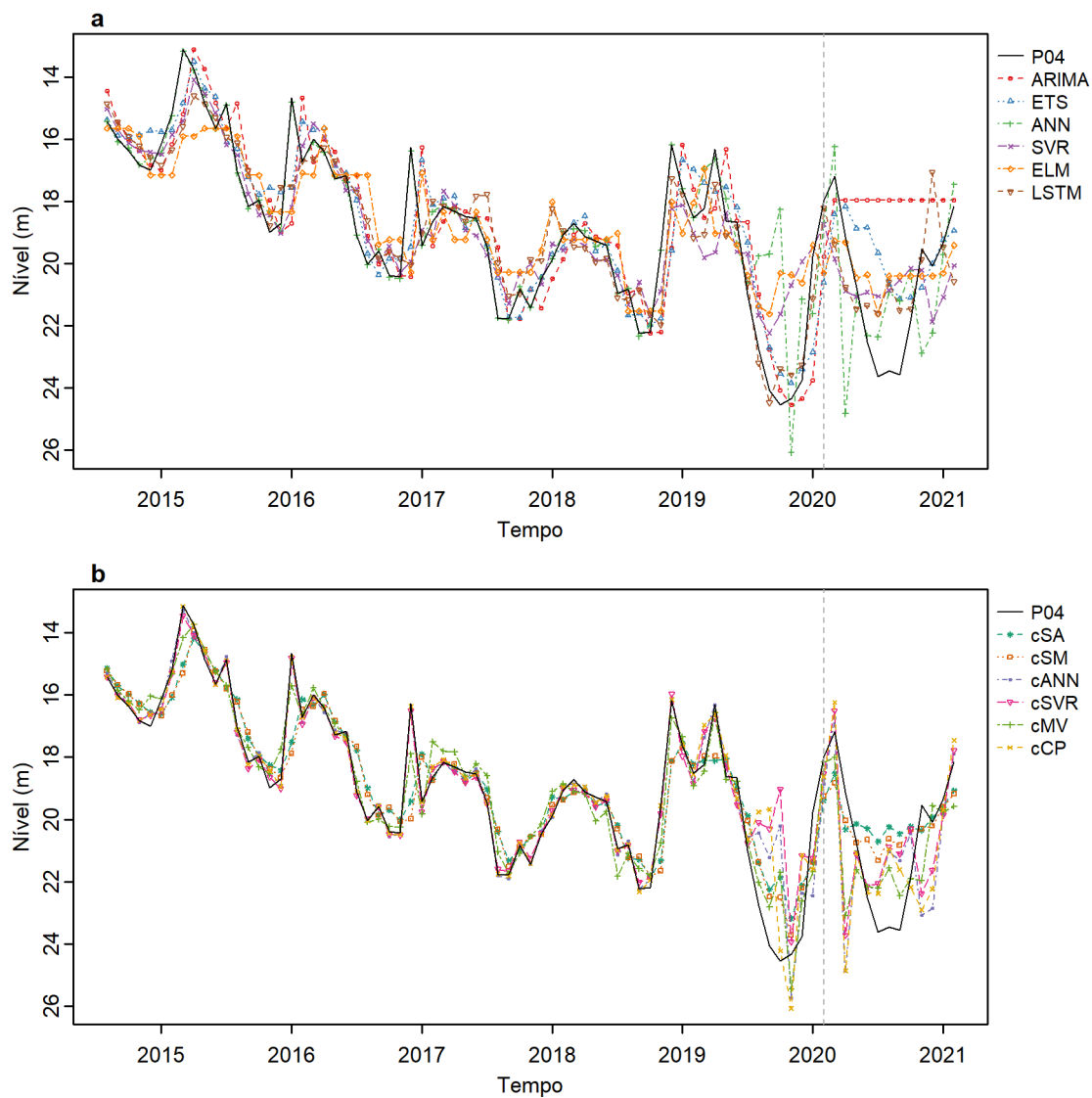
Figura D.3 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P03



Nota: A linha vertical em cor cinza indica a separação em conjuntos de treinamento e teste.

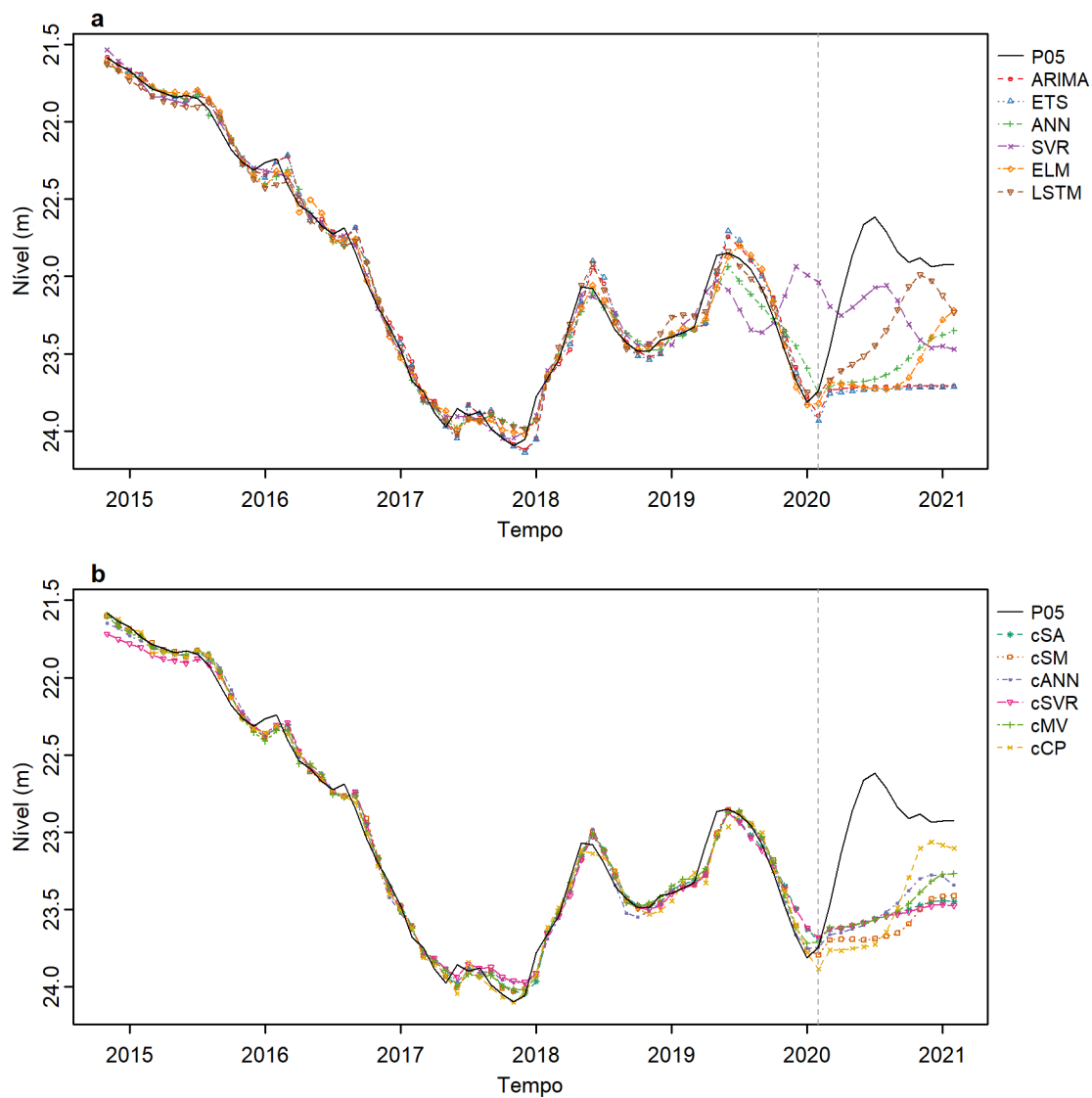
Fonte: Elaborado pelo autor (2022).

Figura D.4 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P04



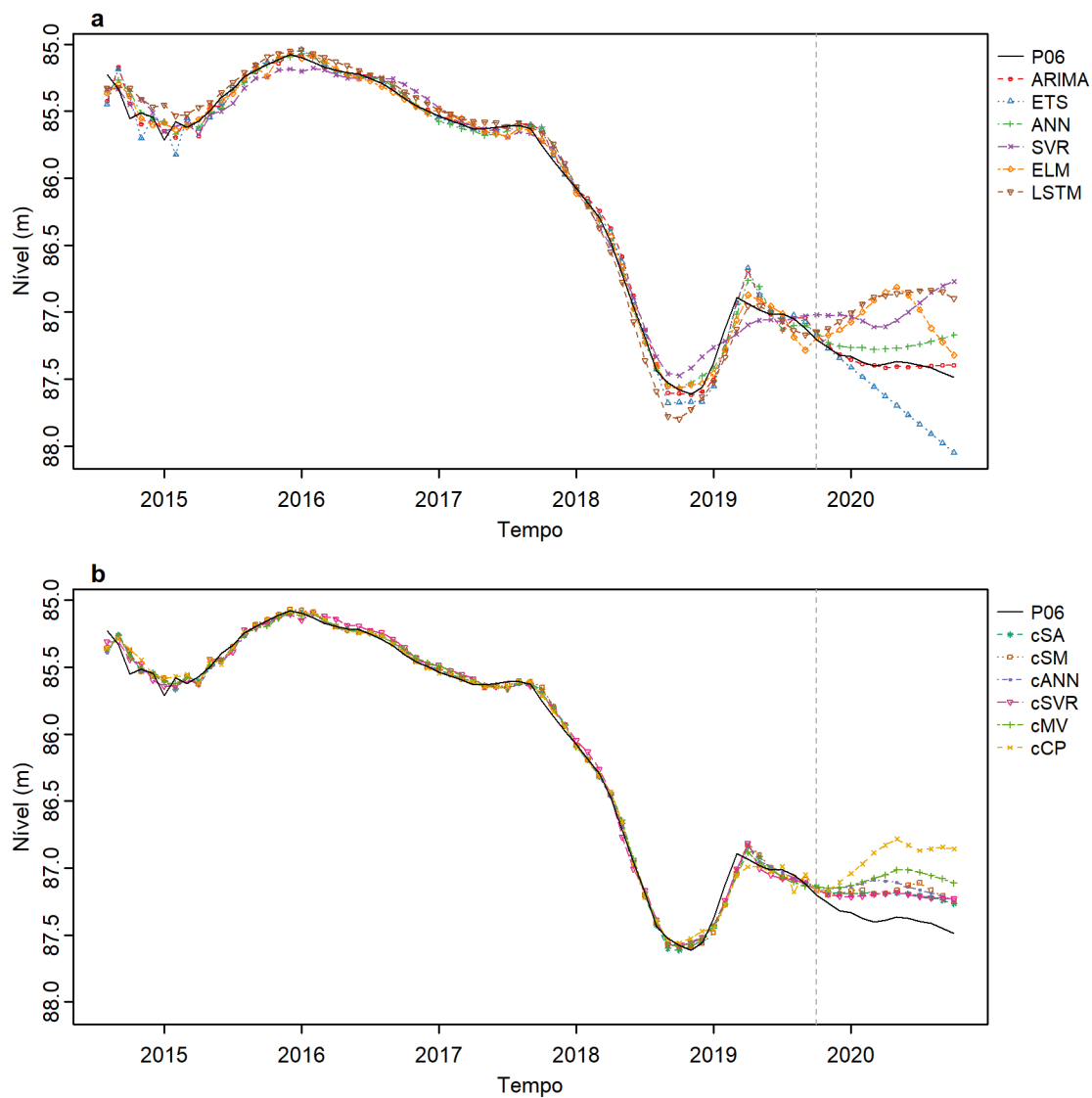
Nota: A linha vertical em cor cinza indica a separação em conjuntos de treinamento e teste.
Fonte: Elaborado pelo autor (2022).

Figura D.5 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P05



Nota: A linha vertical em cor cinza indica a separação em conjuntos de treinamento e teste.
 Fonte: Elaborado pelo autor (2022).

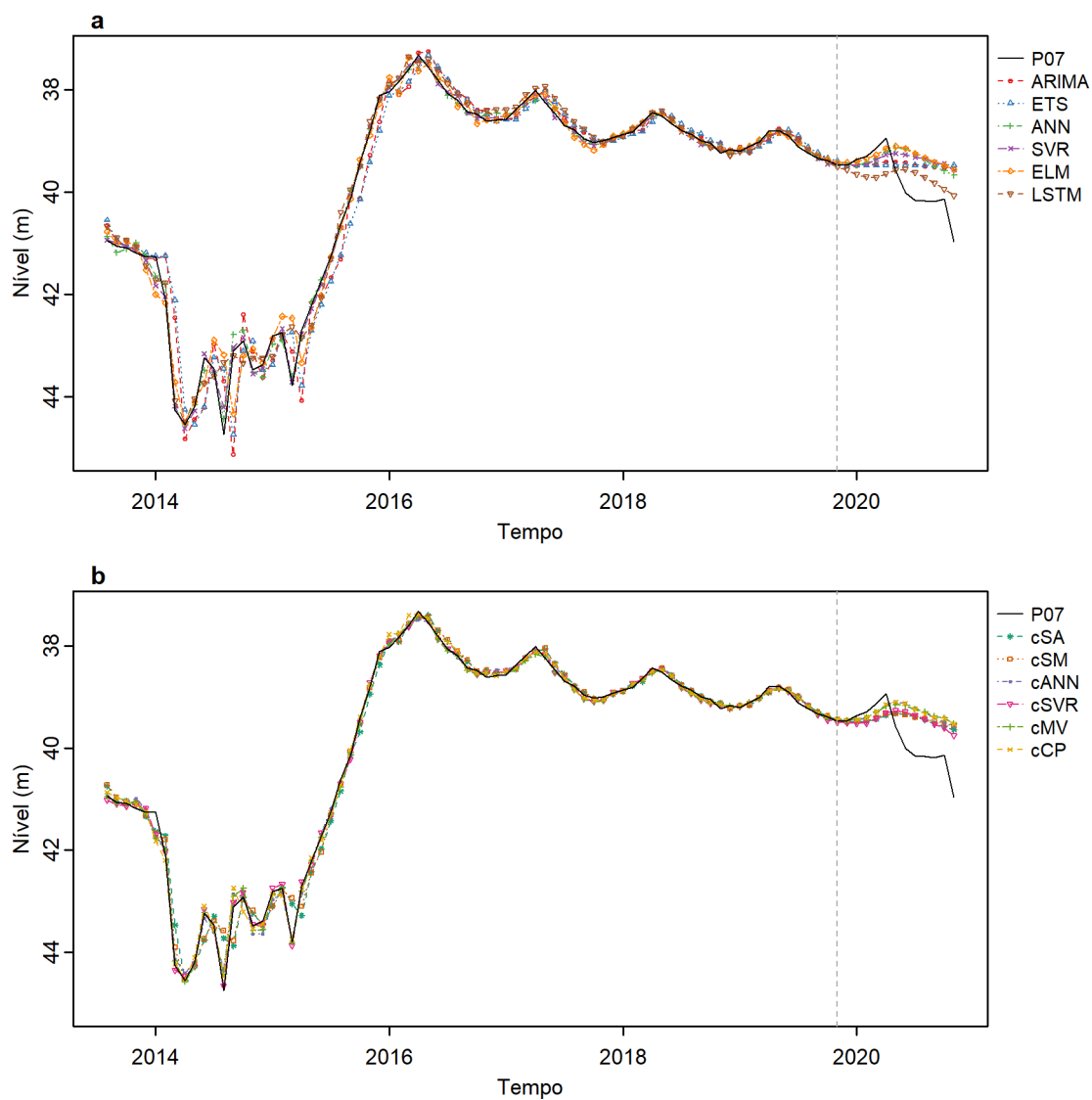
Figura D.6 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P06



Nota: A linha vertical em cor cinza indica a separação em conjuntos de treinamento e teste.

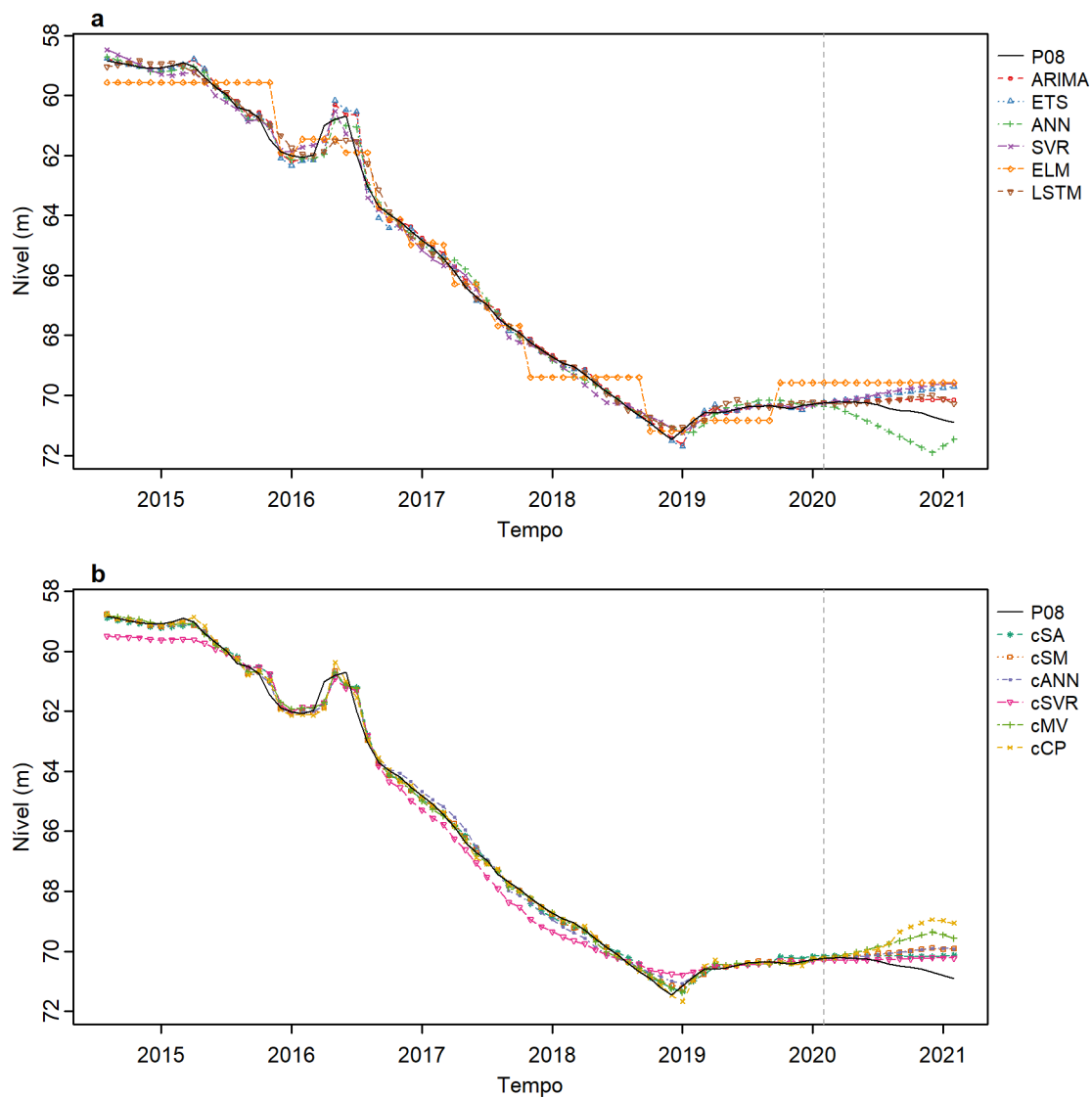
Fonte: Elaborado pelo autor (2022).

Figura D.7 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P07



Nota: A linha vertical em cor cinza indica a separação em conjuntos de treinamento e teste.
Fonte: Elaborado pelo autor (2022).

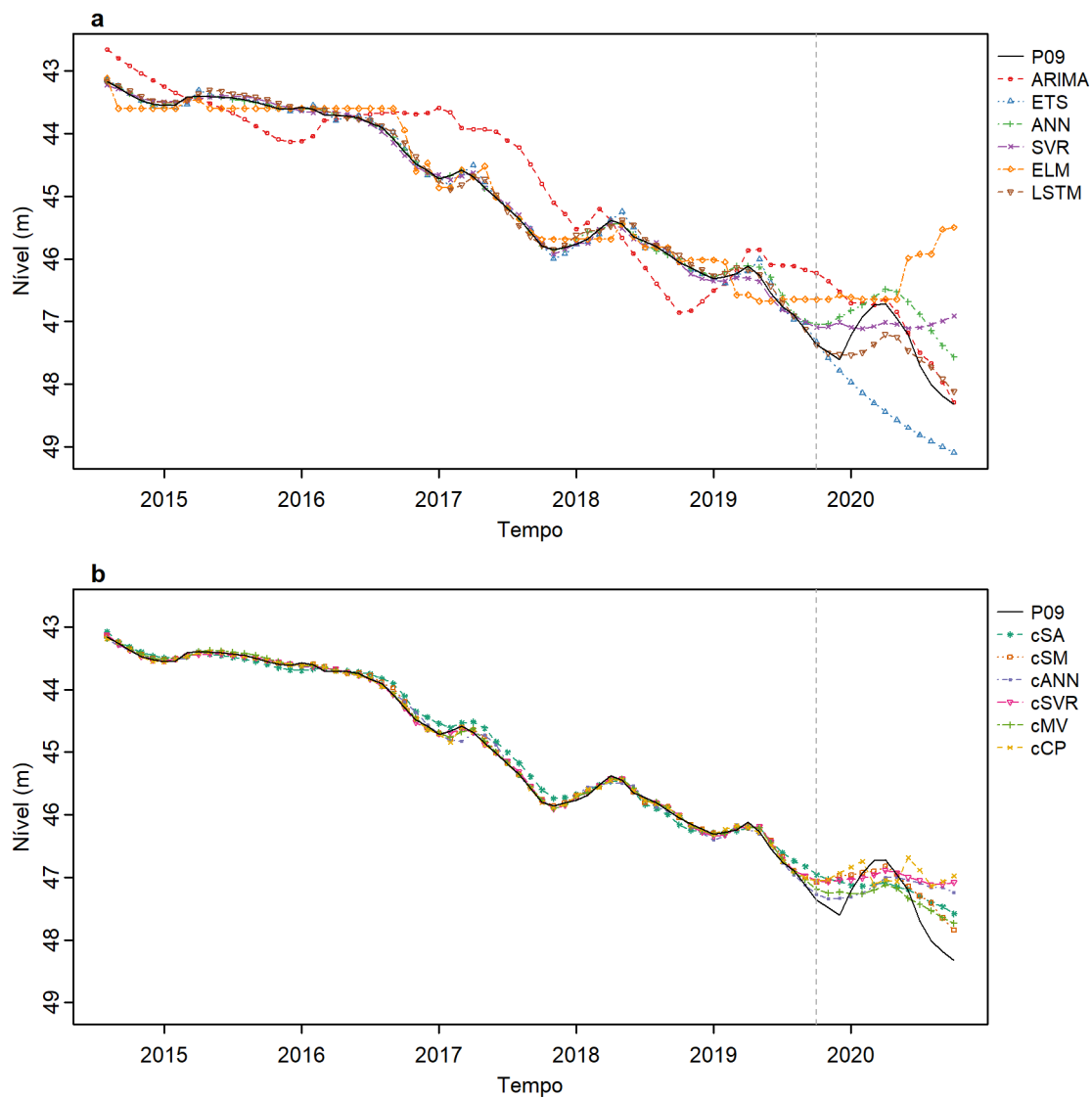
Figura D.8 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P08



Nota: A linha vertical em cor cinza indica a separação em conjuntos de treinamento e teste.

Fonte: Elaborado pelo autor (2022).

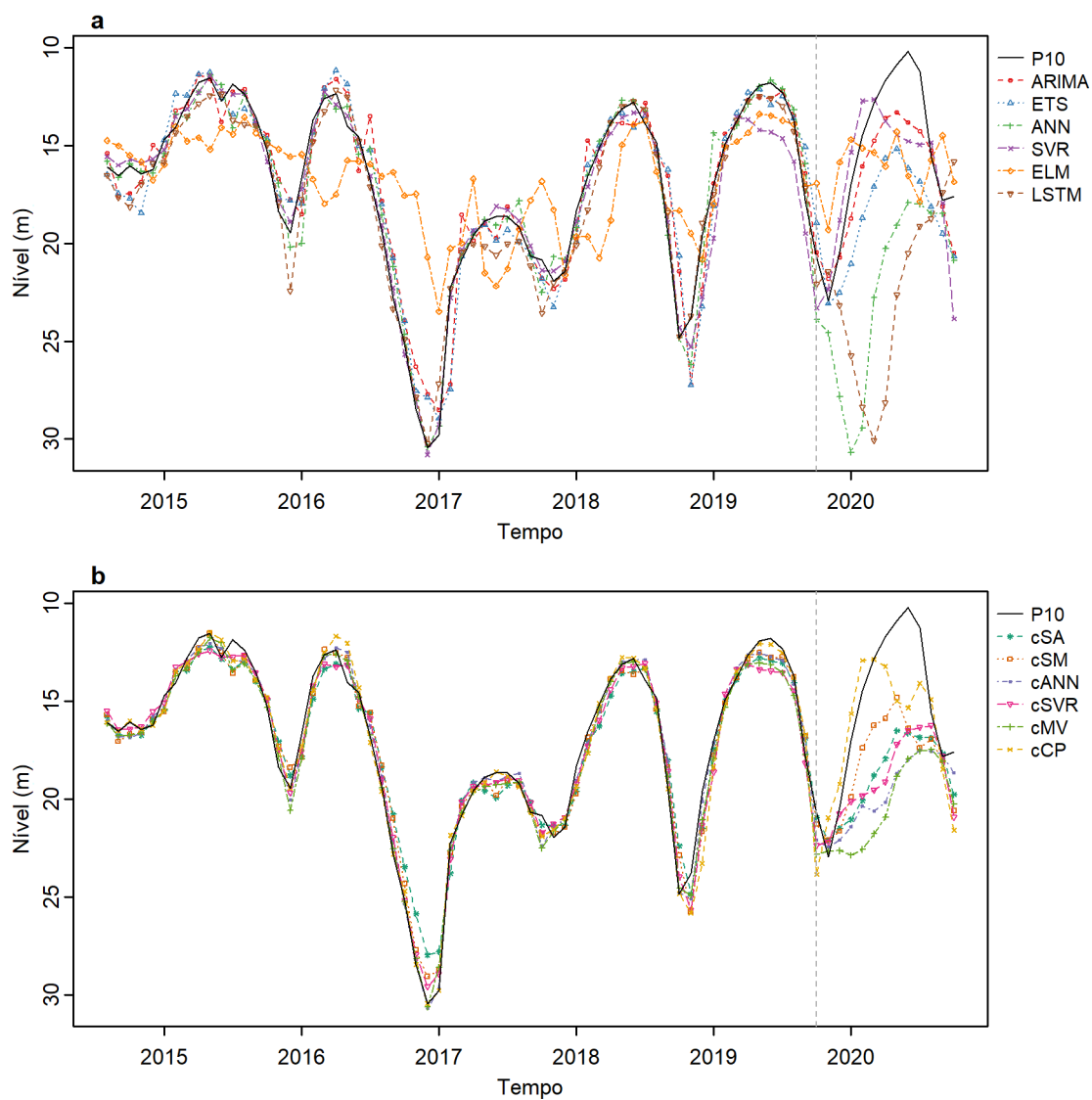
Figura D.9 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P09



Nota: A linha vertical em cor cinza indica a separação em conjuntos de treinamento e teste.

Fonte: Elaborado pelo autor (2022).

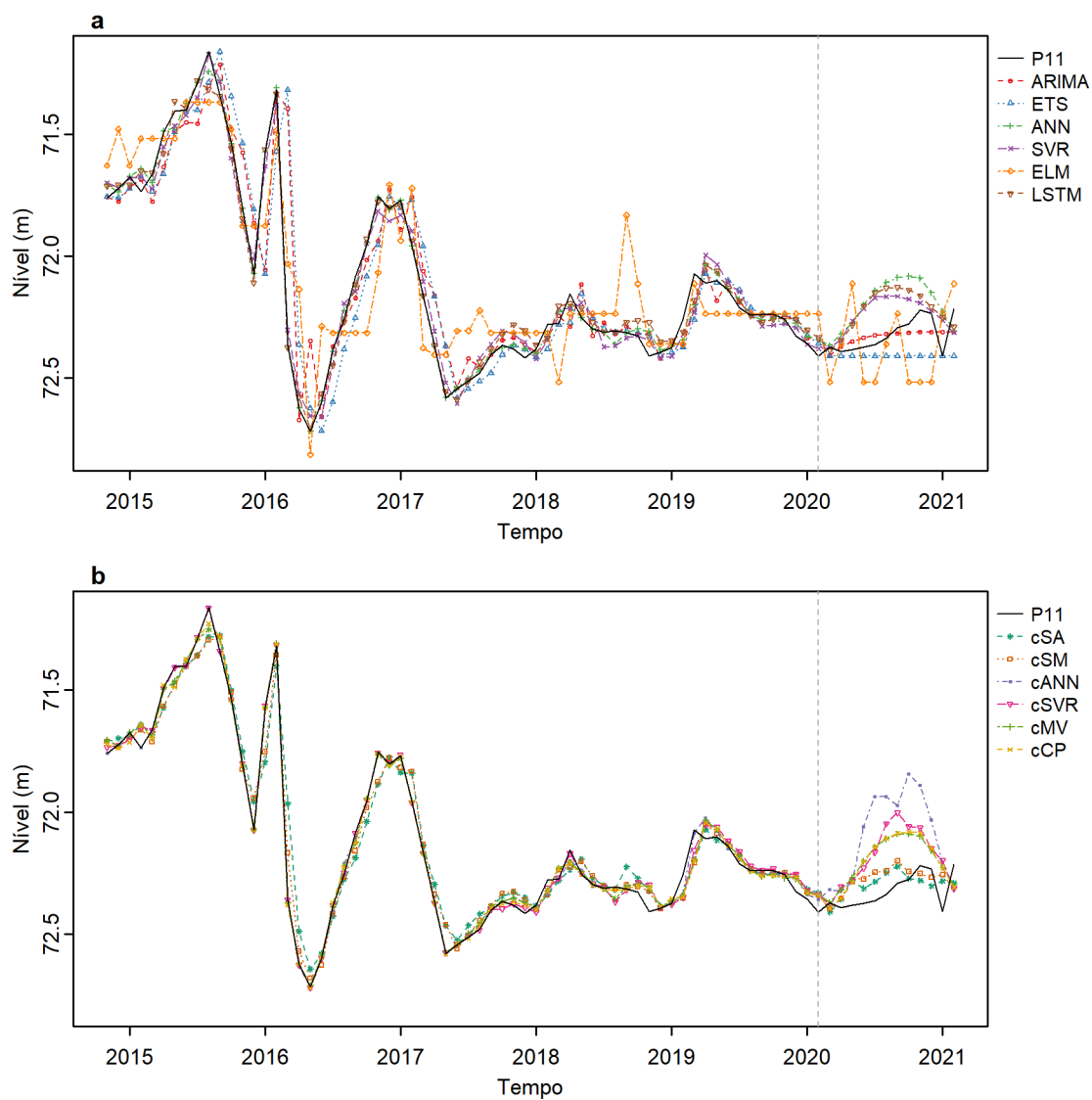
Figura D.10 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P10



Nota: A linha vertical em cor cinza indica a separação em conjuntos de treinamento e teste.

Fonte: Elaborado pelo autor (2022).

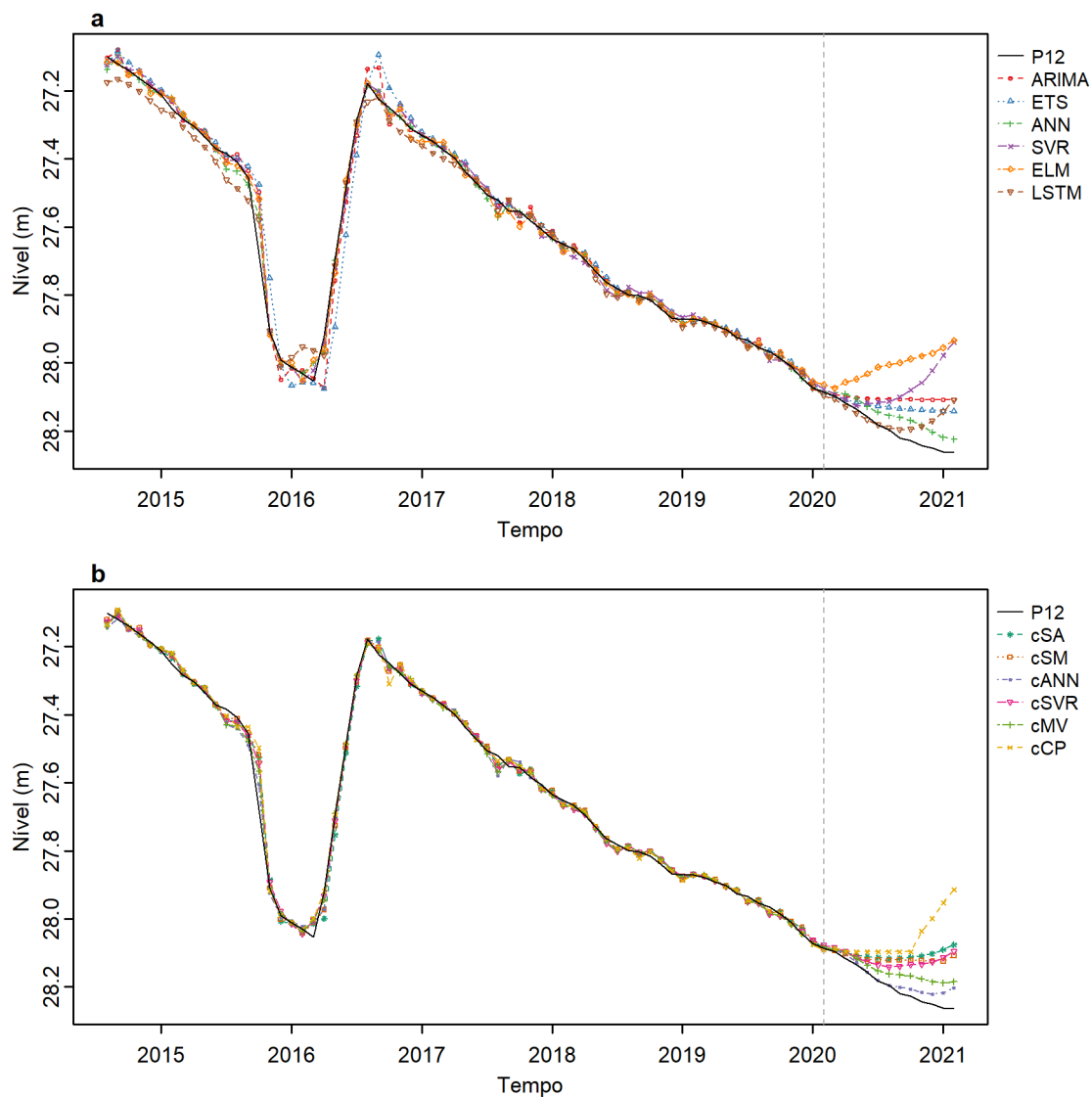
Figura D.11 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P11



Nota: A linha vertical em cor cinza indica a separação em conjuntos de treinamento e teste.

Fonte: Elaborado pelo autor (2022).

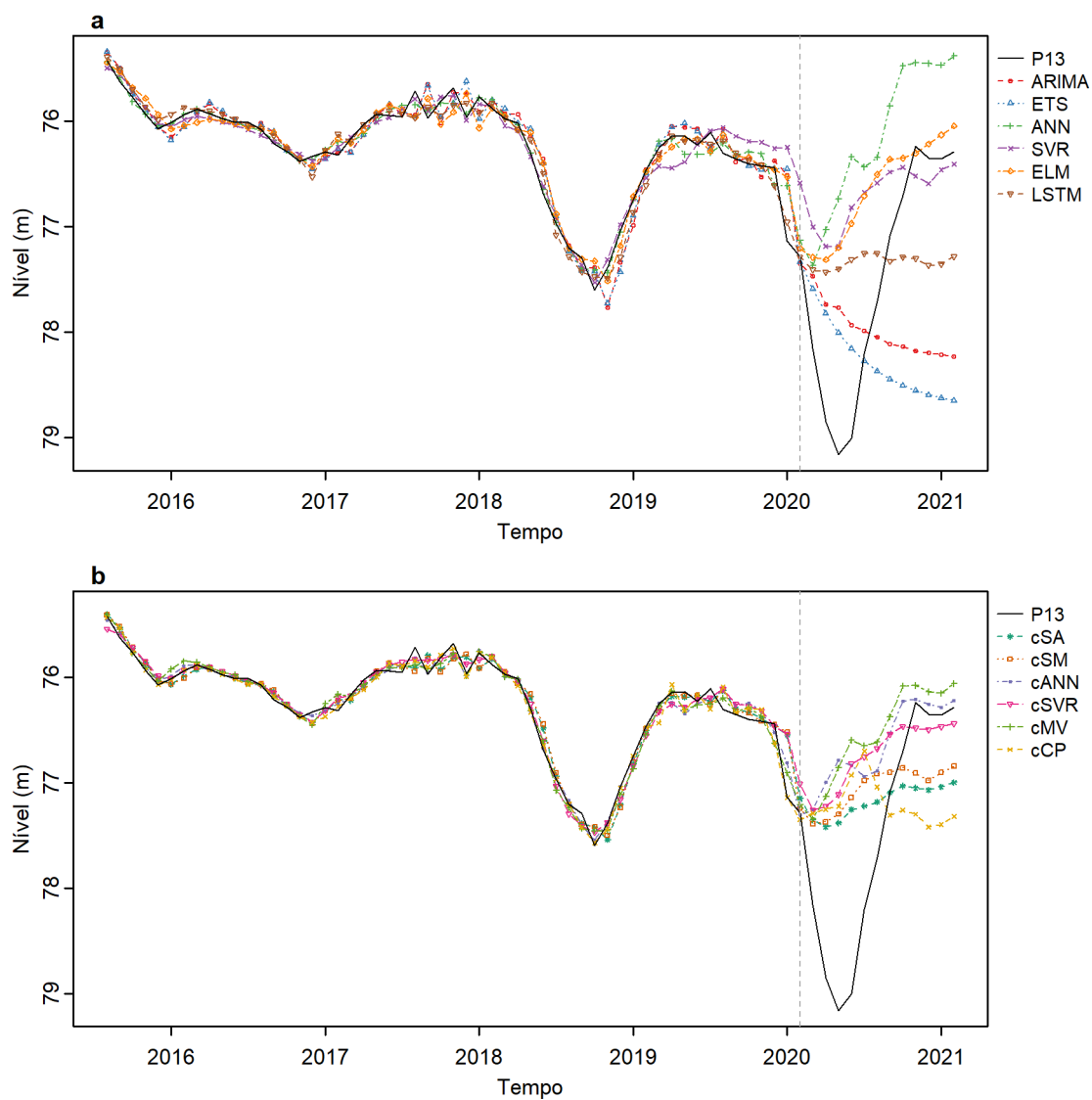
Figura D.12 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P12



Nota: A linha vertical em cor cinza indica a separação em conjuntos de treinamento e teste.

Fonte: Elaborado pelo autor (2022).

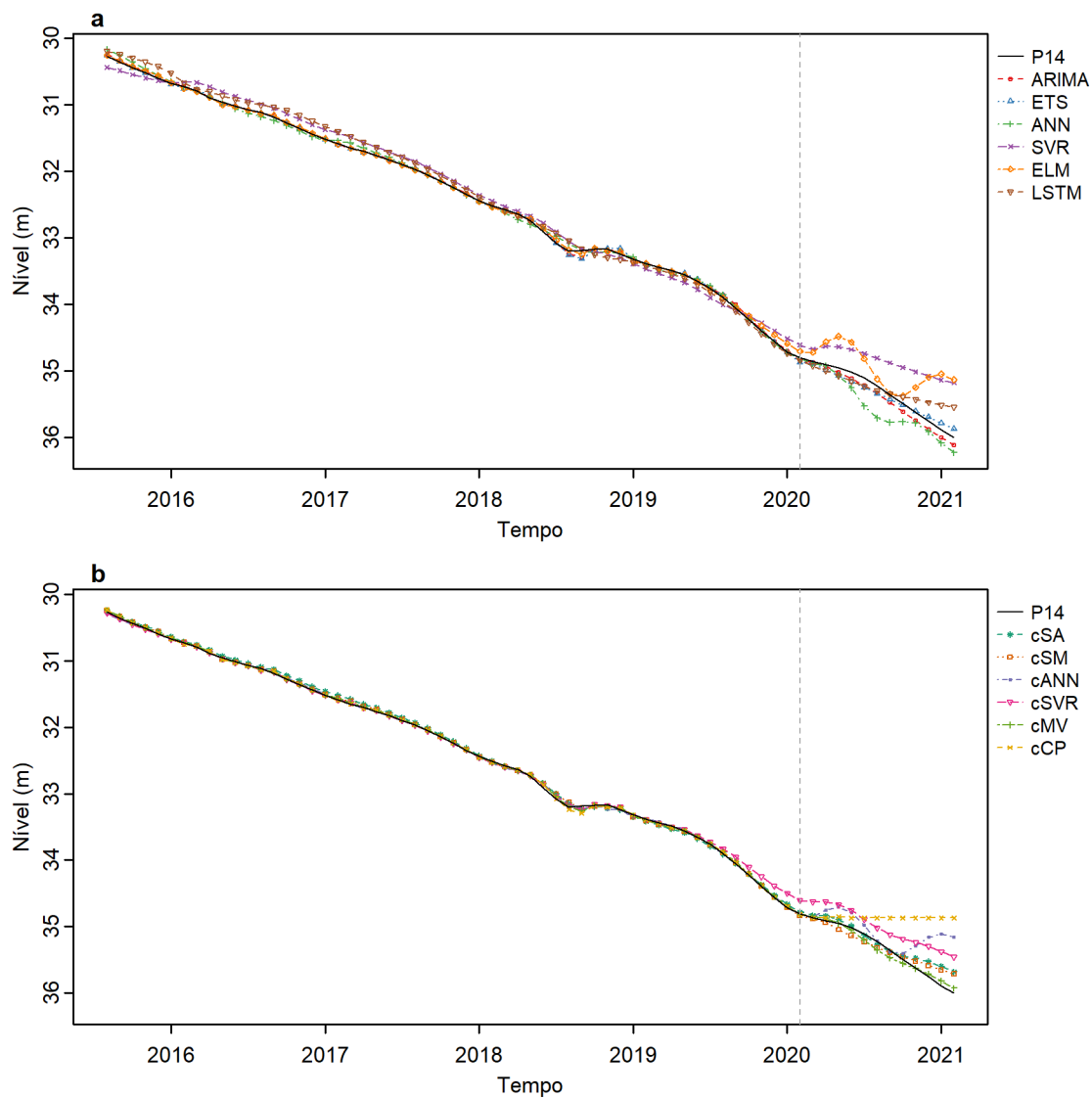
Figura D.13 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P13



Nota: A linha vertical em cor cinza indica a separação em conjuntos de treinamento e teste.

Fonte: Elaborado pelo autor (2022).

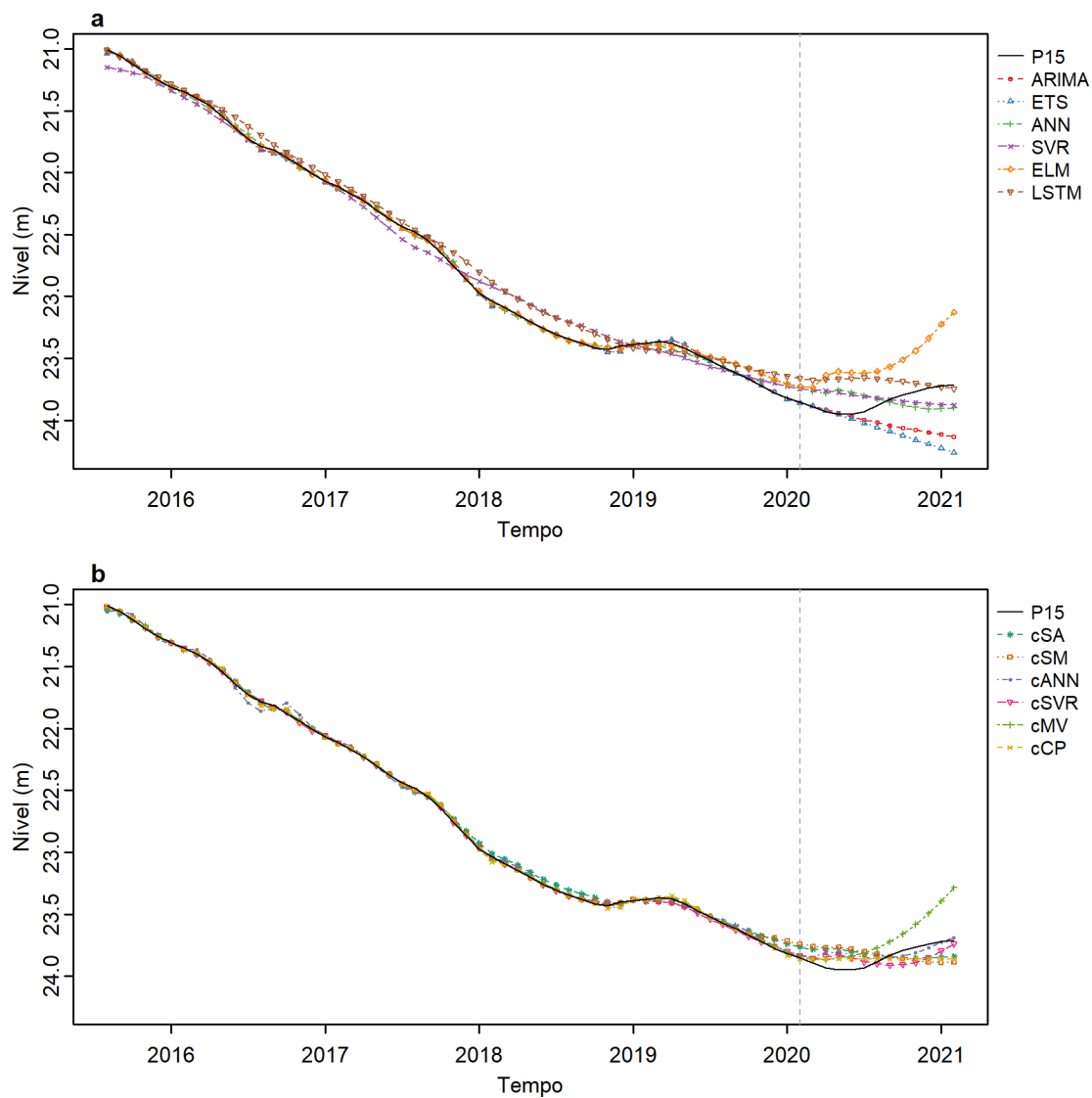
Figura D.14 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P14



Nota: A linha vertical em cor cinza indica a separação em conjuntos de treinamento e teste.

Fonte: Elaborado pelo autor (2022).

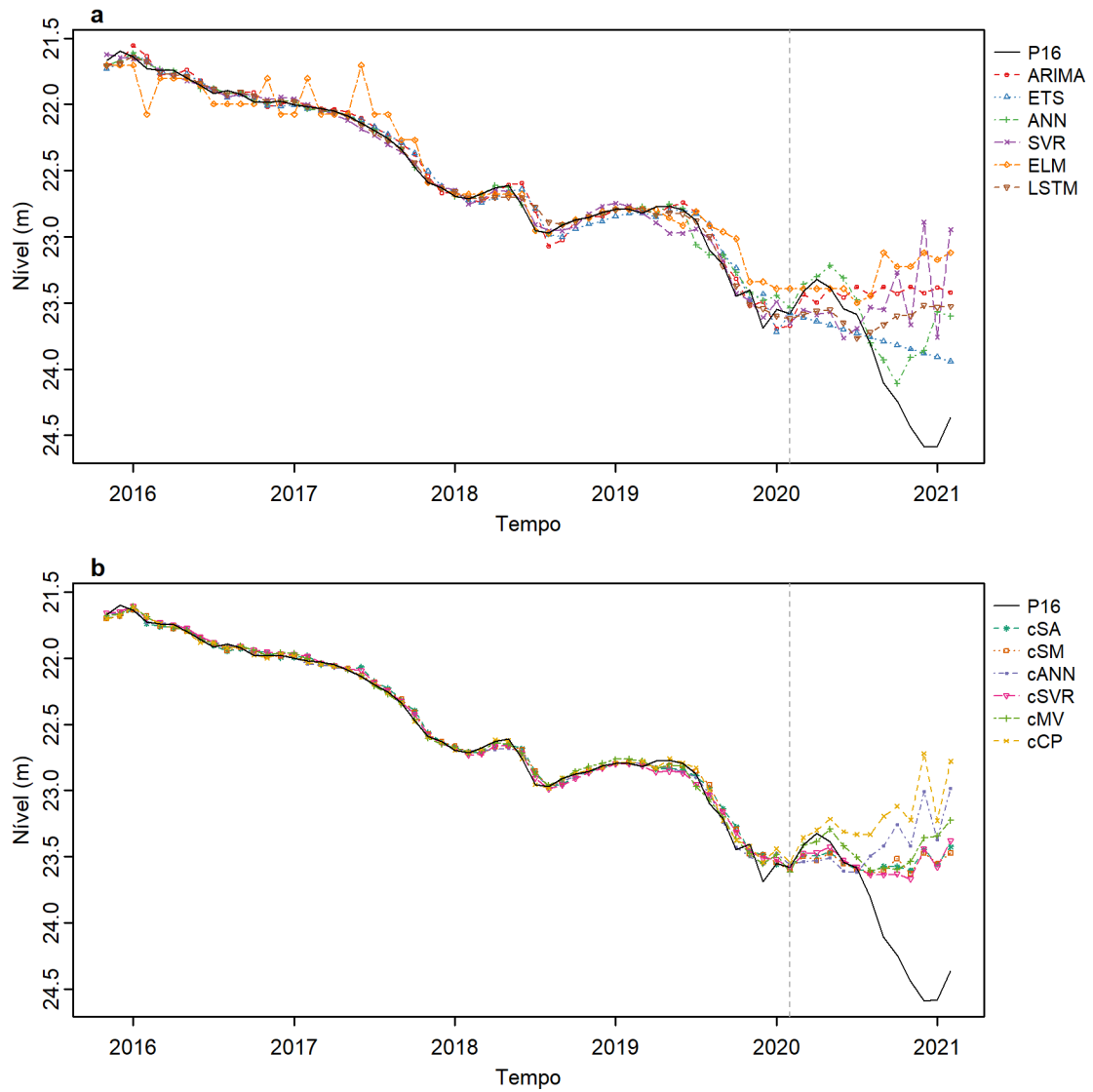
Figura D.15 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P15



Nota: A linha vertical em cor cinza indica a separação em conjuntos de treinamento e teste.

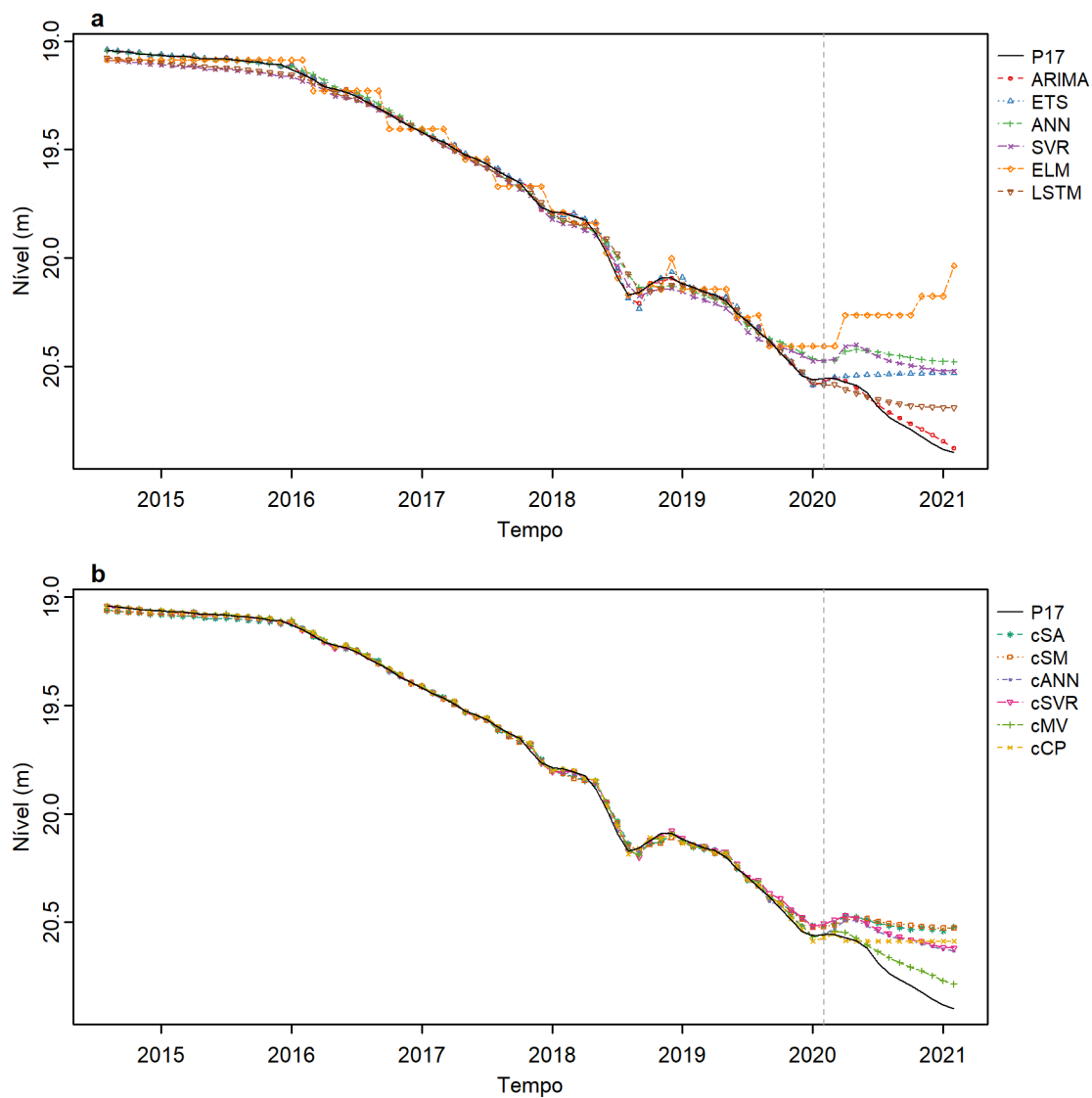
Fonte: Elaborado pelo autor (2022).

Figura D.16 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P16



Nota: A linha vertical em cor cinza indica a separação em conjuntos de treinamento e teste.
 Fonte: Elaborado pelo autor (2022).

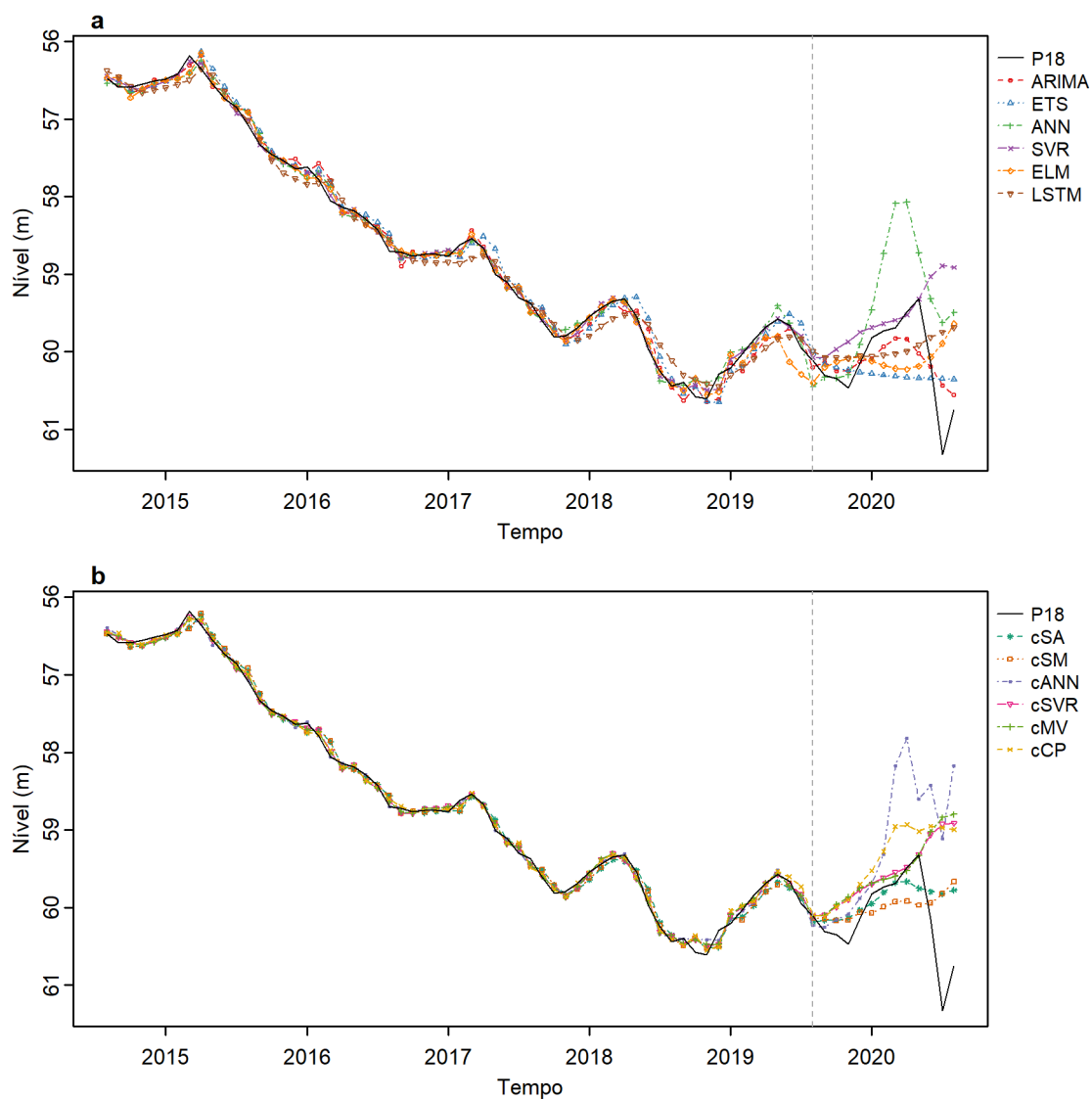
Figura D.17 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P17



Nota: A linha vertical em cor cinza indica a separação em conjuntos de treinamento e teste.

Fonte: Elaborado pelo autor (2022).

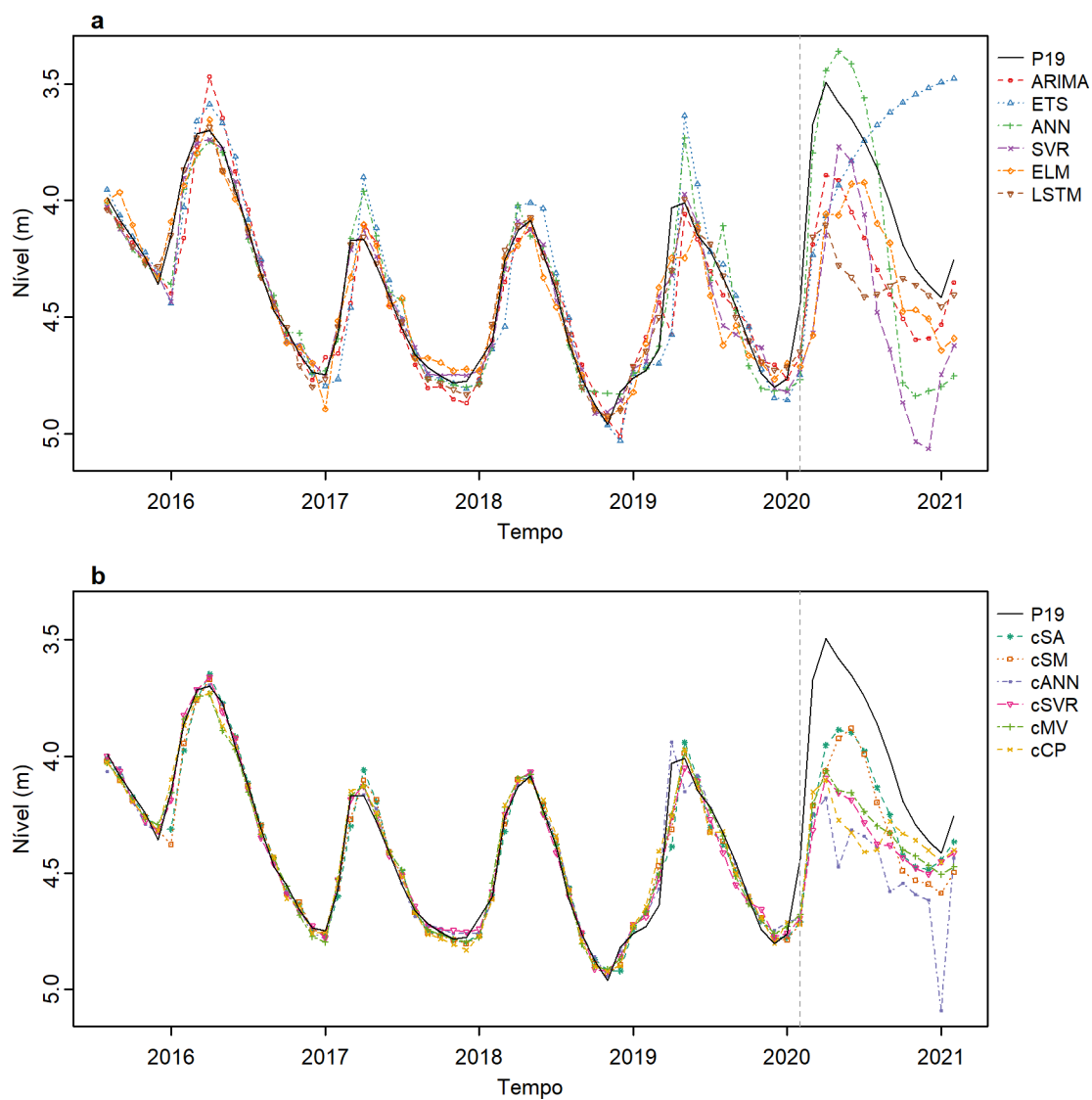
Figura D.18 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P18



Nota: A linha vertical em cor cinza indica a separação em conjuntos de treinamento e teste.

Fonte: Elaborado pelo autor (2022).

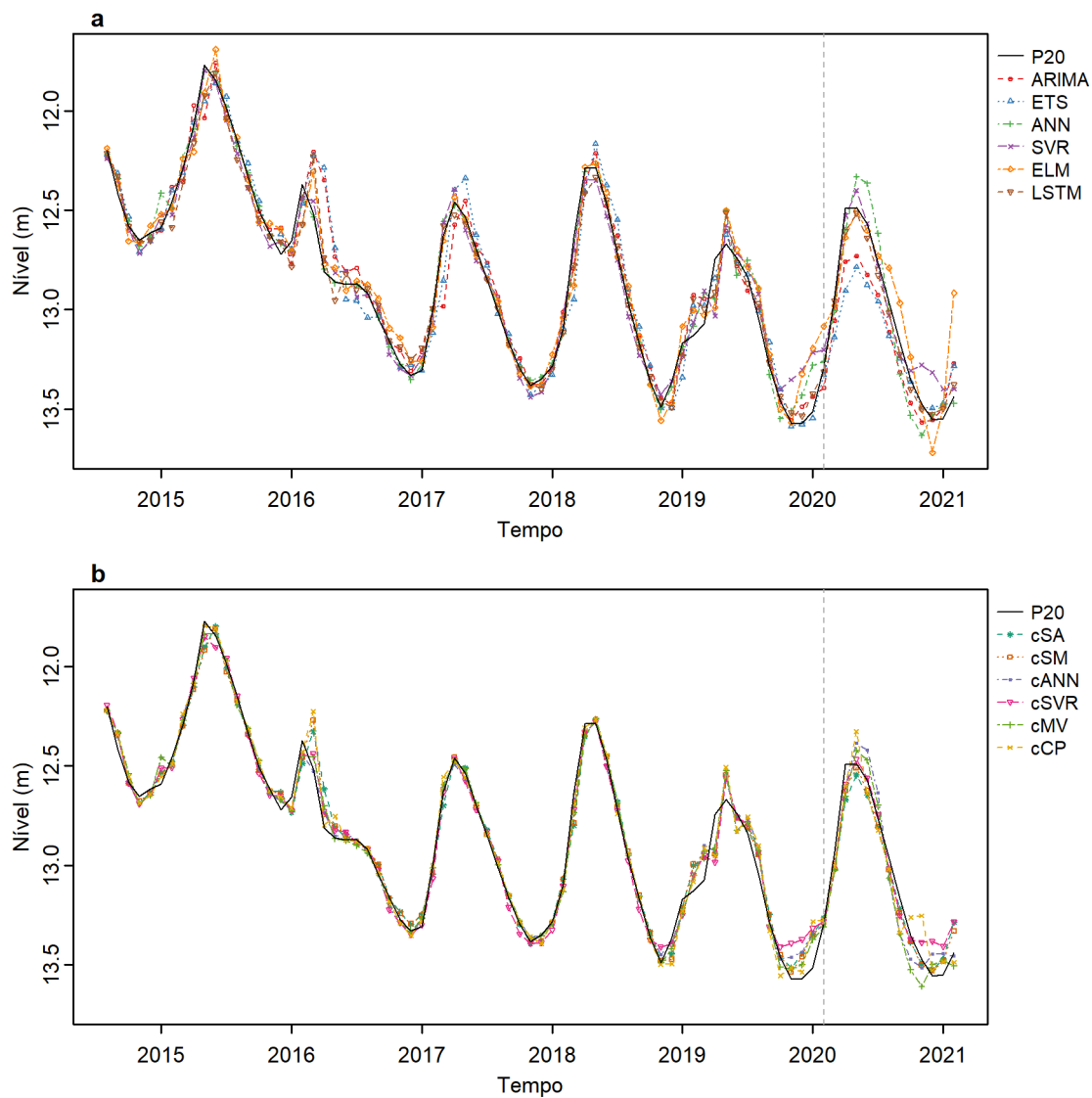
Figura D.19 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P19



Nota: A linha vertical em cor cinza indica a separação em conjuntos de treinamento e teste.

Fonte: Elaborado pelo autor (2022).

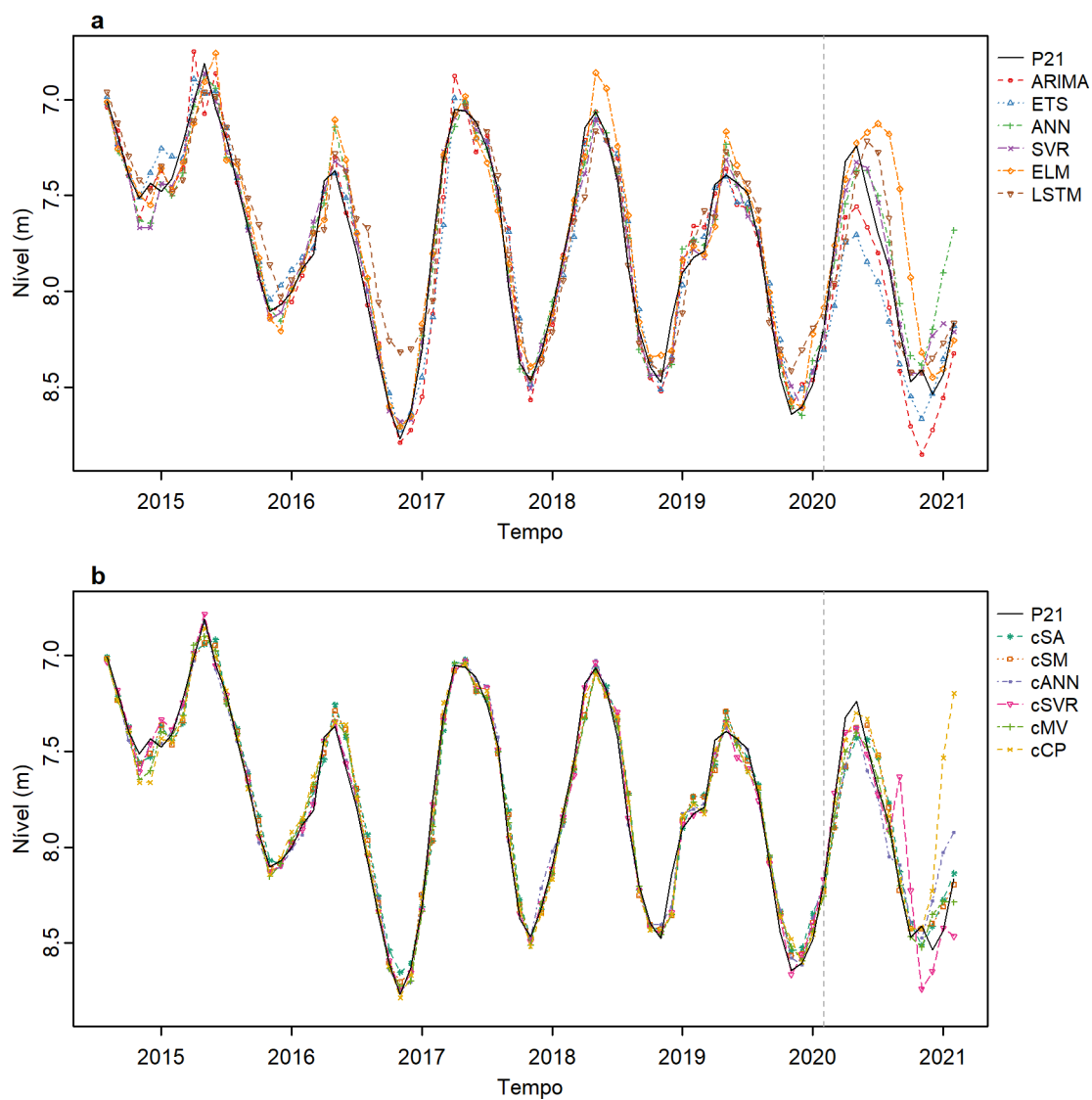
Figura D.20 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P20



Nota: A linha vertical em cor cinza indica a separação em conjuntos de treinamento e teste.

Fonte: Elaborado pelo autor (2022).

Figura D.21 – Gráfico das previsões realizadas pelos modelos a. individuais e b. combinados para a série P21



Nota: A linha vertical em cor cinza indica a separação em conjuntos de treinamento e teste.

Fonte: Elaborado pelo autor (2022).

APÊNDICE E – AVALIAÇÃO DO DESEMPENHO DOS MODELOS

Tabela E.1 – Desempenho dos modelos de previsão para a série P01 no conjunto de teste (mar/2020 – fev/2021)

Modelo	RMSE	MAE	MAPE	U de Theil	R ²	POCID
ARIMA	0,4224	0,3947	2,1599	9,600	0,9168	72,73
ETS	0,4356	0,3999	2,1872	10,595	0,6575	81,82
ANN	0,1665	0,1385	0,7601	1,549	0,9985	90,91
SVR	0,2610	0,2262	1,2394	3,746	0,6522	81,82
ELM	0,3416	0,2914	1,5990	6,513	0,2032	54,55
LSTM	0,1460	0,1275	0,6956	1,143	0,8474	90,91
cSA	0,2738	0,2199	1,2097	4,192	0,9015	90,91
cSM	0,2922	0,2406	1,3219	4,770	0,8638	90,91
cANN	0,2144	0,1784	0,9785	2,571	0,8735	90,91
cSVR	0,2506	0,2218	1,2092	3,504	0,6585	63,64
cMV	0,3128	0,2598	1,4269	5,474	0,5551	54,55
cCP	0,3645	0,3246	1,7778	7,414	0,7216	72,73

Nota: RMSE (em m), MAE (em m), MAPE (em %), Coeficiente U de Theil, Coeficiente de Determinação R² e POCID (em %).

Fonte: Elaborado pelo autor (2022).

Tabela E.2 – Desempenho dos modelos de previsão para a série P02 no conjunto de teste (mar/2020 – fev/2021)

Modelo	RMSE	MAE	MAPE	U de Theil	R ²	POCID
ARIMA	1,3377	1,1474	6,018	67,016	0,6989	81,82
ETS	0,5852	0,4683	2,453	12,539	0,6695	72,73
ANN	0,2498	0,2262	1,198	2,297	0,8832	90,91
SVR	0,4480	0,4164	2,208	7,500	0,7504	90,91
ELM	1,0433	0,9139	4,811	40,775	0,5141	72,73
LSTM	1,3447	1,0737	5,618	67,672	0,6706	63,64
cSA	0,7828	0,6912	3,631	22,957	0,7528	81,82
cSM	0,7248	0,6314	3,321	19,671	0,6800	72,73
cANN	0,5286	0,4717	2,473	10,263	0,7899	81,82
cSVR	0,3601	0,2897	1,528	4,135	0,6788	72,73
cMV	0,8726	0,7614	3,997	28,501	0,7003	72,73
cCP	0,8731	0,7614	3,997	28,537	0,6993	72,73

Nota: RMSE (em m), MAE (em m), MAPE (em %), Coeficiente U de Theil, Coeficiente de Determinação R² e POCID (em %).

Fonte: Elaborado pelo autor (2022).

Tabela E.3 – Desempenho dos modelos de previsão para a série P03 no conjunto de teste (fev/2020 – jan/2021)

Modelo	RMSE	MAE	MAPE	U de Theil	R ²	POCID
ARIMA	1,0981	0,9741	5,719	6,155	0,7154	72,73
ETS	1,0833	0,9479	5,605	5,973	0,7333	72,73
ANN	0,9718	0,8307	4,923	4,741	0,8495	81,82
SVR	1,1341	1,0171	5,965	5,659	0,2933	81,82
ELM	1,4503	1,0211	5,670	10,800	0,2773	45,45
LSTM	0,8391	0,6809	3,982	3,611	0,3227	63,64
cSA	0,8886	0,7708	4,532	3,992	0,5048	72,73
cSM	0,9773	0,8494	5,020	4,859	0,6795	81,82
cANN	1,0135	0,8669	5,135	5,208	0,7287	72,73
cSVR	1,0580	0,9522	5,492	5,601	0,0114	45,45
cMV	0,9622	0,8494	5,017	4,613	0,7558	81,82
cCP	1,0513	0,9293	5,450	5,398	0,1951	72,73

Nota: RMSE (em m), MAE (em m), MAPE (em %), Coeficiente U de Theil, Coeficiente de Determinação R² e POCID (em %).

Fonte: Elaborado pelo autor (2022).

Tabela E.4 – Desempenho dos modelos de previsão para a série P04 no conjunto de teste (mar/2020 – fev/2021)

Modelo	RMSE	MAE	MAPE	U de Theil	R ²	POCID
ARIMA	3,507	2,927	13,307	6,913	0,0000	0,00
ETS	2,071	1,654	7,611	2,352	0,3293	63,64
ANN	2,343	1,784	8,704	3,058	0,1826	72,73
SVR	2,010	1,857	8,954	1,948	0,1018	72,73
ELM	1,774	1,481	6,955	1,573	0,5828	54,55
LSTM	1,848	1,577	7,650	1,702	0,2697	63,64
cSA	1,853	1,533	7,080	1,856	0,6211	72,73
cSM	1,628	1,333	6,213	1,369	0,7972	81,82
cANN	2,393	1,782	8,647	3,223	0,1452	63,64
cSVR	2,045	1,619	7,790	2,337	0,2396	63,64
cMV	1,641	1,269	6,253	1,489	0,4330	63,64
cCP	2,275	1,646	8,087	2,877	0,2315	63,64

Nota: RMSE (em m), MAE (em m), MAPE (em %), Coeficiente U de Theil, Coeficiente de Determinação R² e POCID (em %).

Fonte: Elaborado pelo autor (2022).

Tabela E.5 – Desempenho dos modelos de previsão para a série P05 no conjunto de teste (mar/2020 – fev/2021)

Modelo	RMSE	MAE	MAPE	U de Theil	R²	POCID
ARIMA	0,8350	0,8082	3,537	31,866	0,5544	63,64
ETS	0,8491	0,8231	3,602	32,918	0,2993	63,64
ANN	0,7016	0,6581	2,881	22,463	0,0055	63,64
SVR	0,4224	0,4032	1,763	7,936	0,0764	63,64
ELM	0,7414	0,6808	2,982	25,164	0,0109	36,36
LSTM	0,4918	0,4112	1,802	10,998	0,0742	54,55
cSA	0,6568	0,6231	2,729	19,794	0,0764	54,55
cSM	0,7313	0,6886	3,015	24,451	0,0001	45,45
cANN	0,6120	0,5646	2,473	17,126	0,0347	45,45
cSVR	0,6673	0,6347	2,779	20,435	0,0912	54,55
cMV	0,6215	0,5724	2,508	17,726	0,0067	54,55
cCP	0,6591	0,5536	2,425	19,697	0,0014	36,36

Nota: RMSE (em m), MAE (em m), MAPE (em %), Coeficiente U de Theil, Coeficiente de Determinação R² e POCID (em %).

Fonte: Elaborado pelo autor (2022).

Tabela E.6 – Desempenho dos modelos de previsão para a série P06 no conjunto de teste (nov/2020 – out/2021)

Modelo	RMSE	MAE	MAPE	U de Theil	R²	POCID
ARIMA	0,0368	0,0263	0,0301	1,618	0,5830	54,55
ETS	0,3403	0,2799	0,3202	139,305	0,8227	81,82
ANN	0,1596	0,1389	0,1589	30,514	0,2627	54,55
SVR	0,4279	0,3995	0,4572	214,564	0,4065	45,45
ELM	0,3739	0,3416	0,3909	167,334	0,0424	63,64
LSTM	0,4822	0,4612	0,5277	277,795	0,6144	36,36
cSA	0,1814	0,1774	0,2030	38,869	0,7068	81,82
cSM	0,2113	0,2030	0,2322	53,275	0,0593	54,55
cANN	0,2368	0,2311	0,2644	66,211	0,2569	45,45
cSVR	0,1803	0,1728	0,1978	38,665	0,2341	81,82
cMV	0,3124	0,3003	0,3436	116,147	0,1561	63,64
cCP	0,4895	0,4613	0,5278	287,294	0,5825	45,45

Nota: RMSE (em m), MAE (em m), MAPE (em %), Coeficiente U de Theil, Coeficiente de Determinação R² e POCID (em %).

Fonte: Elaborado pelo autor (2022).

Tabela E.7 – Desempenho dos modelos de previsão para a série P07 no conjunto de teste (dez/2020 – nov/2021)

Modelo	RMSE	MAE	MAPE	U de Theil	R ²	POCID
ARIMA	0,6167	0,4938	1,231	3,307	0,4600	72,73
ETS	0,6369	0,5037	1,255	3,527	0,0000	0,00
ANN	0,6451	0,5222	1,301	3,618	0,1968	72,73
SVR	0,6616	0,5335	1,329	3,805	0,1374	81,82
ELM	0,6823	0,5327	1,325	4,045	0,2556	81,82
LSTM	0,4799	0,4149	1,041	1,995	0,4057	45,45
cSA	0,6073	0,4989	1,244	3,207	0,2930	72,73
cSM	0,6313	0,5094	1,269	3,465	0,2586	72,73
cANN	0,6864	0,5442	1,355	4,096	0,1442	81,82
cSVR	0,5948	0,5025	1,254	3,075	0,2591	72,73
cMV	0,6985	0,5515	1,373	4,241	0,1075	81,82
cCP	0,6962	0,5492	1,367	4,214	0,1311	81,82

Nota: RMSE (em m), MAE (em m), MAPE (em %), Coeficiente U de Theil, Coeficiente de Determinação R² e POCID (em %).

Fonte: Elaborado pelo autor (2022).

Tabela E.8 – Desempenho dos modelos de previsão para a série P08 no conjunto de teste (mar/2020 – fev/2021)

Modelo	RMSE	MAE	MAPE	U de Theil	R ²	POCID
ARIMA	0,4041	0,3201	0,4530	33,21	0,6230	0,00
ETS	0,6467	0,5233	0,7408	85,03	0,9472	0,00
ANN	0,7903	0,7264	1,0299	126,56	0,7203	81,82
SVR	0,7368	0,6023	0,8526	110,38	0,9057	9,09
ELM	0,9269	0,8986	1,2740	167,58	0,0000	0,00
LSTM	0,4151	0,3225	0,4565	34,98	0,2625	18,18
cSA	0,3926	0,3193	0,4521	31,23	0,0036	72,73
cSM	0,5419	0,4346	0,6152	59,70	0,8624	9,09
cANN	0,5377	0,4188	0,5927	58,78	0,8867	9,09
cSVR	0,3339	0,2530	0,3580	22,60	0,9347	18,18
cMV	0,8746	0,7339	1,0392	155,51	0,7638	18,18
cCP	1,1563	0,9224	1,3058	271,91	0,8592	18,18

Nota: RMSE (em m), MAE (em m), MAPE (em %), Coeficiente U de Theil, Coeficiente de Determinação R² e POCID (em %).

Fonte: Elaborado pelo autor (2022).

Tabela E.9 – Desempenho dos modelos de previsão para a série P09 no conjunto de teste (nov/2020 – out/2021)

Modelo	RMSE	MAE	MAPE	U de Theil	R ²	POCID
ARIMA	0,4992	0,3311	0,6967	2,140	0,6162	72,73
ETS	1,1447	1,0227	2,1641	19,640	0,2311	63,64
ANN	0,5769	0,5203	1,0920	4,739	0,8966	90,91
SVR	0,6771	0,5253	1,0989	6,671	0,3113	36,36
ELM	1,4815	1,1484	2,4008	32,039	0,6998	0,00
LSTM	0,3454	0,2938	0,6223	1,789	0,8457	90,91
cSA	0,4534	0,3937	0,8264	2,829	0,6559	81,82
cSM	0,3813	0,3079	0,6450	1,952	0,8350	90,91
cANN	0,5607	0,4413	0,9243	4,690	0,1182	81,82
cSVR	0,6187	0,4785	1,0003	5,527	0,7624	72,73
cMV	0,3751	0,3403	0,7160	2,042	0,8337	81,82
cCP	0,6983	0,5985	1,2544	7,064	0,0321	54,55

Nota: RMSE (em m), MAE (em m), MAPE (em %), Coeficiente U de Theil, Coeficiente de Determinação R² e POCID (em %).

Fonte: Elaborado pelo autor (2022).

Tabela E.10 – Desempenho dos modelos de previsão para a série P10 no conjunto de teste (nov/2020 – out/2021)

Modelo	RMSE	MAE	MAPE	U de Theil	R ²	POCID
ARIMA	2,061	1,766	13,67	0,9124	0,9108	81,82
ETS	3,848	3,503	26,92	3,2582	0,9244	81,82
ANN	8,331	7,144	52,52	15,2248	0,2317	72,73
SVR	2,894	2,237	16,92	1,8374	0,5554	54,55
ELM	3,798	3,223	24,17	2,9348	0,0895	36,36
LSTM	9,864	7,984	62,85	21,3710	0,0566	36,36
cSA	4,421	3,747	29,74	4,2895	0,6480	81,82
cSM	3,535	3,018	23,92	2,7375	0,8226	72,73
cANN	5,437	4,471	36,06	6,5017	0,3212	72,73
cSVR	4,623	3,781	30,36	4,6926	0,5197	72,73
cMV	6,076	5,112	40,36	8,1247	0,2552	63,64
cCP	2,583	2,097	15,93	1,3955	0,6188	54,55

Nota: RMSE (em m), MAE (em m), MAPE (em %), Coeficiente U de Theil, Coeficiente de Determinação R² e POCID (em %).

Fonte: Elaborado pelo autor (2022).

Tabela E.11 – Desempenho dos modelos de previsão para a série P11 no conjunto de teste (mar/2020 – fev/2021)

Modelo	RMSE	MAE	MAPE	U de Theil	R²	POCID
ARIMA	0,0585	0,0506	0,0699	0,5284	0,2945	72,73
ETS	0,1109	0,0876	0,1212	1,9401	0,0000	0,00
ANN	0,1572	0,1421	0,1965	3,9345	0,1659	72,73
SVR	0,1180	0,1006	0,1391	2,2181	0,0735	63,64
ELM	0,1842	0,1594	0,2205	5,1244	0,0060	36,36
LSTM	0,1324	0,1140	0,1576	2,7864	0,1270	63,64
cSA	0,0736	0,0670	0,0926	0,8441	0,1081	45,45
cSM	0,0861	0,0762	0,1053	1,1730	0,0750	54,55
cANN	0,2812	0,2459	0,3400	12,5532	0,1435	72,73
cSVR	0,1777	0,1572	0,2173	5,0249	0,1519	63,64
cMV	0,1543	0,1386	0,1917	3,7895	0,1636	72,73
cCP	0,1551	0,1390	0,1922	3,8216	0,1867	72,73

Nota: RMSE (em m), MAE (em m), MAPE (em %), Coeficiente U de Theil, Coeficiente de Determinação R² e POCID (em %).

Fonte: Elaborado pelo autor (2022).

Tabela E.12 – Desempenho dos modelos de previsão para a série P12 no conjunto de teste (mar/2020 – fev/2021)

Modelo	RMSE	MAE	MAPE	U de Theil	R²	POCID
ARIMA	0,1051	0,0913	0,3235	42,838	0,8155	100,00
ETS	0,0800	0,0681	0,2414	24,796	0,9670	100,00
ANN	0,0427	0,0402	0,1424	7,038	0,9615	100,00
SVR	0,1636	0,1251	0,4431	103,759	0,5247	18,18
ELM	0,2138	0,1913	0,6780	177,072	0,9560	0,00
LSTM	0,0641	0,0428	0,1515	15,918	0,1245	54,55
cSA	0,1089	0,0919	0,3256	45,953	0,0162	45,45
cSM	0,0939	0,0799	0,2832	34,205	0,5028	81,82
cANN	0,0263	0,0194	0,0688	2,654	0,9245	81,82
cSVR	0,0909	0,0755	0,2676	31,971	0,1853	45,45
cMV	0,0500	0,0441	0,1563	9,681	0,9676	90,91
cCP	0,1777	0,1405	0,4975	122,344	0,4842	9,09

Nota: RMSE (em m), MAE (em m), MAPE (em %), Coeficiente U de Theil, Coeficiente de Determinação R² e POCID (em %).

Fonte: Elaborado pelo autor (2022).

Tabela E.13 – Desempenho dos modelos de previsão para a série P13 no conjunto de teste (mar/2020 – fev/2021)

Modelo	RMSE	MAE	MAPE	U de Theil	R ²	POCID
ARIMA	1,367	1,2373	1,603	9,718	0,6088	36,36
ETS	1,583	1,3895	1,804	13,180	0,6634	36,36
ANN	1,528	1,3997	1,797	12,144	0,7491	72,73
SVR	1,190	0,9373	1,196	6,938	0,7389	72,73
ELM	1,140	0,9049	1,156	6,577	0,8380	72,73
LSTM	1,087	0,9927	1,277	6,036	0,1277	63,64
cSA	1,017	0,8773	1,125	5,213	0,8593	72,73
cSM	1,055	0,8944	1,145	5,655	0,6231	63,64
cANN	1,204	0,8895	1,134	7,356	0,7162	63,64
cSVR	1,156	0,8884	1,133	6,744	0,7281	63,64
cMV	1,270	1,0013	1,279	8,285	0,6864	72,73
cCP	1,253	1,1356	1,459	8,012	0,3081	45,45

Nota: RMSE (em m), MAE (em m), MAPE (em %), Coeficiente U de Theil, Coeficiente de Determinação R² e POCID (em %).

Fonte: Elaborado pelo autor (2022).

Tabela E.14 – Desempenho dos modelos de previsão para a série P14 no conjunto de teste (mar/2020 – fev/2021)

Modelo	RMSE	MAE	MAPE	U de Theil	R ²	POCID
ARIMA	0,1032	0,0963	0,2715	0,9801	0,9961	100,00
ETS	0,0957	0,0848	0,2402	0,8349	0,9786	100,00
ANN	0,2714	0,2303	0,6509	6,7710	0,8911	90,91
SVR	0,5177	0,4811	1,3553	24,3961	0,9892	90,91
ELM	0,4734	0,3878	1,0921	20,4812	0,5015	54,55
LSTM	0,2117	0,1691	0,4752	4,0938	0,9135	100,00
cSA	0,1518	0,1099	0,3082	2,1124	0,9452	100,00
cSM	0,1375	0,1114	0,3133	1,7377	0,9509	100,00
cANN	0,4031	0,2870	0,8050	14,9533	0,4421	54,55
cSVR	0,3467	0,3278	0,9244	10,6446	0,9636	90,91
cMV	0,0690	0,0559	0,1580	0,4364	0,9713	100,00
cCP	0,6139	0,4831	1,3552	34,6815	0,0993	9,09

Nota: RMSE (em m), MAE (em m), MAPE (em %), Coeficiente U de Theil, Coeficiente de Determinação R² e POCID (em %).

Fonte: Elaborado pelo autor (2022).

Tabela E.15 – Desempenho dos modelos de previsão para a série P15 no conjunto de teste (mar/2020 – fev/2021)

Modelo	RMSE	MAE	MAPE	U de Theil	R ²	POCID
ARIMA	0,2406	0,1841	0,7745	65,898	0,7948	27,27
ETS	0,3078	0,2352	0,9894	107,909	0,8461	27,27
ANN	0,1443	0,1347	0,5650	22,144	0,9163	36,36
SVR	0,1283	0,1174	0,4922	17,010	0,8201	27,27
ELM	0,3565	0,3399	1,4268	142,405	0,7724	72,73
LSTM	0,1916	0,1608	0,6727	37,402	0,8534	9,09
cSA	0,1140	0,1054	0,4421	13,693	0,6927	45,45
cSM	0,1360	0,1260	0,5286	19,395	0,8915	27,27
cANN	0,0736	0,0590	0,2471	5,967	0,4732	72,73
cSVR	0,0866	0,0784	0,3288	8,414	0,1055	54,55
cMV	0,1968	0,1624	0,6827	44,067	0,8868	72,73
cCP	0,0903	0,0811	0,3405	9,193	0,0000	0,00

Nota: RMSE (em m), MAE (em m), MAPE (em %), Coeficiente U de Theil, Coeficiente de Determinação R² e POCID (em %).

Fonte: Elaborado pelo autor (2022).

Tabela E.16 – Desempenho dos modelos de previsão para a série P16 no conjunto de teste (mar/2020 – fev/2021)

Modelo	RMSE	MAE	MAPE	U de Theil	R ²	POCID
ARIMA	0,7207	0,5657	2,327	20,516	0,2097	45,45
ETS	0,4114	0,3573	1,479	6,554	0,9031	72,73
ANN	0,4622	0,3272	1,345	8,428	0,5344	63,64
SVR	0,8011	0,6203	2,556	25,283	0,2179	36,36
ELM	0,8825	0,6701	2,752	30,759	0,7245	27,27
LSTM	0,6085	0,4875	2,009	14,521	0,1336	36,36
cSA	0,6292	0,4734	1,945	15,615	0,0018	54,55
cSM	0,6217	0,4718	1,940	15,244	0,0019	45,45
cANN	0,8438	0,6451	2,652	28,073	0,5717	45,45
cSVR	0,6157	0,4539	1,864	14,958	0,0518	54,55
cMV	0,7039	0,5217	2,142	19,572	0,0001	63,64
cCP	0,9922	0,7736	3,182	38,876	0,4274	45,45

Nota: RMSE (em m), MAE (em m), MAPE (em %), Coeficiente U de Theil, Coeficiente de Determinação R² e POCID (em %).

Fonte: Elaborado pelo autor (2022).

Tabela E.17 – Desempenho dos modelos de previsão para a série P17 no conjunto de teste (mar/2020 – fev/2021)

Modelo	RMSE	MAE	MAPE	U de Theil	R ²	POCID
ARIMA	0,0246	0,0216	0,1038	5,552e-01	0,9910	100,00
ETS	0,2308	0,1938	0,9312	4,906e+01	0,9044	0,00
ANN	0,2970	0,2774	1,3353	8,070e+01	0,5151	81,82
SVR	0,2713	0,2580	1,2424	6,720e+01	0,8284	72,73
ELM	0,5329	0,4966	2,3902	2,598e+02	0,6245	0,00
LSTM	0,1146	0,0942	0,4525	1,204e+01	0,9041	100,00
cSA	0,2414	0,2194	1,0556	5,349e+01	0,7633	72,73
cSM	0,2480	0,2238	1,0766	5,649e+01	0,7109	72,73
cANN	0,1829	0,1669	0,8032	3,079e+01	0,9365	90,91
cSVR	0,1932	0,1805	0,8686	3,406e+01	0,9850	90,91
cMV	0,0762	0,0655	0,3147	5,342e+00	0,9935	100,00
cCP	0,1867	0,1493	0,7173	3,212e+01	0,2141	18,18

Nota: RMSE (em m), MAE (em m), MAPE (em %), Coeficiente U de Theil, Coeficiente de Determinação R² e POCID (em %).

Fonte: Elaborado pelo autor (2022).

Tabela E.18 – Desempenho dos modelos de previsão para a série P18 no conjunto de teste (set/2020 – ago/2021)

Modelo	RMSE	MAE	MAPE	U de Theil	R ²	POCID
ARIMA	0,3735	0,2753	0,4573	0,6061	0,7580	63,64
ETS	0,5732	0,4778	0,7957	1,4467	0,0045	36,36
ANN	0,9691	0,7652	1,2721	4,1568	0,4083	81,82
SVR	0,9705	0,6133	1,0102	4,1498	0,1009	54,55
ELM	0,6690	0,5235	0,8687	1,9770	0,5265	27,27
LSTM	0,6392	0,4959	0,8211	1,7881	0,2289	72,73
cSA	0,5596	0,3693	0,6096	1,3784	0,1227	72,73
cSM	0,6085	0,4462	0,7386	1,6295	0,0524	54,55
cANN	1,3074	0,9858	1,6353	7,5654	0,0829	63,64
cSVR	0,9587	0,6030	0,9933	4,0520	0,0739	63,64
cMV	1,0001	0,6264	1,0317	4,4110	0,1223	63,64
cCP	0,9998	0,7681	1,2706	4,4118	0,0269	54,55

Nota: RMSE (em m), MAE (em m), MAPE (em %), Coeficiente U de Theil, Coeficiente de Determinação R² e POCID (em %).

Fonte: Elaborado pelo autor (2022).

Tabela E.19 – Desempenho dos modelos de previsão para a série P19 no conjunto de teste (mar/2020 – fev/2021)

Modelo	RMSE	MAE	MAPE	U de Theil	R²	POCID
ARIMA	0,3504	0,3284	8,530	7,263	0,8901	81,82
ETS	0,5844	0,5130	12,647	22,767	0,7580	18,18
ANN	0,3512	0,2984	7,276	8,814	0,9456	72,73
SVR	0,5708	0,5242	13,238	18,700	0,7222	81,82
ELM	0,3938	0,3317	8,683	6,269	0,5634	72,73
LSTM	0,4537	0,3731	9,985	13,453	0,4926	72,73
cSA	0,2889	0,2507	6,602	3,998	0,8446	81,82
cSM	0,3328	0,3098	8,048	6,253	0,8784	81,82
cANN	0,5605	0,5230	13,588	20,682	0,6060	81,82
cSVR	0,4309	0,3788	10,033	10,913	0,8641	90,91
cMV	0,3927	0,3483	9,219	9,421	0,9723	100,00
cCP	0,4463	0,3630	9,732	12,994	0,4725	81,82

Nota: RMSE (em m), MAE (em m), MAPE (em %), Coeficiente U de Theil, Coeficiente de Determinação R² e POCID (em %).

Fonte: Elaborado pelo autor (2022).

Tabela E.20 – Desempenho dos modelos de previsão para a série P20 no conjunto de teste (mar/2020 – fev/2021)

Modelo	RMSE	MAE	MAPE	U de Theil	R²	POCID
ARIMA	0,1672	0,1487	1,1549	0,8269	0,9486	90,91
ETS	0,2048	0,1656	1,2942	1,1902	0,9576	100,00
ANN	0,1284	0,1138	0,8779	0,4972	0,9381	90,91
SVR	0,1079	0,0817	0,6158	0,3571	0,9483	81,82
ELM	0,1854	0,1280	0,9698	1,0592	0,8053	100,00
LSTM	0,0582	0,0528	0,4071	0,1031	0,9908	100,00
cSA	0,0857	0,0717	0,5530	0,2083	0,9752	100,00
cSM	0,0664	0,0580	0,4453	0,1268	0,9837	100,00
cANN	0,1141	0,1061	0,8165	0,3828	0,9285	81,82
cSVR	0,0964	0,0802	0,6065	0,2686	0,9559	81,82
cMV	0,1141	0,1056	0,8100	0,3810	0,9415	81,82
cCP	0,1117	0,0961	0,7374	0,3722	0,9211	72,73

Nota: RMSE (em m), MAE (em m), MAPE (em %), Coeficiente U de Theil, Coeficiente de Determinação R² e POCID (em %).

Fonte: Elaborado pelo autor (2022).

Tabela E.21 – Desempenho dos modelos de previsão para a série P21 no conjunto de teste (mar/2020 – fev/2021)

Modelo	RMSE	MAE	MAPE	U de Theil	R²	POCID
ARIMA	0,2399	0,2228	2,813	1,0629	0,9608	81,82
ETS	0,2732	0,2275	2,962	1,3018	0,9294	81,82
ANN	0,2604	0,2122	2,636	1,3340	0,7685	72,73
SVR	0,1424	0,1100	1,377	0,4001	0,9187	72,73
ELM	0,3869	0,2743	3,431	2,9819	0,6442	72,73
LSTM	0,2284	0,1838	2,375	0,9606	0,7606	81,82
cSA	0,1404	0,1247	1,593	0,3519	0,9113	81,82
cSM	0,1289	0,1079	1,391	0,2893	0,9215	81,82
cANN	0,1949	0,1666	2,084	0,7409	0,8525	81,82
cSVR	0,2305	0,1609	1,979	1,0573	0,7685	63,64
cMV	0,1154	0,0941	1,190	0,2492	0,9466	72,73
cCP	0,4001	0,2375	2,913	3,1865	0,4732	90,91

Nota: RMSE (em m), MAE (em m), MAPE (em %), Coeficiente U de Theil, Coeficiente de Determinação R² e POCID (em %).

Fonte: Elaborado pelo autor (2022).

Tabela E.22 – Classificação dos modelos por medida de desempenho (MD): série P01

MD	ARIMA	ETS	ANN	SVR	ELM	LSTM	cSA	cSM	cANN	cSVR	cMV	cCP
RMSE	11	12	2	5	9	1	6	7	3	4	8	10
MAE	11	12	2	6	9	1	4	7	3	5	8	10
MAPE	11	12	2	6	9	1	5	7	3	4	8	10
U	11	12	2	5	9	1	6	7	3	4	8	10
R ²	2	9	1	10	12	6	3	5	4	8	11	7
POCID	3	2	1	2	5	1	1	1	1	4	5	3

Fonte: Elaborado pelo autor (2022).

Tabela E.23 – Classificação dos modelos por medida de desempenho (MD): série P02

MD	ARIMA	ETS	ANN	SVR	ELM	LSTM	cSA	cSM	cANN	cSVR	cMV	cCP
RMSE	11	5	1	3	10	12	7	6	4	2	8	9
MAE	11	4	1	3	9	10	7	6	5	2	8	8
MAPE	12	4	1	3	10	11	7	6	5	2	9	8
U	11	5	1	3	10	12	7	6	4	2	8	9
R ²	7	11	1	4	12	10	3	8	2	9	5	6
POCID	2	3	1	1	3	4	2	3	2	3	3	3

Fonte: Elaborado pelo autor (2022).

Tabela E.24 – Classificação dos modelos por medida de desempenho (MD): série P03

MD	ARIMA	ETS	ANN	SVR	ELM	LSTM	cSA	cSM	cANN	cSVR	cMV	cCP
RMSE	10	9	4	11	12	1	2	5	6	8	3	7
MAE	9	7	3	10	11	1	2	4	5	8	4	6
MAPE	11	9	3	12	10	1	2	5	6	8	4	7
U	11	10	4	9	12	1	2	5	6	8	3	7
R ²	5	3	1	9	10	8	7	6	4	12	2	11
POCID	2	2	1	1	4	3	2	1	2	4	1	2

Fonte: Elaborado pelo autor (2022).

Tabela E.25 – Classificação dos modelos por medida de desempenho (MD): série P04

MD	ARIMA	ETS	ANN	SVR	ELM	LSTM	cSA	cSM	cANN	cSVR	cMV	cCP
RMSE	12	8	10	6	3	4	5	1	11	7	2	9
MAE	12	8	10	11	3	5	4	2	9	6	1	7
MAPE	12	5	10	11	3	6	4	1	9	7	2	8
U	12	8	10	6	3	4	5	1	11	7	2	9
R ²	12	5	9	11	3	6	2	1	10	7	4	8
POCID	5	3	2	2	4	3	2	1	3	3	3	3

Fonte: Elaborado pelo autor (2022).

Tabela E.26 – Classificação dos modelos por medida de desempenho (MD): série P05

MD	ARIMA	ETS	ANN	SVR	ELM	LSTM	cSA	cSM	cANN	cSVR	cMV	cCP
RMSE	11	12	8	1	10	2	5	9	3	7	4	6
MAE	11	12	8	1	9	2	6	10	4	7	5	3
MAPE	11	12	8	1	9	2	6	10	4	7	5	3
U	11	12	8	1	10	2	6	9	3	7	4	5
R ²	1	2	9	4	7	5	4	11	6	3	8	10
POCID	1	1	1	1	4	2	2	3	3	2	2	4

Fonte: Elaborado pelo autor (2022).

Tabela E.27 – Classificação dos modelos por medida de desempenho (MD): série P06

MD	ARIMA	ETS	ANN	SVR	ELM	LSTM	cSA	cSM	cANN	cSVR	cMV	cCP
RMSE	1	8	2	10	9	11	4	5	6	3	7	12
MAE	1	7	2	10	9	11	4	5	6	3	8	12
MAPE	1	7	2	10	9	11	4	5	6	3	8	12
U	1	8	2	10	9	11	4	5	6	3	7	12
R ²	4	1	7	6	12	3	2	11	8	9	10	5
POCID	3	1	3	4	2	5	1	3	4	1	2	4

Fonte: Elaborado pelo autor (2022).

Tabela E.28 – Classificação dos modelos por medida de desempenho (MD): série P07

MD	ARIMA	ETS	ANN	SVR	ELM	LSTM	cSA	cSM	cANN	cSVR	cMV	cCP
RMSE	4	6	7	8	9	1	3	5	10	2	12	11
MAE	2	5	7	9	8	1	3	6	10	4	12	11
MAPE	2	5	7	9	8	1	3	6	10	4	12	11
U	4	6	7	8	9	1	3	5	10	2	12	11
R ²	1	12	7	9	6	2	3	5	8	4	11	10
POCID	2	4	2	1	1	3	2	2	1	2	1	1

Fonte: Elaborado pelo autor (2022).

Tabela E.29 – Classificação dos modelos por medida de desempenho (MD): série P08

MD	ARIMA	ETS	ANN	SVR	ELM	LSTM	cSA	cSM	cANN	cSVR	cMV	cCP
RMSE	3	7	9	8	11	4	2	6	5	1	10	12
MAE	3	7	9	8	11	4	2	6	5	1	10	12
MAPE	3	7	9	8	11	4	2	6	5	1	10	12
U	3	7	9	8	11	4	2	6	5	1	10	12
R ²	9	1	8	3	12	10	11	5	4	2	7	6
POCID	5	5	1	4	5	3	2	4	4	3	3	3

Fonte: Elaborado pelo autor (2022).

Tabela E.30 – Classificação dos modelos por medida de desempenho (MD): série P09

MD	ARIMA	ETS	ANN	SVR	ELM	LSTM	cSA	cSM	cANN	cSVR	cMV	cCP
RMSE	5	11	7	9	12	1	4	3	6	8	2	10
MAE	3	11	8	9	12	1	5	2	6	7	4	10
MAPE	3	11	8	9	12	1	5	2	6	7	4	10
U	4	11	7	9	12	1	5	2	6	8	3	10
R ²	8	10	1	9	6	2	7	3	11	5	4	12
POCID	3	4	1	6	7	1	2	1	2	3	2	5

Fonte: Elaborado pelo autor (2022).

Tabela E.31 – Classificação dos modelos por medida de desempenho (MD): série P10

MD	ARIMA	ETS	ANN	SVR	ELM	LSTM	cSA	cSM	cANN	cSVR	cMV	cCP
RMSE	1	6	11	3	5	12	7	4	9	8	10	2
MAE	1	6	11	3	5	12	7	4	9	8	10	2
MAPE	1	6	11	3	5	12	7	4	9	8	10	2
U	1	6	11	3	5	12	7	4	9	8	10	2
R ²	2	1	10	6	11	12	4	3	8	7	9	5
POCID	1	1	2	4	5	5	1	2	2	2	3	4

Fonte: Elaborado pelo autor (2022).

Tabela E.32 – Classificação dos modelos por medida de desempenho (MD): série P11

MD	ARIMA	ETS	ANN	SVR	ELM	LSTM	cSA	cSM	cANN	cSVR	cMV	cCP
RMSE	1	4	9	5	11	6	2	3	12	10	7	8
MAE	1	4	9	5	11	6	2	3	12	10	7	8
MAPE	1	4	9	5	11	6	2	3	12	10	7	8
U	1	4	9	5	11	6	2	3	12	10	7	8
R ²	1	12	3	10	11	7	8	9	6	5	4	2
POCID	1	6	1	2	5	2	4	3	1	2	1	1

Fonte: Elaborado pelo autor (2022).

Tabela E.33 – Classificação dos modelos por medida de desempenho (MD): série P12

MD	ARIMA	ETS	ANN	SVR	ELM	LSTM	cSA	cSM	cANN	cSVR	cMV	cCP
RMSE	8	5	2	10	12	4	9	7	1	6	3	11
MAE	8	5	2	10	12	3	9	7	1	6	4	11
MAPE	8	5	2	10	12	3	9	7	1	6	4	11
U	8	5	2	10	12	4	9	7	1	6	3	11
R ²	6	2	3	7	4	11	12	8	5	10	1	9
POCID	1	1	1	6	8	4	5	3	3	5	2	7

Fonte: Elaborado pelo autor (2022).

Tabela E.34 – Classificação dos modelos por medida de desempenho (MD): série P13

MD	ARIMA	ETS	ANN	SVR	ELM	LSTM	cSA	cSM	cANN	cSVR	cMV	cCP
RMSE	10	12	11	6	4	3	1	2	7	5	9	8
MAE	10	11	12	6	5	7	1	4	3	2	8	9
MAPE	10	12	11	6	5	7	1	4	3	2	8	9
U	10	12	11	6	4	3	1	2	7	5	9	8
R ²	10	8	3	4	2	12	1	9	6	5	7	11
POCID	4	4	1	1	1	2	1	2	2	2	1	3

Fonte: Elaborado pelo autor (2022).

Tabela E.35 – Classificação dos modelos por medida de desempenho (MD): série P14

MD	ARIMA	ETS	ANN	SVR	ELM	LSTM	cSA	cSM	cANN	cSVR	cMV	cCP
RMSE	3	2	7	11	10	6	5	4	9	8	1	12
MAE	3	2	7	11	10	6	4	5	8	9	1	12
MAPE	3	2	7	12	10	6	4	5	8	9	1	11
U	3	2	7	11	10	6	5	4	9	8	1	12
R ²	1	3	9	2	10	8	7	6	11	5	4	12
POCID	1	1	2	2	3	1	1	1	3	2	1	4

Fonte: Elaborado pelo autor (2022).

Tabela E.36 – Classificação dos modelos por medida de desempenho (MD): série P15

MD	ARIMA	ETS	ANN	SVR	ELM	LSTM	cSA	cSM	cANN	cSVR	cMV	cCP
RMSE	10	11	7	5	12	8	4	6	1	2	9	3
MAE	10	11	7	5	12	8	4	6	1	2	9	3
MAPE	10	11	7	5	12	8	4	6	1	2	9	3
U	10	11	7	5	12	8	4	6	1	2	9	3
R ²	7	5	1	6	8	4	9	2	10	11	3	12
POCID	5	5	4	5	1	6	3	5	1	2	1	7

Fonte: Elaborado pelo autor (2022).

Tabela E.37 – Classificação dos modelos por medida de desempenho (MD): série P16

MD	ARIMA	ETS	ANN	SVR	ELM	LSTM	cSA	cSM	cANN	cSVR	cMV	cCP
RMSE	8	1	2	9	11	3	6	5	10	4	7	12
MAE	8	2	1	9	11	6	5	4	10	3	7	12
MAPE	8	2	1	9	11	6	5	4	10	3	7	12
U	8	1	2	9	11	3	6	5	10	4	7	12
R ²	7	1	4	6	2	8	11	10	3	9	12	5
POCID	4	1	2	5	6	5	3	4	4	3	2	4

Fonte: Elaborado pelo autor (2022).

Tabela E.38 – Classificação dos modelos por medida de desempenho (MD): série P17

MD	ARIMA	ETS	ANN	SVR	ELM	LSTM	cSA	cSM	cANN	cSVR	cMV	cCP
RMSE	1	7	11	10	12	3	8	9	4	6	2	5
MAE	1	7	11	10	12	3	8	9	5	6	2	4
MAPE	1	7	11	10	12	3	8	9	5	6	2	4
U	1	7	11	10	12	3	8	9	4	6	2	5
R ²	2	5	11	7	10	6	8	9	4	3	1	12
POCID	1	6	3	4	6	1	4	4	2	2	1	5

Fonte: Elaborado pelo autor (2022).

Tabela E.39 – Classificação dos modelos por medida de desempenho (MD): série P18

MD	ARIMA	ETS	ANN	SVR	ELM	LSTM	cSA	cSM	cANN	cSVR	cMV	cCP
RMSE	1	3	8	9	6	5	2	4	12	7	11	10
MAE	1	4	10	8	6	5	2	3	12	7	9	11
MAPE	1	4	11	8	6	5	2	3	12	7	9	10
U	1	3	9	8	6	5	2	4	12	7	10	11
R ²	1	12	3	7	2	4	5	10	8	9	6	11
POCID	3	5	1	4	6	2	2	4	3	3	3	4

Fonte: Elaborado pelo autor (2022).

Tabela E.40 – Classificação dos modelos por medida de desempenho (MD): série P19

MD	ARIMA	ETS	ANN	SVR	ELM	LSTM	cSA	cSM	cANN	cSVR	cMV	cCP
RMSE	3	12	4	11	6	9	1	2	10	7	5	8
MAE	4	10	2	12	5	8	1	3	11	9	6	7
MAPE	4	10	2	11	5	8	1	3	12	9	6	7
U	4	12	5	10	3	9	1	2	11	7	6	8
R ²	3	7	2	8	10	11	6	4	9	5	1	12
POCID	3	5	4	3	4	4	3	3	3	2	1	3

Fonte: Elaborado pelo autor (2022).

Tabela E.41 – Classificação dos modelos por medida de desempenho (MD): série P20

MD	ARIMA	ETS	ANN	SVR	ELM	LSTM	cSA	cSM	cANN	cSVR	cMV	cCP
RMSE	9	11	8	5	10	1	3	2	7	4	7	6
MAE	11	12	9	5	10	1	3	2	8	4	7	6
MAPE	11	12	9	5	10	1	3	2	8	4	7	6
U	10	12	9	5	11	1	3	2	8	4	7	6
R ²	6	4	9	7	12	1	3	2	10	5	8	11
POCID	2	1	2	3	1	1	1	1	3	3	3	4

Fonte: Elaborado pelo autor (2022).

Tabela E.42 – Classificação dos modelos por medida de desempenho (MD): série P21

MD	ARIMA	ETS	ANN	SVR	ELM	LSTM	cSA	cSM	cANN	cSVR	cMV	cCP
RMSE	8	10	9	4	11	6	3	2	5	7	1	12
MAE	9	10	8	3	12	7	4	2	6	5	1	11
MAPE	9	11	8	2	12	7	4	3	6	5	1	10
U	8	9	10	4	11	6	3	2	5	7	1	12
R ²	1	3	8	5	10	9	6	4	7	8	2	11
POCID	2	2	3	3	3	2	2	2	2	4	3	1

Fonte: Elaborado pelo autor (2022).