

Patrícia de Oliveira e Lucas

Previsão de Séries Temporais de
Evapotranspiração de Referência com
Redes Neurais Convolucionais

Belo Horizonte - Minas Gerais

Dezembro, 2019

Patrícia de Oliveira e Lucas

Previsão de Séries Temporais de Evapotranspiração de Referência com Redes Neurais Convolucionais

Dissertação de mestrado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Escola de Engenharia da Universidade Federal de Minas Gerais, como obtenção do Título de Mestre em Engenharia Elétrica.

Universidade Federal de Minas Gerais - UFMG

Programa de Pós-Graduação em Engenharia Elétrica - PPGEE

Machine Intelligence and Data Science Laboratory - MINDS

Orientador: Frederico Gadelha Guimarães

Belo Horizonte - Minas Gerais

Dezembro, 2019

L933p

Lucas, Patrícia de Oliveira e.

Previsão de séries temporais de evapotranspiração de referência com redes neurais convolucionais [recurso eletrônico] / Patrícia de Oliveira e Lucas. - 2019.

1 recurso online (100 f. : il., color.) : pdf.

Orientador: Frederico Gadelha Guimarães.

Dissertação (mestrado) - Universidade Federal de Minas Gerais, Escola de Engenharia.

Bibliografia: f. 93-100.

Exigências do sistema: Adobe Acrobat Reader.

1. Engenharia elétrica - Teses. 2. Previsão - Teses. 3. Redes neurais convolucionais - Teses. 4. Séries temporais - Teses. I. Guimarães, Frederico Gadelha. II. Universidade Federal de Minas Gerais. Escola de Engenharia. III. Título.

CDU: 621.3(043)



UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE ENGENHARIA
Programa de Pós-Graduação em Engenharia Elétrica

ATA DA 1162ª DEFESA DE DISSERTAÇÃO DE MESTRADO
DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

ATA DE DEFESA DE DISSERTAÇÃO DE MESTRADO da aluna PATRÍCIA DE OLIVEIRA E LUCAS - registro de matrícula de número 2018687110. Às 14:00 horas do dia 18 do mês de dezembro de 2019, reuniu-se na Escola de Engenharia da UFMG a Comissão Examinadora da DISSERTAÇÃO DE MESTRADO para julgar, em exame final, o trabalho intitulado "Previsão de Séries Temporais de Evapotranspiração de Referência com Redes Neurais Convolucionais" da Área de Concentração em Sistemas de Computação e Telecomunicações, Linha de Pesquisa Inteligência Computacional. O Prof. Frederico Gadelha Guimarães, orientador da aluna, abriu a sessão apresentando os membros da Comissão e, dando continuidade aos trabalhos, informou aos presentes que, de acordo com o Regulamento do Programa no seu Art. 8.16, será considerado APROVADO na defesa da Dissertação de Mestrado o candidato que obtiver a aprovação unânime dos membros da Comissão Examinadora. Em seguida deu início à apresentação do trabalho pela Candidata. Ao final da apresentação seguiu-se a arguição da candidata pelos examinadores. Logo após o término da arguição a Comissão Examinadora se reuniu, sem a presença da Candidata e do público, e elegeu o Prof. Frederico Gadelha Guimarães para presidir a fase de avaliação do trabalho, constituída de deliberação individual de APROVAÇÃO ou de REPROVAÇÃO e expedição do resultado final. As deliberações individuais de cada membro da Comissão Examinadora foram as seguintes:

Membro da Comissão Examinadora	Instituição de Origem	Deliberação	Assinatura
Prof. Dr. Frederico Gadelha Guimarães - Orientador	DEE (UFMG)	Aprovada	
Prof. Dr. Eduardo Mazoni Andrade Marçal Mendes	DELT (UFMG)	Aprovada	
Prof. Dr. Petrônio Cândido de Lima e Silva	(IFNMG)	Aprovada	
Prof. Dr. Rodrigo César Pedrosa Silva	DECOM (UFOP)	Aprovada	

Tendo como base as deliberações dos membros da Comissão Examinadora a Dissertação de Mestrado foi aprovada. O resultado final de aprovada foi comunicado publicamente à Candidata pelo Presidente da Comissão, ressaltando que a obtenção do Grau de Mestre em ENGENHARIA ELÉTRICA fica condicionada à entrega do TEXTO FINAL da Dissertação de Mestrado. A Candidata terá um prazo máximo de 30 (trinta) dias, a partir desta data, para fazer as CORREÇÕES DE FORMA e entregar o texto final da Dissertação de Mestrado na secretaria do PPGE/UFMG. As correções de forma exigidas pelos membros da Comissão Examinadora deverão ser registradas em um exemplar do texto da Dissertação de Mestrado, cuja verificação ficará sob a responsabilidade do Presidente da Banca Examinadora. Nada mais havendo a tratar o Presidente encerrou a reunião e lavrou a presente ATA, que será assinada pelo Presidente da Comissão Examinadora. Belo Horizonte, 18 de dezembro de 2019.

ASSINATURA DO PRESIDENTE DA COMISSÃO EXAMINADORA

À Duda, ao Marcelo e ao Loki, pela colaboração, incentivo e amor.

Agradecimentos

Primeiro de tudo, gostaria de agradecer a Deus por me guiar, iluminar e me dar tranquilidade para seguir em frente com os meus objetivos e não desanimar com as dificuldades. Agradeço a Ele também por manter a minha mãe ao meu lado, com a saúde que ela está hoje.

Agradeço aos meus pais e às minhas irmãs por sempre terem me apoiado na minha formação, me proporcionando a continuidade nos estudos. A vocês, meus eternos agradecimentos!

À Duda, Marcelo e Loki por estarem sempre comigo, incondicionalmente, sempre me fazendo acreditar que chegaria ao final desta difícil, porém gratificante etapa. Sou grata por cada gesto carinhoso, pelo apoio, pela compreensão e alegria. Vocês contribuíram imensamente para a realização desse trabalho!

Ao Bergston, Jaciele e Niara pela amizade, pelos momentos de descontração e principalmente por terem me proporcionado um lar!

Minha gratidão especial ao Prof. Dr. Frederico Gadelha Guimarães, meu orientador, pela pessoa e profissional que é. Obrigada pela sua dedicação, paciência, ensinamentos e pelo incentivo em todas as nossas reuniões!

Aos colegas do MINDS, LACSED e do LITC, que me acolheram, que muitas vezes me ajudaram e me orientaram nessa jornada, que compartilham conhecimento comigo e me ensinaram muitas coisas. Sou imensamente grata a vocês!

Resumo

O crescimento da população e as mudanças climáticas estão fazendo com que o setor agrícola busque abordagens mais precisas e eficientes para garantir um suprimento adequado e regular de alimentos para a sociedade com menor consumo de água. A agricultura 4.0 surge nessa conjuntura de escassez de recursos como uma gestão que busca através de tecnologias como Big Data, Internet das Coisas (IoT), Inteligência Artificial e Robótica fornecer às plantas e animais exatamente o que eles precisam e quando precisam, aumentando a produtividade e reduzindo os impactos ambientais. O manejo da irrigação, prática essencial para o desenvolvimento de uma agricultura sustentável, busca, através da previsão da evapotranspiração de referência, conhecer antecipadamente as necessidades de água das culturas para planejar e gerenciar os recursos hídricos. Esta dissertação se insere nesse contexto, com o objetivo de investigar o uso de modelos de aprendizagem profunda, em especial das redes neurais convolucionais, na previsão de séries temporais de evapotranspiração de referência (ET_o). Para isso, foram implementadas três redes neurais convolucionais com diferentes estruturas para prever uma série temporal diária de ET_o . Para otimizar os hiperparâmetros desses modelos foi utilizado um algoritmo genético que buscou equilibrar dois objetivos, precisão e parcimônia. A validação dos modelos CNN foi feita através da comparação com modelos conhecidos de previsão de séries temporais como o ARIMA, WFTS e LSTM. Para fins de comparação foram implementados também *ensemble learning* compostos pelos modelos CNN. Os resultados mostraram que os modelos CNN são viáveis para previsão de séries temporais de ET_o e que os modelos *ensemble* melhoram as previsões em termos de variância, precisão e custo computacional em relação aos modelos individuais.

Palavras-chave: evapotranspiração de referência, previsão, redes neurais convolucionais, *ensemble learning*.

Abstract

Population growth and climate change are causing the agricultural sector to seek more accurate and efficient approaches to ensure an adequate and regular supply of food to society with less water consumption. Agriculture 4.0 comes at this context of resource scarcity as a management that seeks through technologies such as Big Data, Internet of Things (IoT), Artificial Intelligence and Robotics to provide plants and animals with exactly what they need and when they need it, increasing productivity and reducing environmental impacts. Irrigation management, an essential practice for the development of sustainable agriculture, seeks, through reference evapotranspiration forecasting, to know in advance the water needs of crops to plan and manage water resources. This dissertation is inserted in this context, aiming to investigate the use of deep learning models, especially convolutional neural networks, in the prediction of reference evapotranspiration time series (ET_o). For this, three convolutional neural networks with different structures were implemented to predict a daily time series of ET_o . To optimize the hyperparameters of these models a genetic algorithm was used, it sought to balance two objectives, precision and parsimony. The CNN models were validated by comparing them with known time series forecasting models such as ARIMA, WFTS and LSTM. For comparison purposes, ensemble learning with the CNN models was also implemented. The results showed that CNN models are feasible for ET_o time series forecasting and that ensemble models improve predictions in terms of variance, accuracy, and computational cost over individual models.

Keywords: reference evapotranspiration, time series forecasting, convolutional neural networks, ensemble learning.

Lista de Ilustrações

Figura 1 – Variação anual da radiação extraterrestre (R_a) no Equador, a 20° e 40° de latitude norte e sul.	32
Figura 2 – Pressão de saturação de vapor em função da temperatura.	33
Figura 3 – Ilustração do efeito da velocidade do vento sobre a evapotranspiração em condições climáticas quente e seca e úmida.	34
Figura 4 – Série de Evapotranspiração de Referência da estação meteorológica de Sete Lagoas - MG.	40
Figura 5 – Gráfico das funções de autocorrelação e autocorrelação parcial da série temporal de Evapotranspiração de Referência.	41
Figura 6 – Arquitetura da LeNet-5.	44
Figura 7 – Operação de agrupamento.	46
Figura 8 – Arquitetura TCN.	48
Figura 9 – Camada convolucional causal.	49
Figura 10 – Camada convolucional causal dilatada.	49
Figura 11 – Etapas do ensemble learning	50
Figura 12 – Construção do KDE.	52
Figura 13 – Efeito da largura de banda (h) sobre a suavização da curva resultante do KDE.	52
Figura 14 – Fluxograma do algoritmo genético	56
Figura 15 – Método de janela deslizante.	58
Figura 16 – Convergência dos objetivos 1 e 2 durante a otimização de hiperparâmetros da CNN1.	60
Figura 17 – Interface dos modelos preditivos	63
Figura 18 – Boxplot do RMSE dos modelos CNN1, CNN2, CNN3 e LSTM para os horizontes de previsão 1, 3, 7 e 10 do ano de 2012.	65
Figura 19 – Visualização para o Teste de Friedman dos resultados dos modelos CNN e LSTM do ano de 2012.	65
Figura 20 – Boxplot do RMSE dos modelos CNN1, CNN2, CNN3 e LSTM para os horizontes de previsão 1, 3, 7 e 10 do ano de 2013.	66
Figura 21 – Visualização para o Teste de Friedman dos resultados dos modelos CNN e LSTM do ano de 2013.	66

Figura 22 – Boxplot do RMSE dos modelos CNN1, CNN2, CNN3 e LSTM para os horizontes de previsão 1, 3, 7 e 10 do ano de 2014.	67
Figura 23 – Visualização para o Teste de Friedman dos resultados dos modelos CNN e LSTM do ano de 2014.	67
Figura 24 – Boxplot do RMSE dos modelos CNN1, CNN2, CNN3 e LSTM para os horizontes de previsão 1, 3, 7 e 10 do ano de 2015.	68
Figura 25 – Visualização para o Teste de Friedman dos resultados dos modelos CNN e LSTM do ano de 2015.	68
Figura 26 – Boxplot do RMSE dos modelos CNNs e Ensembles para os horizontes de previsão 1, 3, 7 e 10 do ano de 2012.	71
Figura 27 – Visualização para o Teste de Friedman dos resultados dos modelos CNNs e Ensembles do ano de 2012.	72
Figura 28 – Desempenho do modelo Ensemble CNN1 para horizonte de previsão 1 do ano de teste 2012.	72
Figura 29 – Desempenho do modelo Ensemble CNN2 para horizonte de previsão 1 do ano de teste 2012.	73
Figura 30 – Desempenho do modelo Ensemble CNN3 para horizonte de previsão 1 do ano de teste 2012.	73
Figura 31 – Desempenho do modelo Ensemble Híbrido para horizonte de previsão 1 do ano de teste 2012.	73
Figura 32 – Boxplot do RMSE dos modelos CNNs e Ensembles para os horizontes de previsão 1, 3, 7 e 10 do ano de 2013.	74
Figura 33 – Visualização para o Teste de Friedman dos resultados dos modelos CNNs e Ensembles do ano de 2013.	74
Figura 34 – Desempenho do modelo Ensemble CNN1 para no horizontes de previsão 1 do ano de teste 2013.	75
Figura 35 – Desempenho do modelo Ensemble CNN2 para no horizontes de previsão 1 do ano de teste 2013.	75
Figura 36 – Desempenho do modelo Ensemble CNN3 para no horizontes de previsão 1 do ano de teste 2013.	75
Figura 37 – Desempenho do modelo Ensemble Híbrido para no horizontes de previsão 1 do ano de teste 2013.	76
Figura 38 – Boxplot do RMSE dos modelos CNNs e Ensembles para os horizontes de previsão 1, 3, 7 e 10 do ano de 2014.	76
Figura 39 – Visualização para o Teste de Friedman dos resultados dos modelos CNNs e Ensembles do ano de 2014.	77
Figura 40 – Desempenho do modelo Ensemble CNN1 para no horizontes de previsão 1 do ano de teste 2014.	77

Figura 41 – Desempenho do modelo Ensemble CNN2 no horizontes de previsão 1 do ano de teste 2014.	77
Figura 42 – Desempenho do modelo Ensemble CNN3 para no horizontes de previsão 1 do ano de teste 2014.	78
Figura 43 – Desempenho do modelo Ensemble Híbrido para no horizontes de previsão 1 do ano de teste 2014.	78
Figura 44 – Boxplot do RMSE dos modelos CNNs e Ensembles para os horizontes de previsão 1, 3, 7 e 10 do ano de 2015.	78
Figura 45 – Visualização para o Teste de Friedman dos resultados dos modelos CNNs e Ensembles do ano de 2015.	79
Figura 46 – Desempenho do modelo Ensemble CNN1 para o horizonte de previsão 1 do ano de teste 2015.	79
Figura 47 – Desempenho do modelo Ensemble CNN2 para o horizonte de previsão 1 do ano de teste 2015.	79
Figura 48 – Desempenho do modelo Ensemble CNN3 para o horizonte de previsão 1 do ano de teste 2015.	80
Figura 49 – Desempenho do modelo Ensemble Híbrido para o horizonte de previsão 1 do ano de teste 2015.	80
Figura 50 – Desempenho do modelo Ensemble CNN3 e do método Naïve para o horizontes de previsão 10 do ano de teste 2015.	82
Figura 51 – Resíduos do modelo CNN1.	83
Figura 52 – Resíduos do modelo CNN2.	84
Figura 53 – Resíduos do modelo CNN3.	85
Figura 54 – Resíduos do modelo Ensemble CNN1.	86
Figura 55 – Resíduos do modelo Ensemble CNN2.	87
Figura 56 – Resíduos do modelo Ensemble CNN3.	88
Figura 57 – Resíduos do modelo Ensemble CNN Híbrido.	89

Lista de Tabelas

Tabela 1 – Estatísticas da série temporal de Evapotranspiração de Referência. . .	40
Tabela 2 – Hiperparâmetros em comum das redes CNN1, CNN2 e CNN3.	55
Tabela 3 – Hiperparâmetros específicos da CNN1.	55
Tabela 4 – Hiperparâmetro específico da CNN2.	55
Tabela 5 – Hiperparâmetros específicos da CNN3.	55
Tabela 6 – Hiperparâmetros fixos em comum das redes CNN1, CNN2 e CNN3. . .	55
Tabela 7 – Hiperparâmetros fixos da CNN1.	55
Tabela 8 – Espaços de busca dos hiperparâmetros da CNN1.	57
Tabela 9 – Espaços de busca dos hiperparâmetros da CNN2.	57
Tabela 10 – Espaços de busca dos hiperparâmetros da CNN3.	57
Tabela 11 – Resultado da otimização dos hiperparâmetros da CNN1.	59
Tabela 12 – Resultado da otimização dos hiperparâmetros da CNN2.	60
Tabela 13 – Resultado da otimização dos hiperparâmetros da CNN3.	60
Tabela 14 – Espaços de busca e resultado da otimização dos hiperparâmetros da LSTM.	61
Tabela 15 – Espaços de busca e resultado da otimização dos hiperparâmetros do WFTS.	61
Tabela 16 – RMSE em mmd^{-1} da previsão de 1, 3, 7 e 10 dias a frente dos modelos CNN1, CNN2, CNN3, LSTM, ARIMA e WFTS para o ano de 2012. . .	64
Tabela 17 – RMSE em mmd^{-1} da previsão de 1, 3, 7 e 10 dias a frente dos modelos CNN1, CNN2, CNN3, LSTM, ARIMA e WFTS para o ano de 2013. . .	65
Tabela 18 – RMSE em mmd^{-1} da previsão de 1, 3, 7 e 10 dias a frente dos modelos CNN1, CNN2, CNN3, LSTM, ARIMA e WFTS para o ano de 2014. . .	66
Tabela 19 – RMSE em mmd^{-1} da previsão de 1, 3, 7 e 10 dias a frente dos modelos CNN1, CNN2, CNN3, LSTM, ARIMA e WFTS para o ano de 2015. . .	67
Tabela 20 – Composição dos <i>ensemble</i>	69
Tabela 21 – RMSE da previsão dos modelos CNN1, Ensemble CNN1, CNN2, En- semble CNN2, CNN3, Ensemble CNN3 e Ensemble híbrido para o ano de 2012.	70

Tabela 22 – Scaled RMSE da previsão de 1, 3, 7 e 10 dias a frente dos modelos Ensemble CNN1, Ensemble CNN2, Ensemble CNN3 e Ensemble híbrido para o ano de 2012.	71
Tabela 23 – RMSE da previsão dos modelos CNN1, Ensemble CNN1, CNN2, Ensemble CNN2, CNN3, Ensemble CNN3 e Ensemble híbrido para o ano de 2013.	71
Tabela 24 – RMSE da previsão dos modelos CNN1, Ensemble CNN1, CNN2, Ensemble CNN2, CNN3, Ensemble CNN3 e Ensemble híbrido para o ano de 2014.	71
Tabela 25 – RMSE da previsão dos modelos CNN1, Ensemble CNN1, CNN2, Ensemble CNN2, CNN3, Ensemble CNN3 e Ensemble híbrido para o ano de 2015.	72
Tabela 26 – Scaled RMSE da previsão de 1, 3, 7 e 10 dias a frente dos modelos CNN1, CNN2 e CNN3 para os anos de 2012, 2013, 2014 e 2015.	81
Tabela 27 – Scaled RMSE da previsão de 1, 3, 7 e 10 dias a frente dos modelos Ensemble CNN1, Ensemble CNN2, Ensemble CNN3 e Ensemble híbrido para os anos de 2012, 2013, 2014 e 2015.	81
Tabela 28 – Teste Ljung-Box para os resíduos do modelo CNN1.	82
Tabela 29 – Teste Ljung-Box para os resíduos do modelo CNN2.	82
Tabela 30 – Teste Ljung-Box para os resíduos do modelo CNN3.	82
Tabela 31 – Teste Ljung-Box para os resíduos do modelo Ensemble CNN1.	83
Tabela 32 – Teste Ljung-Box para os resíduos do modelo Ensemble CNN2.	83
Tabela 33 – Teste Ljung-Box para os resíduos do modelo Ensemble CNN3.	83
Tabela 34 – Teste Ljung-Box para os resíduos do modelo Ensemble CNN Híbrido.	84

Lista de Abreviaturas e Siglas

ABNT	Associação Brasileira de Normas Técnicas
<i>ET</i>	Evapotranspiração
ET_{FAO}	Evapotranspiração de referência padrão FAO
ET_o	Evapotranspiração de referência
FAO	Organização das Nações Unidas para Alimentação e Agricultura
MINDS	Machine Intelligence and Data Science
PM-FAO	Equação Penman-Monteith parametrizada
UFMG	Universidade Federal de Minas Gerais
INMET	Instituto Nacional de Meteorologia
ONU	Organização das Nações Unidas
IoT	Internet das coisas
DL	<i>Deep Learning</i>
ARIMA	Modelo auto-regressivo integrado de médias móveis
LSTM	<i>Long Short Term Memory</i>
CNN	<i>Convolutional Neural network</i>
TCN	<i>Temporal Convolutional Neural Network</i>
WFST	<i>Weighted Fuzzy Times Series</i>

Lista de Símbolos

Δ_e	Deficit de saturação de vapor de água
Δ	Declividade da curva de pressão de vapor
γ	Constante psicrométrica
α	Albedo
σ	Constante de Stefan-Boltzmann

Sumário

1	Introdução	25
1.1	Motivação	26
1.2	Objetivos	27
1.3	Publicações	28
1.4	Organização da dissertação	28
2	Evapotranspiração	31
2.1	Fatores que afetam a evapotranspiração	31
2.1.1	Parâmetros climáticos	31
2.1.1.1	Radiação solar	32
2.1.1.2	Umidade do ar	32
2.1.1.3	Temperatura do ar	33
2.1.1.4	Velocidade do vento	33
2.1.2	Características da cultura	34
2.1.3	Manejo e os aspectos ambientais	34
2.2	Evapotranspiração de referência	35
2.3	Série temporal de evapotranspiração de referência	39
2.4	Previsão da evapotranspiração de referência	40
2.5	Resumo	41
3	Redes neurais convolucionais	43
3.1	Redes neurais convolucionais	43
3.2	Redes neurais convolucionais temporais	46
3.3	<i>Ensemble learning</i>	50
3.4	Resumo	52
4	Otimização de hiperparâmetros	53
4.1	O problema de otimização	53
4.2	O algoritmo genético	54
4.3	Resultado da otimização dos hiperparâmetros das CNNs	59
4.4	Resultado da otimização dos hiperparâmetros dos modelos de comparação	59
4.5	Resumo	62
5	Resultados	63
5.1	Comparação entre os modelos CNN, LSTM, WFTS e ARIMA	64
5.2	Comparação entre os modelos CNN e <i>Ensemble Learnings</i>	68
5.3	Comparação entre os modelos CNN e o método Naïve	76
5.4	Análise dos resíduos	80
5.5	Resumo	84
6	Conclusão	91

6.1 Trabalhos futuros	93
Referências	95

Capítulo 1

Introdução

A Agenda 2030 para o Desenvolvimento Sustentável Global ([United Nations 2015](#)), estabelecida pelas Nações Unidas (ONU) em 2015, teve como um de seus objetivos “*Acabar com a fome, alcançar a segurança alimentar e melhorar a nutrição e promover a agricultura sustentável.*” Até 2030, os países devem garantir sistemas sustentáveis de produção através de práticas agrícolas que: promovam o aumento da produtividade mantendo os ecossistemas; fortaleçam a capacidade de adaptação às mudanças climáticas e às condições meteorológicas extremas; e melhorem progressivamente a qualidade da terra e do solo. Esses desafios têm impulsionado o uso de novas tecnologias nos sistemas de produção agrícola como, por exemplo, *Big Data*, Internet das Coisas (IoT), Inteligência Artificial e Robótica.

Todas essas tecnologias fazem parte de um conceito amplo, denominado Agricultura 4.0, definida em ([Zarco-Tejada et al. 2014](#)) como uma gestão agrícola baseada na observação, medição e resposta à variabilidade de aspectos das lavouras ou da criação de animais. Uma gestão agrícola eficiente deve ser capaz de fornecer às plantas e animais exatamente o que eles precisam e quando precisam, aumentando a produtividade e reduzindo os impactos ambientais.

A Agricultura 4.0 pode beneficiar o sistema de produção agrícola de diversas formas. Na gestão das culturas, por exemplo, a previsão de produtividade pode ser feita a partir de imagens de satélite, como apresentado em ([Pantazi et al. 2016](#)). Nesse trabalho os autores concluíram que é possível prever o rendimento do trigo a partir do processamento de imagens com redes Kohonen supervisionadas. Outra atividade beneficiada é o controle de pragas e doenças, onde a pulverização de agrotóxicos passa a ser direcionada apenas para as plantas afetadas, diminuindo o custo financeiro e ambiental. Por exemplo, em ([Moshou et al. 2004](#)), redes neurais artificiais foram usadas para detectar trigo infectado por ferrugem amarela.

Na gestão da água e do solo, práticas importantes têm sido desenvolvidas não só para aumentar a produtividade das culturas, mas também para que esses recursos sejam usados de forma eficiente. Trabalhos como de ([Patil and Deka 2016](#)) fazem uso do algoritmo

máquina de aprendizado extremo para estimar a evapotranspiração semanal de uma colheita na Índia. Na gestão do solo, (Coopersmith et al. 2014) apresentou um novo método para a estimativa da umidade do solo, baseado em modelos de redes neurais artificiais. Como no caso das culturas, a Agricultura 4.0 permite que criadores de animais monitorem melhor as necessidades de cada animal. Um exemplo é o sistema de monitoramento automatizado para rastreamento do comportamento animal para detectar sinais precoces de problemas, conforme apresentado por (Matthews et al. 2017).

1.1 Motivação

A água é um dos recursos fundamentais para a agricultura. Sua disponibilidade controla as operações agrícolas em escala espacial e temporal, além de ser um dos principais limitantes de produtividade, já que tanto sua escassez quanto seu excesso afetam diretamente o desenvolvimento e a fitossanidade da cultura. Através da irrigação, a água permite a suplementação do regime de chuvas, viabilizando o cultivo em regiões com fortes restrições hídricas, como o semiárido, ou em locais com períodos específicos de seca, como a região central do Brasil.

Segundo a Organização das Nações Unidas para Alimentação e Agricultura (FAO) (Food and Agriculture Organization of the United Nations 2017), a agricultura é o principal usuário de água na maioria dos países, com a irrigação representando em média 70% das retiradas globais de água doce. Nos próximos anos esse número deve aumentar ainda mais para atender as crescentes demandas alimentares de uma população em crescimento. A FAO estima que a produção irrigada de alimentos deverá aumentar em mais de 50% até 2050 (Food and Agriculture Organization of the United Nations 2017), mas a quantidade de água retirada pela agricultura não poderá ultrapassar 10% de aumento devido às crescentes pressões sobre os recursos hídricos com urbanização, industrialização e mudanças climáticas. Isso fará com que irrigantes sejam forçados a se tornarem mais eficientes para garantir um suprimento adequado e regular de alimentos para a sociedade com menor consumo de água.

Nesse contexto, o manejo da irrigação é uma alternativa para melhorar o desempenho e a sustentabilidade dos sistemas de irrigação. O manejo da irrigação é o processo de tomada de decisão que determina a frequência de irrigação, a lâmina (ou volume) de água que deve ser aplicada em cada evento de irrigação e a intensidade dessa aplicação (Frizzone et al. 2012). Definir com precisão essas variáveis é essencial para um bom desenvolvimento da colheita e para a minimização de impactos ambientais.

Para enfrentar esses desafios, a agricultura inteligente (*smart farming*) surgiu como um novo campo científico que emprega, entre outras tecnologias, algoritmos de aprendizado de máquina para avaliar e acompanhar as condições das áreas cultivadas,

como solo e condições climáticas, levando a resultados mais precisos e tomada de decisão mais rápida (Liakos et al. 2018).

No manejo da irrigação, o aprendizado de máquina vem sendo aplicado na previsão e estimativa de evapotranspiração de referência (ET_o) diária, semanal ou mensal para definir o volume de água a ser aplicado em um evento de irrigação com base em dados climatológicos. A estimativa precisa da ET_o é essencial para o planejamento de irrigação e gerenciamento dos recursos hídricos. No entanto, prever o valor dessa variável permite melhorar o gerenciamento dos cronogramas de irrigação, bem como as operações das instalações de abastecimento de água, como canais e reservatórios. Em (Bachour et al. 2016), por exemplo, foi utilizada uma metodologia que combina a análise de multirresolução de *wavelet* com o algoritmo MVRVM (máquina de vetores de relevância multivariada), para prever 16 dias de ET_o . O MVRVM foi utilizado também por (Torres et al. 2011) para previsões diárias diretas e indiretas (onde a previsão dos parâmetros climáticos necessários para o cálculo da ET_o é feita com base em dados históricos e depois é calculada a ET_o). Já em (Landeras et al. 2009), utilizou-se uma RNA (rede neural artificial) para fazer previsões semanais de ET_o e em (Manikumari N et al. 2017), a previsão diária de ET_o foi feita usando um *ensemble* de RNAs.

Mais recentemente, uma técnica de aprendizado de máquina que vem apresentando bons resultados na resolução de problemas agrícolas é o aprendizado profundo (*deep learning* (DL)). Em (Kamilaris and Prenafeta-Boldú 2018), por exemplo, é apresentada uma revisão de 40 pesquisas relevantes na área. Segundo o autor, além das melhorias no desempenho dos problemas de classificação/previsão, também verificou-se que o DL reduziu a engenharia de recursos em muitos dos trabalhos.

Além do que já foi exposto, outra motivação para este trabalho é o fato de não ter sido encontrada nenhuma pesquisa com aplicação de DL na previsão de séries temporais de evapotranspiração de referência.

1.2 Objetivos

O principal objetivo desta pesquisa foi investigar o uso de modelos de aprendizagem profunda, em especial das redes neurais convolucionais, na previsão de séries temporais de evapotranspiração de referência. A partir desse objetivo geral, pode-se detalhar os objetivos específicos a seguir.

1. Buscar, através da pesquisa bibliográfica, conhecimento acerca do processo de evapotranspiração e de modelos de redes neurais convolucionais gerais e específicos para aplicações com dados sequenciais.

2. Implementar modelos de redes neurais convolucionais para previsão de séries temporais de evapotranspiração de referência, usando um algoritmo genético como abordagem para a otimização dos hiperparâmetros das redes.
3. Realizar experimentos para quantificar a precisão dos modelos de redes neurais convolucionais implementados e compará-los aos modelos ARIMA, WFTS, LSTM e Naïve.
4. Avaliar o desempenho da combinação de modelos de redes neurais convolucionais em modelos *ensemble* e obter previsões probabilísticas de evapotranspiração de referência.

1.3 Publicações

1. Patrícia de Oliveira e Lucas; Petrônio Cândido de Lima e Silva; Frederico Gadelha Guimarães. “*Otimização Evolutiva e Hiperparâmetros para Modelos de Séries Temporais Nebulosas*”. 14° Simpósio Brasileiro de Automação Inteligente (SBAI 2019), Ouro Preto, MG. 2019, p. 1-7. <https://proceedings.science/p/111141>
2. Petrônio Cândido de Lima e Silva; Patrícia de Oliveira e Lucas; Frederico Gadelha Guimarães. “*A Distributed Algorithm for Scalable Fuzzy Time Series*”. 14th International Conference on Green, Pervasive, and Cloud Computing. GPC 2019, Uberlândia. Lecture Notes in Computer Science. 14 ed.: Springer International Publishing, 2019, v.11484, p. 42-56.

1.4 Organização da dissertação

O restante desta dissertação está organizado da seguinte forma: o capítulo 2 apresenta o conceito de evapotranspiração de referência, fatores e variáveis climatológicas que a influenciam e a importância da previsão dessa variável no planejamento de irrigação e no gerenciamento dos recursos hídricos. Além disso, apresenta a série temporal de evapotranspiração de referência usada nos experimentos realizados, bem como as equações usadas para gerá-la.

No capítulo 3, são apresentadas as redes neurais convolucionais, suas variações para aplicação de previsão de séries temporais e uma breve descrição sobre *ensemble learning*. Já no capítulo 4, é descrita a abordagem usada para otimizar os parâmetros dessas redes e dos modelos usados para comparação.

No capítulo 5, os resultados e discussões da previsão da série temporal diária de evapotranspiração de referência para diferentes horizontes de previsão são apresentados.

É feita uma comparação entre o desempenho das redes convolucionais com os 3 tipos de estratégias de previsão e com múltiplos modelos de redes convolucionais.

Finalmente, no capítulo 6, as conclusões do trabalho e propostas para futuras pesquisas são apresentadas.

Capítulo 2

Evapotranspiração

A evapotranspiração (ET) é a variável mais ativa do ciclo hidrológico e a principal componente no balanço hídrico em ecossistemas agrícolas (Pereira et al. 2013). Durante a evapotranspiração a água passa do estado líquido para o de vapor e retorna à atmosfera por meio de dois processos físicos semelhantes: a evaporação e a transpiração. Esses processos se diferenciam unicamente quanto ao tipo de superfície evaporante. A evaporação pode ocorrer a partir do solo úmido ou da copa das plantas molhadas. Já na transpiração a água contida nos tecidos das plantas é perdida através das folhas, caule, flores e/ou raízes (Anjitha 2019).

Dessa forma, a ET , segundo (Bernardo et al. 2019), pode ser definida como “a quantidade de água evaporada e transpirada por uma superfície com vegetal, durante um determinado período do tempo”. A ET pode ser expressa como altura equivalente de água evaporada em $mm\ t^{-1}$, onde t denota uma unidade de tempo (horas, dia, mês, estações de crescimento, ou anos) (Frizzone et al. 2012).

2.1 Fatores que afetam a evapotranspiração

Os fatores que afetam a evapotranspiração são os parâmetros climáticos, as características das culturas, o manejo e os aspectos ambientais (Allen et al. 2006).

2.1.1 Parâmetros climáticos

Como definido anteriormente, a ET é um fenômeno que tem como base a passagem da água do estado líquido para o estado de vapor. Essa mudança de fases exige um suprimento de energia para a evaporação da água que, na natureza, é proveniente da radiação solar e, em menor escala, da temperatura do ar. A força motriz para liberar o vapor de água da superfície de evaporação para a atmosfera é dada pelo déficit de saturação de vapor de água. A continuidade desse processo ocorre à medida que o vento proporciona a troca do ar saturado, próximo a superfície evaporante, por ar mais seco

(Frizzone et al. 2012; Pereira et al. 2013). Dessa forma, os principais parâmetros climáticos que afetam a ET são radiação solar, temperatura do ar, umidade e velocidade do vento (Allen et al. 2006).

2.1.1.1 Radiação solar

A disponibilidade de energia por radiação solar depende do local (latitude, altitude e topografia da região) e da época do ano (Allen et al. 2006). A latitude determina o total diário de radiação solar potencialmente passivo de ser utilizado no processo de evapotranspiração. A época do ano determina o ângulo de incidência dos raios solares na região que, dependendo da topografia, poderá ter diferentes disponibilidades de energia (Pereira et al. 2013).

O gráfico da Figura 1 mostra os valores diários de radiação extraterrestre (R_a) ao longo do ano para diferentes latitudes. Percebe-se a existência de um efeito das estações do ano sobre a variação da radiação, que se intensifica à medida que a área se afasta do Equador (Bernardo et al. 2019).

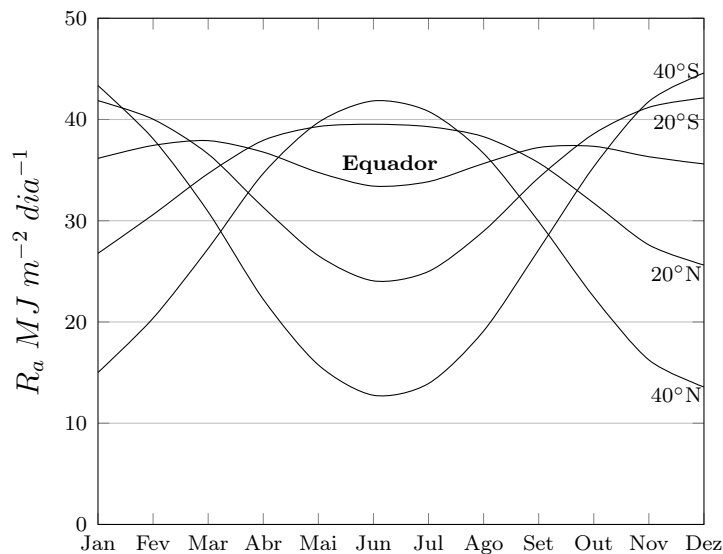


Figura 1 – Variação anual da radiação extraterrestre (R_a) no Equador, a 20° e 40° de latitude norte e sul.

Fonte: Adaptado de (Allen et al. 2006).

2.1.1.2 Umidade do ar

O déficit de saturação de vapor de água (Δ_e) representa o espaço livre na atmosfera para abrigar mais vapor de água até atingir a saturação para uma determinada temperatura. Isso indica que o aumento (decréscimo) do Δ_e implica em aumento (decréscimo) do potencial evapotranspirativo do ambiente (Pereira et al. 2013). O Δ_e é medido pela diferença entre a pressão exercida pelo ar saturado (e_s) e a pressão parcial de vapor no ar (e_a), ou seja,

$\Delta_e = e_s - e_a$ (Allen et al. 2006). O termo $e_s - e_a$ é função da umidade relativa média (UR), sendo que, quando $UR = 100\%$ tem-se $e_s = e_a$ e o Δ_e é zero (Frizzone et al. 2012).

2.1.1.3 Temperatura do ar

A temperatura do ar é influenciada pela radiação solar, de modo que os meses nos quais seus valores são mais altos são aqueles em que se recebe mais radiação (Frizzone et al. 2012). O efeito da temperatura sobre a ET provém da sua ação sobre o Δ_e . Uma vez que o aumento (diminuição) da temperatura provoca a expansão (contração) do ar, o potencial de absorção de vapor d'água (e_s) depende da temperatura (Pereira et al. 2013). Esse fenômeno é ilustrado no gráfico da Figura 2, que mostra a pressão de saturação em função da temperatura do ar.

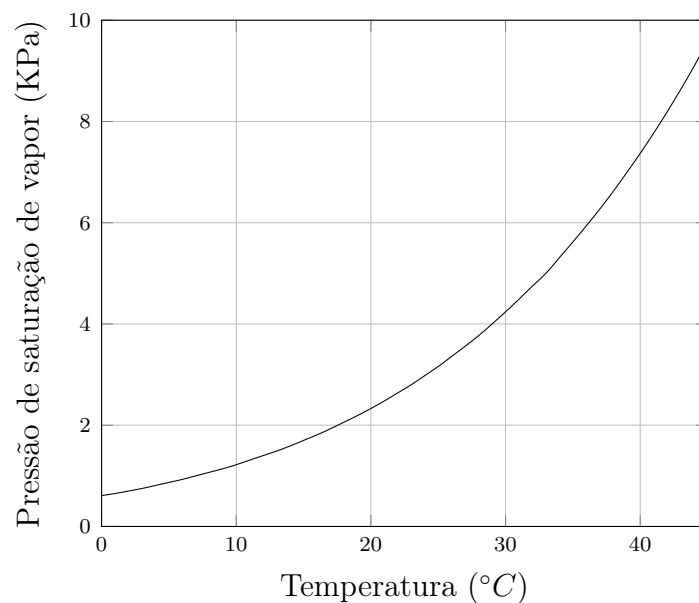


Figura 2 – Pressão de saturação de vapor em função da temperatura.

Fonte: Adaptado de (Allen et al. 2006).

2.1.1.4 Velocidade do vento

A ação do vento consiste em deslocar as parcelas de ar mais úmidas, substituindo-as por outras mais secas. Inexistindo o vento, o processo de evaporação cessaria tão logo o ar atingisse a saturação, uma vez que estaria esgotada sua capacidade de absorver vapor d'água (Frizzone et al. 2012).

A Figura 3 ilustra o efeito da velocidade do vento sobre a evapotranspiração para duas condições climáticas diferentes. É possível verificar que, em climas secos e quentes a taxa de evapotranspiração é maior que em condições climáticas úmidas, já que a disponibilidade de energia e a secura do ar aumentam o potencial de retenção de vapor de água pela atmosfera. Além disso, devido às condições climáticas, a velocidade do vento

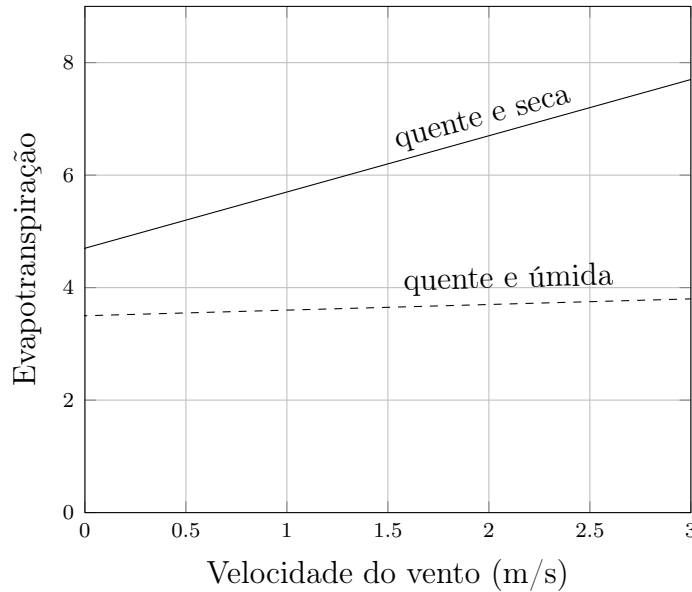


Figura 3 – Ilustração do efeito da velocidade do vento sobre a evapotranspiração em condições climáticas quente e seca e úmida.

Fonte: Adaptado de (Allen et al. 2006).

produz efeitos diferentes sobre a *ET*. Para condições úmidas, o vento só pode substituir o ar saturado por ar um pouco menos saturado e remover a energia térmica. Consequentemente, a velocidade do vento afeta a taxa de evapotranspiração em uma extensão muito menor do que em condições áridas, onde pequenas variações na velocidade do vento podem resultar em maiores variações na taxa de evapotranspiração (Allen et al. 2006).

2.1.2 Características da cultura

O tipo de cultura, variedade e estágio de desenvolvimento devem ser considerados ao avaliar a evapotranspiração. Isso porque a arquitetura das plantas, hábito de crescimento, forma, coloração e espessura das folhas resultam em diferentes níveis de *ET* (Allen et al. 2006). Tomemos a característica coloração das folhas para exemplificar: superfícies mais claras têm maior poder refletor de radiação solar (albedo) que aquelas mais escuras e, portanto, um potencial evapotranspirativo menor por ter menos energia disponível (Pereira et al. 2013).

2.1.3 Manejo e os aspectos ambientais

Em geral, todos os fatores que afetam o crescimento e o desenvolvimento das plantas afetam também a evapotranspiração. São exemplos desses fatores: salinidade e baixa fertilidade do solo, ausência de controle de doenças e pragas, manejo inadequado do solo, excesso ou o déficit de água, densidade da planta e cobertura do solo (Waller et al. 2015).

2.2 Evapotranspiração de referência

Todo sistema de irrigação tem como objetivo suprir de água o sistema solo-planta para satisfazer totalmente o déficit de evapotranspiração de uma cultura, de modo a complementar as precipitações efetivas. De acordo com (Allen et al. 2006), a quantidade de água requerida por uma cultura é calculada a partir da evapotranspiração de referência (ET_o) e do coeficiente da cultura (K_c).

A ET_o foi definida em (Allen et al. 2006) como sendo a taxa de evapotranspiração que ocorre em uma superfície de referência cujas características são: extensa superfície com vegetação em crescimento ativo, cobrindo totalmente o solo, com altura da vegetação de 0.12 m, índice de área foliar de 2.88, albedo igual a 0.23 e resistência ao transporte de vapor de água constante e igual a $70sm^{-1}$.

Os métodos para determinar a ET_o podem ser de medição direta ou indireta. Enquadram-se em métodos diretos os lisímetros e os métodos de balanço de água no solo. Como indiretos, os modelos micrometeorológicos, teóricos e empíricos, que são baseados na utilização de dados climáticos (Frizzone et al. 2012).

O método mais preciso para medir a ET_o é o lisímetro, porém esse equipamento além de ser grande e caro, é de realização demorada e de difícil execução em condições de campo (Anjitha 2019), justificando-se seu uso apenas para desenvolver e calibrar métodos de estimativas (Pereira et al. 2013).

Existem vários modelos disponíveis para estimar a ET_o por métodos indiretos, desenvolvidos e adotados em diferentes condições agro-climáticas. Dentre eles, o método Penman-Monteith foi selecionado pela FAO por sua capacidade de proporcionar resultados similares aos que se obtém pelas medições diretas em diversos tipos de clima (Frizzone et al. 2012). Assim, a evapotranspiração de referência padrão FAO (ET_{FAO}), estimada pela equação Penman-Monteith parametrizada de acordo com a superfície de referência (PM-FAO) (2.1), tornou-se referência em substituição às medidas lisimétricas (Pereira et al. 2013).

$$ET_{FAO} = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T+273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad (2.1)$$

em que:

ET_{FAO} = evapotranspiração de referência padrão, $mm d^{-1}$;

R_n = saldo de radiação na superfície da cultura, $MJ m^{-2} d^{-1}$;

G = fluxo de calor no solo, $MJ m^{-2} d^{-1}$;

T = temperatura média diária, C ;

u_2 = velocidade do vento à 2m de altura, ms^{-1} ;

e_s = pressão de valor de saturação, kPa ;

e_a = pressão de vapor, kPa ;

$e_s - e_a$ = déficit da pressão de valor de saturação, kPa ;

Δ = declividade da curva de pressão de vapor, $kPa^\circ C^{-1}$;

γ = constante psicrométrica, $kPa^\circ C^{-1}$;

A Equação 2.1 utiliza parâmetros climáticos que podem ser medidos por estações meteorológicas automáticas, são eles: temperatura do ar, umidade do ar, insolação e velocidade do vento. Os demais parâmetros podem ser estimados pelas equações descritas abaixo, de acordo com (Allen et al. 2006).

Constante Psicrométrica (γ): representa, segundo (Pereira et al. 2013), a taxa de conversão do calor sensível do ar em calor latente de evaporação. γ pode ser estimada pela Equação 2.3, onde P representa a pressão atmosférica calculada pela Equação 2.2.

$$P = 101.3 \left(\frac{293 - 0.0065Z}{293} \right)^{5.26} \quad (2.2)$$

onde:

Z = altitude [m]

$$\gamma = 0.000663P \quad [P \text{ em } KPa] \quad (2.3)$$

Fluxo de calor no solo (G): pode ser desprezado para cálculos diários, pois é pequena a sua importância para solos completamente cobertos com cultivo rasteiro, denso e em desenvolvimento ativo, como é o caso da superfície de referência.

Pressão de valor de saturação (e_s): é a pressão de vapor do ar quando, para uma temperatura T ($^\circ C$), o ar se encontra saturado (Equação 2.4).

$$e_s = \frac{0,6108 e^{\left(\frac{17,27T_{max}}{T_{max}+237,3}\right)} + 0,6108 e^{\left(\frac{17,27T_{min}}{T_{min}+237,3}\right)}}{2} \quad (2.4)$$

onde:

T_{max} = temperatura máxima [$^\circ C$]

T_{min} = temperatura mínima [$^\circ C$]

Pressão de vapor (e_a): é a pressão de saturação de vapor à temperatura do ponto de orvalho, temperatura em que o ar precisa ser resfriado a pressão constante para que se sature. e_a pode ser calculada de diferentes formas, sendo apresentadas duas delas.

A Equação 2.5 pode ser usada quando se dispõe unicamente de dados de umidade relativa média diária (UR) e a Equação 2.6 em caso de ausência de dados de UR .

$$e_a = \frac{UR}{100} e_s \quad (2.5)$$

$$e_a = 0.611 \left(\frac{17.27 T_{min}}{T_{min} + 237.3} \right) \quad (2.6)$$

Δ : é a declividade da curva de pressão de vapor d'água da atmosfera no ponto correspondente à temperatura média do ar. É estimada pela Equação 2.7, onde T é a temperatura média ($T = \frac{T_{max} + T_{min}}{2}$).

$$\Delta = \frac{4098 \left[0.6108 \frac{17.27T}{T+237.3} \right]}{(T + 237.3)^2} \quad [T \text{ em } ^\circ C] \quad (2.7)$$

Saldo de radiação na superfície da cultura (R_n): é o equilíbrio entre a energia absorvida, refletida e emitida pela superfície da Terra. A R_n pode ser estimada pela diferença entre a radiação curta líquida recebida (R_{ns}) e a radiação de saída de onda longa líquida (R_{nl}) (Equação 2.8). R_n normalmente é positivo durante o dia e negativo durante a noite. O valor total diário para R_n é quase sempre positivo ao longo de um período de 24 horas, exceto em condições extremas de altas latitudes.

$$R_n = R_{ns} - R_{nl} \quad (2.8)$$

A R_{ns} é a fração da radiação solar (R_s) que não é refletida da superfície. Seu valor é estimado pela Equação 2.9.

$$R_{ns} = (1 - \alpha) R_s \quad (2.9)$$

onde:

R_{ns} = radiação de onda curta líquida [$MJ \ m^{-2} \ dia^{-1}$]

R_s = radiação solar [$MJ \ m^{-2} \ dia^{-1}$]

α = albedo ou coeficiente de reflexão do dossel, é de 0.23 para a cultura de referência grama hipotética.

A R_{nl} é a diferença entre radiação de onda longa emitida e recebida (Equação 2.10).

$$R_{nl} = \sigma \left[\frac{T_{max,k^4} + T_{min,k^4}}{2} \right] (0.34 - 0.14 \sqrt{e_a}) \left(1.35 \frac{R_s}{R_{so}} - 0.35 \right) \quad (2.10)$$

onde:

R_{nl} = radiação de onda longa líquida [$MJ m^{-2} dia^{-1}$]

σ = constante de Stefan-Boltzmann [$4.903 \cdot 10^{-9} MJ K^{-4} m^{-2} dia^{-1}$]

$T_{max,k^{-4}}$ = Temperatura absoluta máxima durante 24h em Kelvin

$T_{min,k^{-4}}$ = Temperatura absoluta mínima durante 24h em Kelvin

$\frac{R_s}{R_{so}}$ = Radiação relativa de onda curta (deve ser ≤ 1.0)

R_s é a radiação solar que atinge a superfície da Terra num dado período. Como o sol emite energia por meio de ondas eletromagnéticas caracterizadas por comprimentos de onda curtos, a radiação solar também é conhecida como radiação de ondas curtas. A R_s pode ser estimada através da Equação 2.11, quando o valor da insolação é conhecido, ou através da Equação 2.12 quando esse valor é desconhecido.

$$R_s = \left(a_s + b_s \frac{n}{N} \right) R_a \quad (2.11)$$

onde:

n = insolação (duração real da luz do dia) [h]

N = duração máxima de insolação no dia [h]

R_a = radiação extraterrestre [$MJ m^{-2} dia^{-1}$]

a_s e b_s são 0.25 e 0.5 respectivamente. Esses valores são recomendados quando a radiação solar não está disponível.

$$R_s = k_{R_s} \sqrt{T_{max} - T_{min}} R_a \quad (2.12)$$

onde:

k_{R_s} = coeficiente de ajuste (0,16 .. 0,19) [$^{\circ}C^{-0,5}$]. Para localidades interiores, onde a massa de terra domina, e as massas de ar não são influenciadas por um corpo d'água grande, o $k_{R_s} \approx 0,16$.

A Radiação extraterrestre (R_a) é a radiação solar recebida no topo da atmosfera terrestre em uma superfície horizontal. Para períodos diários R_a pode ser estimada pela Equação 2.13.

$$R_a = \frac{24.6}{\pi} G_{sc} d_r [\omega_s \text{sen}(\phi) \text{sen}(\delta) + \cos(\phi) \cos(\delta) \text{sen}(\omega_s)] \quad (2.13)$$

onde:

R_a = radiação extraterrestre para períodos diários [$MJ m^{-2} dia^{-1}$]

G_{sc} = constante solar [$MJ m^{-2} min^{-1}$]

d_r = inverso da distância relativa da terra-sol

δ = declinação do sol [rad]

ω_s = ângulo horário do pôr-do-sol [rad]

ϕ = latitude [rad]

A Radiação Solar de Céu Claro (R_{so}) é a radiação solar que alcançaria a mesma superfície durante o mesmo período, mas sob condições sem nuvens. Para valores não calibrados de a_s e b_s a R_{so} pode ser estimada pela Equação 2.14.

$$R_{so} = (0.75 + 2 \cdot 10^{-5} Z) R_a \quad (2.14)$$

onde:

R_{so} = radiação solar de céu claro [$MJ m^{-2} dia^{-1}$]

Z = altitude acima do nível do mar [m]

R_a = radiação extraterrestre [$MJ m^{-2} dia^{-1}$]

2.3 Série temporal de evapotranspiração de referência

Os dados climáticos utilizados para gerar a série temporal de ET_o são originários da estação número 83586 do INMET - Instituto Nacional de Meteorologia, localizada no município de Sete Lagoas, estado de Minas Gerais.

A série histórica registrada nesta estação compreende o período de 1993 a 2015, constituindo 22 anos de dados diários, referentes às seguintes variáveis meteorológicas:

T_{max} = temperatura máxima [$^{\circ}C$];

T_{min} = temperatura mínima [$^{\circ}C$];

$U2$ = velocidade do vento a 2m de altura [ms^{-1}];

RH = umidade relativa média [%];

I = insolação [h].

A evapotranspiração de referência foi calculada para cada um dos dias da série histórica utilizando a Equação 2.1, gerando a série temporal de evapotranspiração de referência (Figura 4). A base de dados e o código para geração da série estão disponíveis em <https://github.com/PatriciaLucas/Evapotranspiracao>. A Tabela 1 apresenta as estatísticas e a Figura 5 mostra os gráficos da função de autocorrelação e da função de autocorrelação parcial da série.

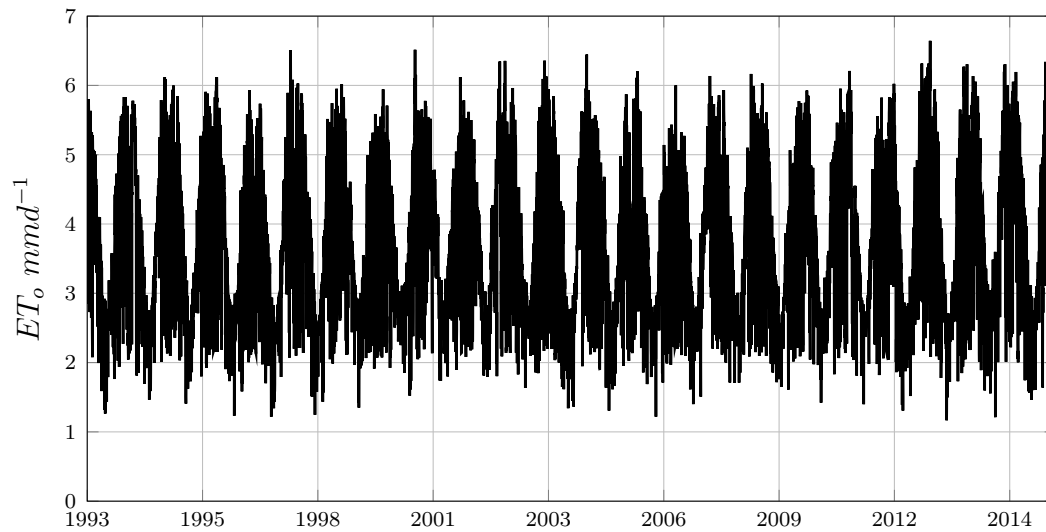


Figura 4 – Série de Evapotranspiração de Referência da estação meteorológica de Sete Lagoas - MG.

Estatísticas	ET_0
Número de amostras	8.400
Média	3.65
Desvio padrão	1.11
Valor mínimo	1.16
25%	2.70
50%	3.47
75%	4.53
Valor máximo	6.66

Tabela 1 – Estatísticas da série temporal de Evapotranspiração de Referência.

2.4 Previsão da evapotranspiração de referência

A estimativa precisa da ET_0 é essencial para o planejamento de irrigação e gerenciamento dos recursos hídricos. No entanto, prever o valor dessa variável permite conhecer antecipadamente as necessidades de água das culturas, auxiliando agricultores na tomada de decisão de atividades como gerenciamento dos sistemas de abastecimento de água em áreas onde a produção agrícola é o principal usuário da água (Mariño et al. 1993), redução do custo da energia elétrica, visto que é possível planejar a irrigação em horários com tarifas mais baratas (Perea et al. 2015), identificação de possível estresse futuro das culturas (Zhao et al. 2019) e operação de canais e reservatórios, sendo possível a previsão da quantidade de água que deve ser entregue (Bachour et al. 2016). Previsões de ET_0 são especialmente importantes para sistemas de irrigação de alta frequência e manejo da irrigação em tempo real (Yang et al. 2016) (Xiong et al. 2016).

Em geral, a tomada de decisão dessas atividades depende de previsões de 1 a 10 passos a frente. Horizontes de previsão mais curtos, de até 7 dias, por exemplo, podem ser

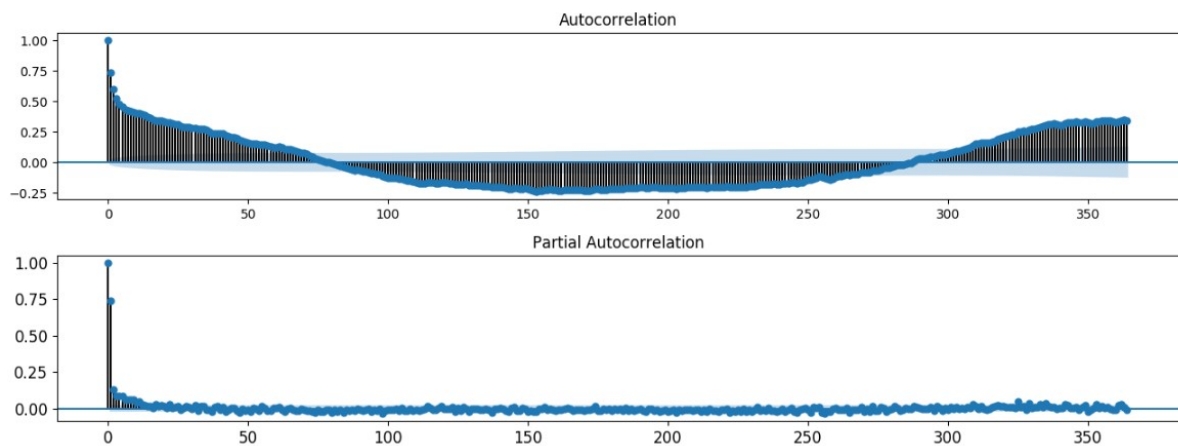


Figura 5 – Gráfico das funções de autocorrelação e autocorrelação parcial da série temporal de Evapotranspiração de Referência.

utilizados em planejamento do uso de sistemas de irrigação na propriedade e na otimização de energia elétrica. Já horizontes de previsão mais longos, de 7 à 10 dias, auxiliam na gestão hídrica de canais de irrigação de perímetros irrigados e em reservatórios para uso de irrigação.

2.5 Resumo

Este capítulo descreveu o conceito de evapotranspiração, bem como as variáveis climatológicas e outros fatores que influenciam nesse processo. Descreveu-se também a evapotranspiração de referência e a importância de sua previsão para o planejamento de irrigação e gerenciamento dos recursos hídricos. Dentre as abordagens pesquisadas até o momento, não foi encontrada uma pesquisa que utilize o aprendizado profundo para previsão de séries temporais de ET_o . Dessa forma, o próximo capítulo tem como objetivo descrever as redes convolucionais, tendo como foco o uso dessas redes para previsão de séries temporais.

Capítulo 3

Redes neurais convolucionais

3.1 Redes neurais convolucionais

Nas últimas décadas, pesquisadores têm se esforçado no sentido de permitir que, através de dados brutos, computadores consigam extrair informações suficientes para modelar o mundo real. Para conseguir isso, muitos recorreram a algoritmos de aprendizado de máquina para capturar uma grande quantidade de informações e descobrir automaticamente as representações necessárias para detectar ou classificar padrões de entrada (Bengio 2009; Guo et al. 2016).

O aprendizado profundo (*Deep Learning*) é, segundo (Lecun et al. 2015), um subcampo do aprendizado de máquina que tenta aprender abstrações através de múltiplos níveis de representação, obtidos pela composição de módulos não lineares simples. Com múltiplas camadas não lineares, um sistema pode implementar funções extremamente complexas de suas entradas, sensíveis a detalhes minuciosos e insensíveis a grandes variações, como o fundo e a iluminação em uma imagem.

Uma rede neural convolucional é um método de aprendizagem profunda inspirada no mecanismo de percepção visual natural dos seres vivos. Proposta inicialmente por (LeCun et al. 1990) e depois melhorada em (Lecun et al. 1998), a LeNet-5, é uma rede neural multicamadas para classificação de dígitos manuscritos a partir de pixels brutos, com pouco ou nenhum pré-processamento.

No entanto, devido à falta de grandes quantidades de dados de treinamento e poder de computação da época, as redes não apresentavam bom desempenho em problemas mais complexos, como por exemplo, classificação de imagem e vídeo em larga escala.

Nos últimos anos o aumento da capacidade de processamento de chips, custo reduzido de hardware para computação, conjuntos de dados maiores e principalmente algoritmos e redes mais aprimoradas permitiram que isso mudasse, tornando o aprendizado profundo uma abordagem emergente e amplamente aplicada (Guo et al. 2016; Szegedy et al. 2015).

Desde então, redes convolucionais cada vez mais profundas estão sendo propostas, superando resultados em muitos campos como: visão computacional (Russakovsky et al. 2015; Simonyan and Zisserman 2015; Szegedy et al. 2015; He et al. 2016), detecção de objetos em imagens (Girshick et al. 2014; Ren et al. 2016), reconhecimento da fala (Ossama et al. 2014), processamento de sinais de áudio (Qu et al. 2016; Kumar and Raj 2017), entre outros.

Tomando como base a arquitetura da LeNet-5 (Figura 6) (Lecun et al. 1998), uma rede neural convolucional possui 3 tipos de camadas, a convolucional, a de agrupamento (ou Pooling) e as camadas totalmente conectadas.

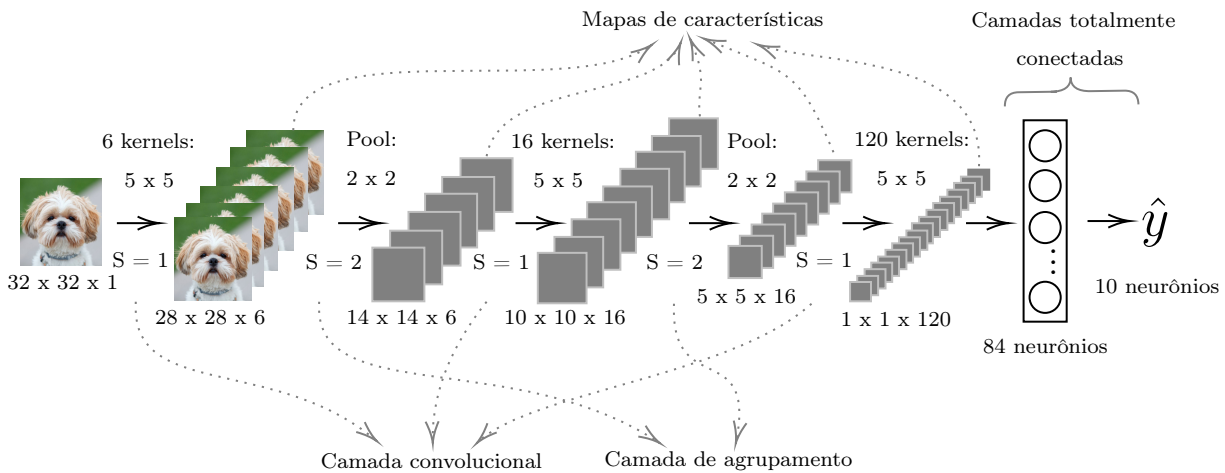


Figura 6 – Arquitetura da LeNet-5.

A camada convolucional tem como objetivo extrair as características mais importantes da entrada. Dessa forma, sua saída é um mapa de características obtido a partir da convolução da entrada com um kernel aprendido, seguido da aplicação de uma função de ativação não linear (Gu et al. 2018). Os mapas de características completos são obtidos pela Equação 3.1:

$$Z_{i,j,k}^L = W_k^L \cdot X_{i,j}^L + b_k^L \quad (3.1)$$

onde:

$Z_{i,j,k}^L$ é o mapa de características obtido pela convolução do k-ésimo filtro da L-ésima camada com a célula de entrada centrada na localização (i,j).

W_k^L vetor de pesos do k-ésimo filtro da L-ésima camada.

b_k^L termo de polarização do k-ésimo filtro da L-ésima camada.

$X_{i,j}^L$ é a célula de entrada centrada na localização (i,j) da L-ésima camada.

A profundidade dos mapas de características é dada pelo número de kernels (ou filtros) de convolução. Observe na Figura 6 que a 1ª camada de convolução com 6 kernels

gera uma saída de profundidade 6. Isso porque, cada kernel possui pesos diferentes para extrair diferentes características da entrada.

Uma vantagem das camadas de convolução é o compartilhamento do vetor de pesos para toda a circunvolução na construção de um mapa de características, pois reduz o número de parâmetros na rede, resultando em treinamento e previsões mais eficientes (Borovykh et al. 2017).

A largura e a altura desses mapas são definidas pelo tamanho do kernel e do *stride* (passo da circunvolução) (Equação 3.2). Voltando à Figura 6, a 1ª camada convolucional gera uma saída 28×28 , pois $(32 - 5)/1 + 1 = 28$.

$$T_{map} = (I - F)/S + 1 \quad (3.2)$$

onde:

T_{map} é a altura ou largura do mapa de características;

I é a altura ou largura da entrada;

F é a altura ou largura do kernel de convolução;

S é o tamanho do *stride*.

As funções de ativação aplicadas nos mapas de características (Equação 3.3) são usadas para detectar recursos não lineares. Normalmente são usadas as funções sigmóide, tangente hiperbólica ou Unidade Linear Retificada (Relu) (Gu et al. 2018).

$$a_{i,j,k}^L = a(Z_{i,j,k}^L) \quad (3.3)$$

em que a é o mapa de característica após a aplicação da função de ativação.

A camada de agrupamento (Equação 3.4) é colocada geralmente entre duas camadas de convolução para reduzir progressivamente o tamanho espacial da representação, reduzindo também a quantidade de parâmetros na rede e, portanto, controlar o *overfitting* (Krizhevsky et al. 2012). As operações típicas de agrupamento são o máximo e o valor médio (Gu et al. 2018).

$$Y_{i,j,k}^L = Pool(a_{m,n,k}^L), \forall (m, n) \in R_{i,j} \quad (3.4)$$

onde:

Y é a saída da camada de agrupamento;

$R_{i,j}$ são os vizinhos em torno da localização (i, j) .

Para ilustrar a operação de agrupamento, observe a Figura 7. Nela a matriz 2×2 é a saída de uma operação de agrupamento máximo com kernel 2×2 na matriz 4×4 . Isso

porque, para cada janela 2×2 (tamanho do kernel) na matriz de entrada, existe um valor na matriz de saída que corresponde ao valor máximo da janela. Para a 1ª janela (com o fundo de bolinhas), seu valor correspondente é 6, que é o valor máximo da janela. Como no exemplo o *stride* é 2 e sem sobreposição, foram formadas 4 janelas e conseqüentemente uma matriz de saída 2×2 .

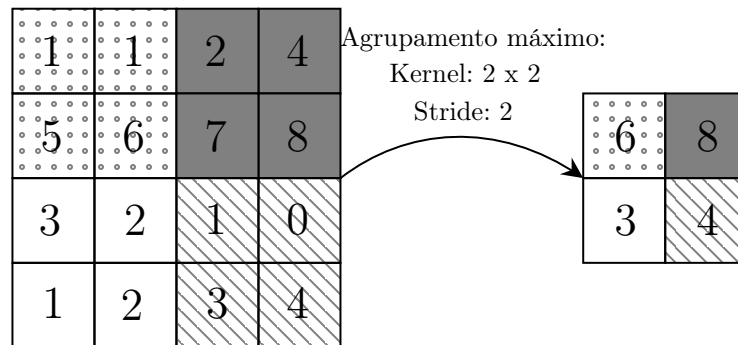


Figura 7 – Operação de agrupamento.

A largura e a altura do mapa de características de saída da camada de agrupamento é dado pela Equação 3.5. A profundidade permanecerá a mesma da entrada.

$$l_2 = (l_1 - F)/S + 1 \quad (3.5)$$

onde:

l_2 é a largura ou altura do mapa de características de saída;

l_1 é a largura ou altura do mapa de características de entrada.

Podemos observar que, na Figura 6, a 2ª camada da LeNet-5 possui um kernel 2×2 e *stride* igual a 2. Dessa forma, o mapa de característica de saída tem a mesma profundidade da entrada e largura e altura iguais a $(28 - 2)/2 + 1 = 14$.

E por último, a camada totalmente conectada, que é uma rede perceptron multicamadas. Sua função é usar os recursos de alto nível das saídas das camadas convolucionais e de agrupamento para classificar a entrada em várias classes com base no conjunto de dados de treinamento (Krizhevsky et al. 2012).

3.2 Redes neurais convolucionais temporais

A previsão de séries temporais é um problema de modelagem de sequências que busca prever como uma sequência evoluirá com o tempo com base em dados passados (Bai et al. 2018a).

Diversas abordagens para previsão de séries temporais foram propostas, no entanto, como esses dados são normalmente não lineares, não estacionários e possuem alta

dimensionalidade, essas abordagens não conseguem extrair recursos suficientes para obter resultados precisos (Shen et al. 2018). Ao adicionar mais informações para melhorar as previsões, torna-se necessário o uso de modelos mais poderosos, capazes de lidar com dados tão complexos e de alta dimensionalidade (Längkvist et al. 2014).

Segundo (Längkvist et al. 2014), métodos de aprendizagem profunda se encaixam nessa descrição, oferecendo melhor representação e classificação em vários problemas de séries temporais em comparação com abordagens rasas quando configurados e treinados adequadamente.

A ideia de usar métodos de aprendizagem profunda na previsão de séries temporais remonta ao início das arquiteturas convolucionais, como mostrado em (LeCun and Bengio 1998). No entanto, praticantes do aprendizado profundo geralmente consideram arquiteturas recorrentes como o ponto de partida padrão para tarefas de modelagem de sequências (Bai et al. 2018b). Essas arquiteturas e suas variantes têm tido muito sucesso em um vasto número de domínios que dependem de dados sequenciais. Entretanto trabalhos recentes, como (Du et al. 2017), (Lea et al. 2016), (Bai et al. 2018a), (Nair et al. 2018), (Radwan and Burgard 2018) e (Bai et al. 2018b), mostram que redes convolucionais podem apresentar desempenho competitivo sem depender da recorrência e, portanto, com menor custo computacional.

Desde a LeNet-5, muitas melhorias foram feitas nas arquiteturas de redes convolucionais para melhorar sua capacidade de processar dados temporais, o que fez com que uma nova classe de modelos temporais surgisse, as chamadas Redes Convolucionais Temporais (TCN). TCNs como WaveNet (Van Den Oord et al. 2016), ByteNet (Kalchbrenner et al. 2017) e ConvS2S (Gehring et al. 2017) apresentaram resultados promissores em modelagem de dados de sequência como áudio de música, reconhecimento de fala e tradução automática.

As TCNs, segundo (Bai et al. 2018a) possuem as seguintes características:

1. As circunvoluções na arquitetura são causais, ou seja, não há vazamento de informações do futuro para o passado.
2. A rede produz uma saída com o mesmo comprimento da entrada, como nas redes recorrentes.

Para descrever a TCN, foi escolhida uma arquitetura publicada em (Bai et al. 2018b), ilustrada na Figura 8. Nessa arquitetura os blocos residuais são inseridos no lugar de camadas convolucionais. Sendo assim, cada camada é um bloco residual formado por duas camadas convolucionais causais dilatadas, com fator de dilatação d . Para ambas as camadas usa-se a função de ativação ReLU, normalização nos pesos dos filtros e *Dropout* para regularização. Para considerar as larguras de entrada e saída discrepantes, é usada

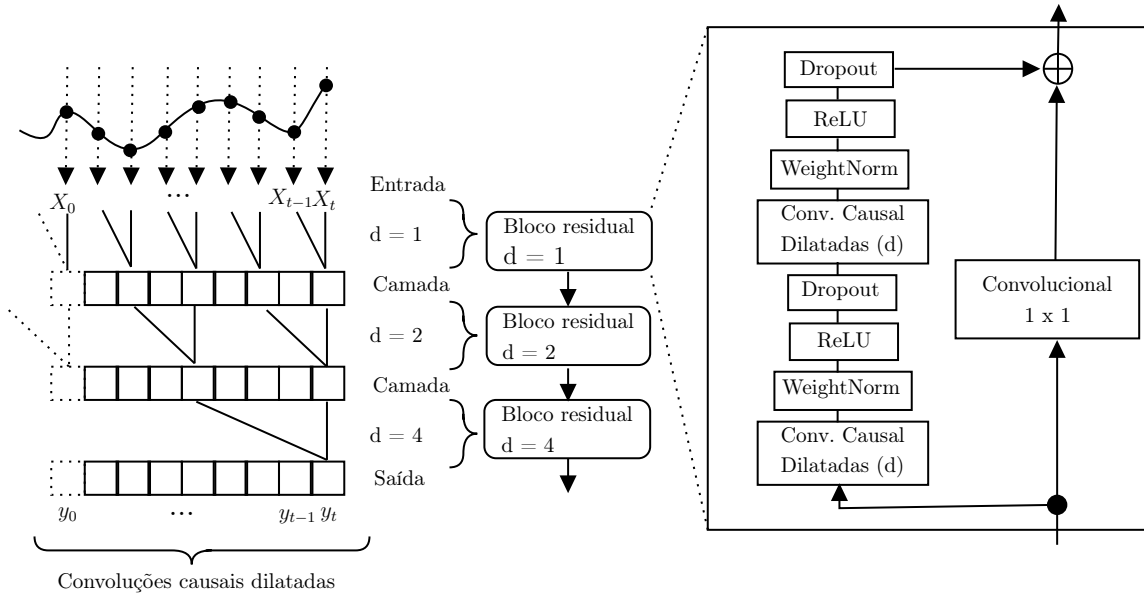


Figura 8 – Arquitetura TCN.

uma convolução adicional de 1×1 para garantir que a adição elementar receba os tensores da mesma forma. Abaixo serão descritos os elementos que compõe essa arquitetura.

Camadas convolucionais causais

Uma convolução causal não permite que o modelo viole a ordenação na qual os dados foram modelados. Dessa forma, a previsão $p(X_{t+1}|X_1 \dots X_t)$ dada pelo modelo no intervalo de tempo t não pode depender de nenhuma das futuras previsões $X_{t+1}, X_{t+2} \dots X_T$. Percebe-se aqui uma vantagem da TCN em relação às redes recorrentes no tempo de processamento, pois durante o treinamento as previsões condicionais para todos os intervalos de tempo podem ser feitas em paralelo, já que todo X já é conhecido. (Van Den Oord et al. 2016).

Uma ilustração da convolução causal é mostrada na Figura 9. Podemos observar que para um campo receptivo de tamanho 5 e kernel de tamanho 2 (que considera apenas a relação entre dois pontos de dados), precisamos de 4 camadas para fazer uma previsão. Se quiséssemos capturar todas as 16 propriedades, seriam necessárias 15 camadas. Ou seja, para tarefas que precisam de um campo receptivo muito grande, precisaríamos de uma rede extremamente profunda. Para resolver esse problema, (Van Den Oord et al. 2016) usou convoluções dilatadas propostas em (Yu and Koltun 2016) que permitem um campo receptivo exponencialmente grande.

Camadas convolucionais causais dilatadas

Uma convolução dilatada é uma convolução onde o filtro é aplicado sobre uma área maior que seu comprimento, pulando valores de entrada com uma certa etapa (Van Den Oord et al. 2016), permitindo que a rede cubra mais informações com o aumento

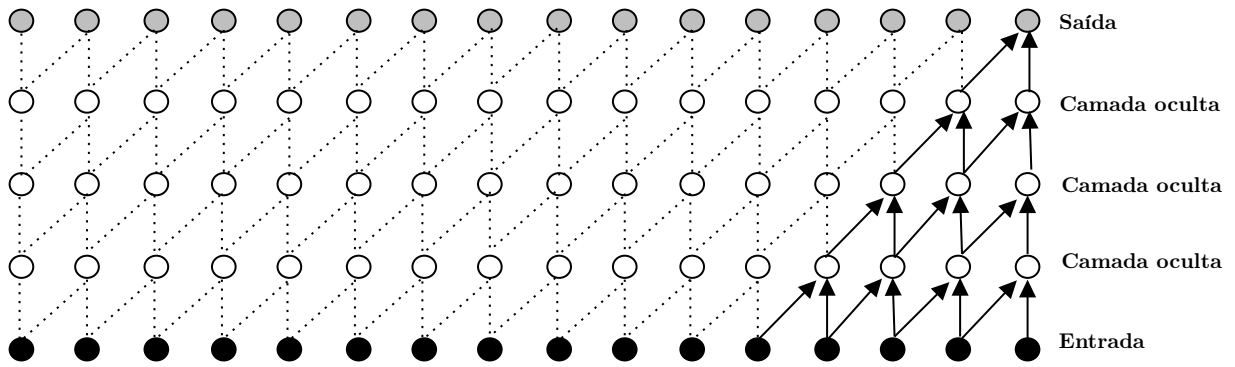


Figura 9 – Camada convolucional causal.

Adaptado de: (Van Den Oord et al. 2016)

do tamanho do campo receptivo (Gu et al. 2018). O fator de dilatação é controlado por um hiperparâmetro adicional d , de forma que $d = 1$ produz uma convolução padrão. Na Figura 10 são ilustradas convoluções causais dilatadas d iguais a 1, 2, 4 e 8. Com o mesmo número de camadas e tamanho do kernel da Figura 9, as dilatações permitem um campo receptivo de tamanho 16.

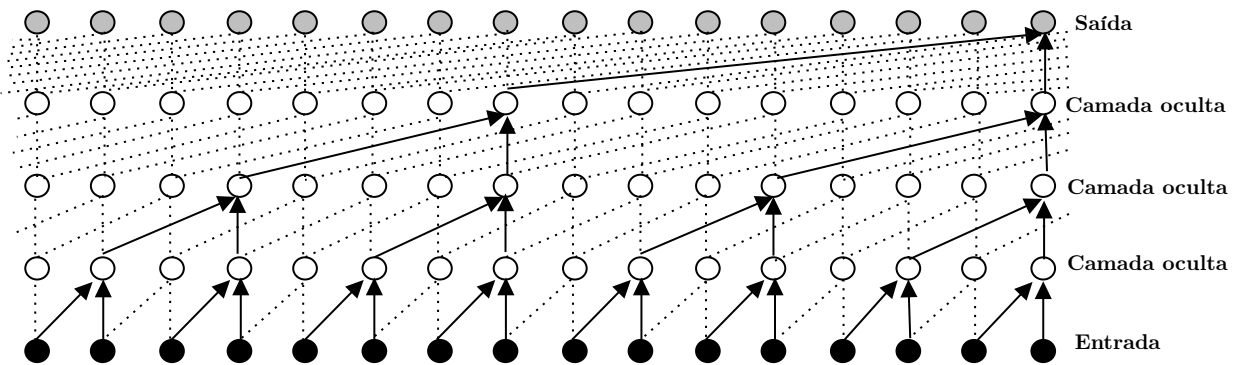


Figura 10 – Camada convolucional causal dilatada.

Adaptado de: (Van Den Oord et al. 2016)

Segundo (Van Den Oord et al. 2016), empilhar convoluções dilatadas aumentando exponencialmente o fator de dilatação resulta em crescimento exponencial do campo receptivo com apenas algumas camadas, preservando a resolução de entrada em toda a rede, bem como a eficiência computacional.

Blocos residuais

Os blocos residuais, propostos por (He et al. 2016), são compostos por uma série de transformações F , cujas saídas são adicionadas à entrada X do bloco (ver Figura 8). Em (Bai et al. 2018b), o uso da função residual estabilizou o treinamento com convergência mais rápida e melhores resultados.

Os blocos residuais da TCN (Figura 8), como dito anteriormente, fazem normali-

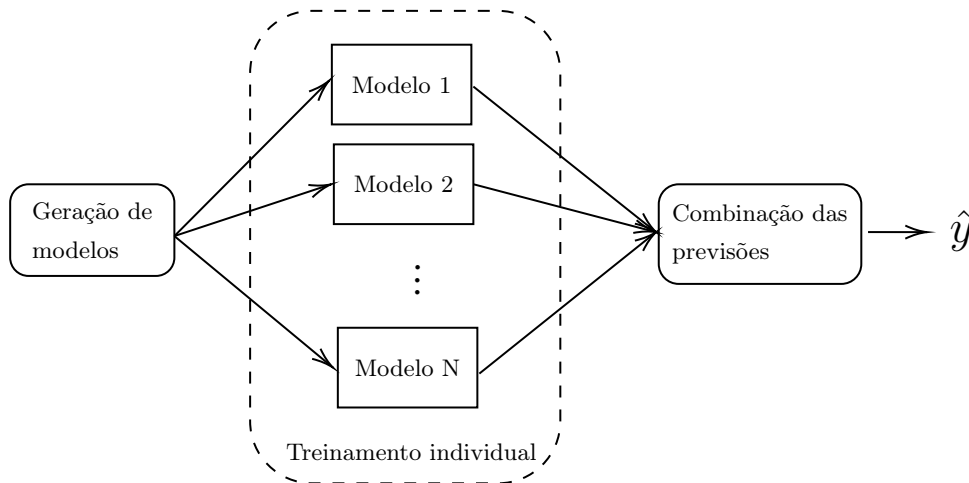


Figura 11 – Etapas do ensemble learning

zação nos pesos dos filtros e regularização através do *Dropout*. A normalização em redes profundas é importante, pois à medida que os dados fluem através da rede, a distribuição da entrada para as camadas internas será alterada, fazendo com que a rede perca sua capacidade de aprendizado e precisão (Gu et al. 2018).

Já o *Dropout* é usado para evitar *overfitting* (quando o modelo se ajusta muito bem aos dados de treinamento, mas se mostra ineficaz para prever novos resultados). A TCN usa SpatialDropout, proposto por (Tompson et al. 2015), que estende o valor de *dropout* em todo o mapa de características.

Campo receptivo

O campo receptivo em um TCN é a quantidade de *lags* (atrasos) que a rede consegue ver para prever sua saída. O tamanho do campo receptivo (r) depende do tamanho do kernel, do fator de dilatação e do número de pilhas de blocos residuais. Para um tamanho de kernel (k) e dilatações exponencialmente crescentes (d), um TCN com n pilhas de blocos residuais terá um campo receptivo $r = nkd$ (Bai et al. 2018b).

3.3 Ensemble learning

Os métodos de aprendizagem em conjunto (*ensemble learning*) treinam vários modelos individualmente e combinam suas saídas para fazer previsões (Opitz and Maclin 1999). Segundo (Hansen and Salamon 1990), as previsões combinadas são frequentemente mais precisas do que as previsões únicas feitas pelo melhor modelo do conjunto. Além disso, o efeito de suavização causado pela combinação das saídas reduz a variância dos métodos em conjunto (Polikar 2012).

Ensembles são construídos em três etapas: geração, treinamento e combinação dos modelos preditores (Polikar 2012), como ilustra a Figura 11.

A geração dos preditores deve ser feita de forma que permita encontrar modelos diversificados. Ao mostrar que ensembles compostos de preditores idênticos não melhoraram o desempenho, (Tumer and Ghosh 1995) evidencia que a diversidade é uma questão fundamental do aprendizado em conjunto. A diversidade pode ser alcançada usando conjuntos de treinamento diferentes, ou métodos iguais com diferentes hiperparâmetros ou métodos diferentes para cada modelo (Polikar 2012). Outro ponto importante colocado por (Zhi-Hua 2012) é que devem ser gerados modelos fortes e fracos, uma vez que a complementariedade é mais importante que a pura precisão.

A combinação dos modelos pode ser feita de diversas formas. Por exemplo, (Allende and Valle 2017) cita 4 métodos, são eles: média simples, média ponderada, média simples com remoção das maiores e menores previsões individuais e média simples com substituição das maiores e menores previsões individuais pelas previsões mais próximas a elas. No método média ponderada é usual calcular os pesos de acordo com alguma abordagem baseada em erros (RMSE, MAPE, MSE, etc.), como em (Adhikari and Verma 2016).

Previsões probabilísticas também podem ser usadas para combinar os modelos do *ensemble*, como em (Cândido De Lima E Silva et al. 2017) e (Xie and Hong 2016). Em contraste com as previsões de valor único apresentadas anteriormente, a previsão probabilística representa uma função de densidade de probabilidade, possibilitando quantificar a incerteza em uma previsão.

Uma abordagem bastante conhecida para estimar a função de densidade de probabilidade subjacente de um conjunto de dados é o KDE (*Kernel density estimation*). A Figura 12 ilustra como o KDE funciona a partir de seis observações, indicadas pelas linhas pretas. Primeiro é centralizado um kernel em cada um dos n pontos, onde um kernel é uma distribuição de probabilidade (curva vermelha). Em seguida, soma-se cada um desses kernels individuais para obter a estimativa de densidade $\hat{p}(x)$ (curva azul), como expresso na Equação 3.6 (Chen 2017).

$$\hat{p}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (3.6)$$

Dois parâmetros devem ser definidos nesse processo, a função kernel (K) e a largura de banda h . Segundo (Chen 2017), o efeito de K sobre o erro de estimativa é muito pequeno entre as funções comuns do kernel (gaussiana, esférica, etc.), o contrário ocorre com o parâmetro h com seu efeito sobre a suavização da curva resultante, como ilustra a Figura 13. Quando h é muito pequeno, há muitas oscilações na estimativa de densidade. Quando h é muito grande, suavizamos recursos importantes. Quando h está na quantidade correta, podemos ver uma imagem clara da densidade subjacente.

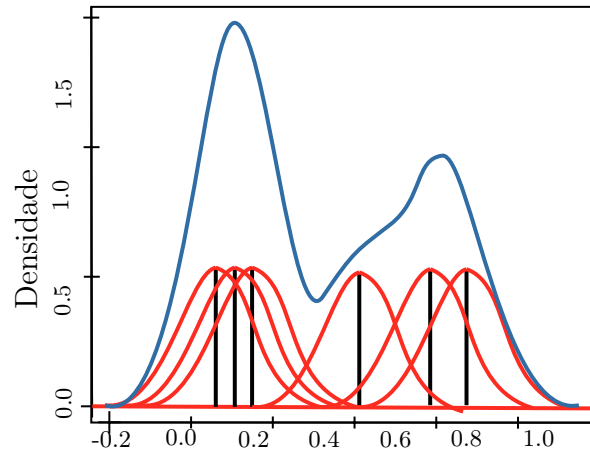


Figura 12 – Construção do KDE.

Adaptado de: (Chen 2017)

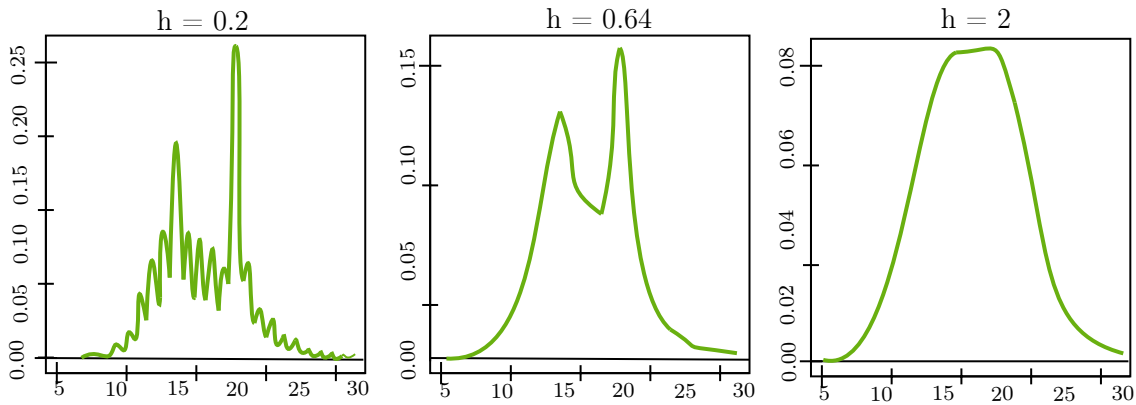


Figura 13 – Efeito da largura de banda (h) sobre a suavização da curva resultante do KDE.

Adaptado de: (Chen 2017)

3.4 Resumo

Este capítulo apresentou as redes convolucionais e suas variações para aplicação em problemas de modelagem de sequências. Essas redes são altamente sensíveis à arquitetura, sendo a escolha dos hiperparâmetros uma tarefa crucial para seu desempenho. Portanto, o próximo capítulo descreve um algoritmo genético para otimização dos hiperparâmetros das redes CNN e TCN na aplicação de previsão de séries temporais de evapotranspiração de referência.

Capítulo 4

Otimização de hiperparâmetros

Todos os sistemas de aprendizado de máquina possuem hiperparâmetros. Essas variáveis são usadas para configurar vários aspectos desses sistemas, devendo assim, serem definidas antes do treinamento dos modelos. A definição dos hiperparâmetros deve ser feita de forma que um nível apropriado de complexidade do modelo seja encontrado, uma vez que um modelo complexo exhibe grande variação (*overfitting*), enquanto um modelo excessivamente simples é fortemente tendencioso (*underfitting*).

As redes neurais convolucionais exigem um conjunto de hiperparâmetros que vão desde a estrutura até a regularização e otimização. A escolha desses hiperparâmetros tem grande impacto no modelo final, sendo sua otimização uma tarefa crucial para precisão e desempenho da rede.

Este capítulo tem como principal objetivo apresentar a otimização dos hiperparâmetros das redes CNN para a previsão da série temporal de ET_o utilizando um algoritmo genético, bem como seus resultados. Para isso, foram escolhidas três CNNs que, por dependerem de hiperparâmetros específicos, foram tratadas de forma diferente, são elas: CNN1 – redes neurais convolucionais que utilizam os elementos descritos na Seção 3.1, CNN2 – redes neurais convolucionais que utilizam os elementos descritos na Seção 3.2 sem o uso de blocos residuais e CNN3 – redes neurais convolucionais que utilizam os elementos descritos na Seção 3.2 com o uso de blocos residuais.

Além disso, a última seção deste capítulo traz o resultado da otimização dos hiperparâmetros dos modelos escolhidos para comparar com os resultados das CNNs na previsão de ET_o , são eles: LSTM (*Long short term memory networks*), WFTS (*Weighted fuzzy time series*) e ARIMA (*Autoregressive integrated moving average*).

4.1 O problema de otimização

O problema de otimização de hiperparâmetros para as CNNs é apresentado abaixo. O problema foi formulado com a intenção de encontrar um equilíbrio entre dois objetivos

conflitantes, a precisão e a parcimônia (a complexidade estrutural geralmente medida pelo número de parâmetros do modelo). A função objetivo f_1 (4.1) controla a precisão, aqui representada pela média do RMSE (*Root mean square error*) (4.3) das N avaliações do modelo e a função objetivo f_2 (4.2) controla a parcimônia, representada pela norma L_0 do vetor w de parâmetros da rede.

Otimizar:

$$\text{minimize } f_1 = N^{-1} \sum_{i=0}^N RMSE(i) \quad (4.1)$$

$$\text{minimize } f_2 = \|w\|_0 \quad (4.2)$$

Onde:

$$RMSE = \sqrt{\sum_{t=0}^n (y(t) - \hat{y}(t))^2} \quad (4.3)$$

4.2 O algoritmo genético

O algoritmo genético (AG) foi escolhido como abordagem para otimizar os hiperparâmetros das CNNs. Apesar desse algoritmo resolver originalmente problemas mono-objetivo, foi feita uma adaptação no operador de seleção para tratar o problema multi-objetivo descrito na seção anterior. Seguindo o fluxograma da Figura 14, serão descritas todas as etapas do desenvolvimento do AG de forma genérica, incluindo especificidades no tratamento dos hiperparâmetros das redes CNN1, CNN2 e CNN3 quando necessário.

1. Geração da população inicial

O algoritmo inicia com a escolha da rede, para definir quais hiperparâmetros serão otimizados. Ao escolher a rede, será gerada a população inicial com 10 indivíduos representados por um vetor (o genótipo) contendo todos os hiperparâmetros da rede escolhida. Para a otimização dos hiperparâmetros da CNN1, um indivíduo é composto pelos hiperparâmetros da Tabela 2 e da Tabela 3, para a CNN2 pelos hiperparâmetros da Tabela 2 e da Tabela 4 e para CNN3 pelos hiperparâmetros da Tabela 2 e da Tabela 5. As tabelas 6 e 7 apresentam os hiperparâmetros fixos utilizados para os 3 modelos CNN.

A população inicial é gerada aleatoriamente dentro dos limites do espaço de busca de cada hiperparâmetro. Os espaços de busca estão definidos nas Tabelas 8, 9 e 10 para a CNN1, CNN2 e CNN3, respectivamente. As redes CNN2 e CNN3 apresentam uma estrutura que limita seu campo receptivo em função do número de camadas convolucionais causais dilatadas (C_d), pelo tamanho do filtro (F) e, no caso da CNN3, também pelo número de pilhas de blocos residuais (P). Portanto, o espaço de busca do tamanho do canal de entrada (L) foi limitado por essas variáveis, como mostrado na Seção 3.2 do Capítulo 3.

Variável	Parâmetro	Tipo
K	Número de filtros da camada de convolução	\mathbb{N}^+
F	Tamanho do filtro de convolução	\mathbb{N}^+
L	Tamanho do canal de entrada (quantidade de lags usado para previsão)	\mathbb{N}^+
D	Porcentagem de dropout	\mathbb{R}^+
N	Normalização em lote	\mathbb{N}^+

Tabela 2 – Hiperparâmetros em comum das redes CNN1, CNN2 e CNN3.

Variável	Parâmetro	Tipo
P_p	Probabilidade do pooling	\mathbb{N}^+
J_p	Tamanho da janela de pooling	\mathbb{N}^+
C_s	Número de camadas convolucionais simples	\mathbb{N}^+

Tabela 3 – Hiperparâmetros específicos da CNN1.

Variável	Parâmetro	Tipo
C_d	Número de camadas convolucionais causais dilatadas	\mathbb{N}^+

Tabela 4 – Hiperparâmetro específico da CNN2.

Variável	Parâmetro	Tipo
C_d	Número de camadas convolucionais causais dilatadas	\mathbb{N}^+
P	Número de pilhas de blocos residuais	\mathbb{N}^+

Tabela 5 – Hiperparâmetros específicos da CNN3.

Variável	Parâmetro	Valor
S_c	Stride de convolução	1
R	Regularização dos pesos	3
T	Taxa de aprendizado	0.1
O	Algoritmo de otimização	Adam
A	Função de ativação	ReLU

Tabela 6 – Hiperparâmetros fixos em comum das redes CNN1, CNN2 e CNN3.

Variável	Parâmetro	Valor
S_p	Stride da janela de pooling	2
T_p	Tipo de pooling	Máx

Tabela 7 – Hiperparâmetros fixos da CNN1.

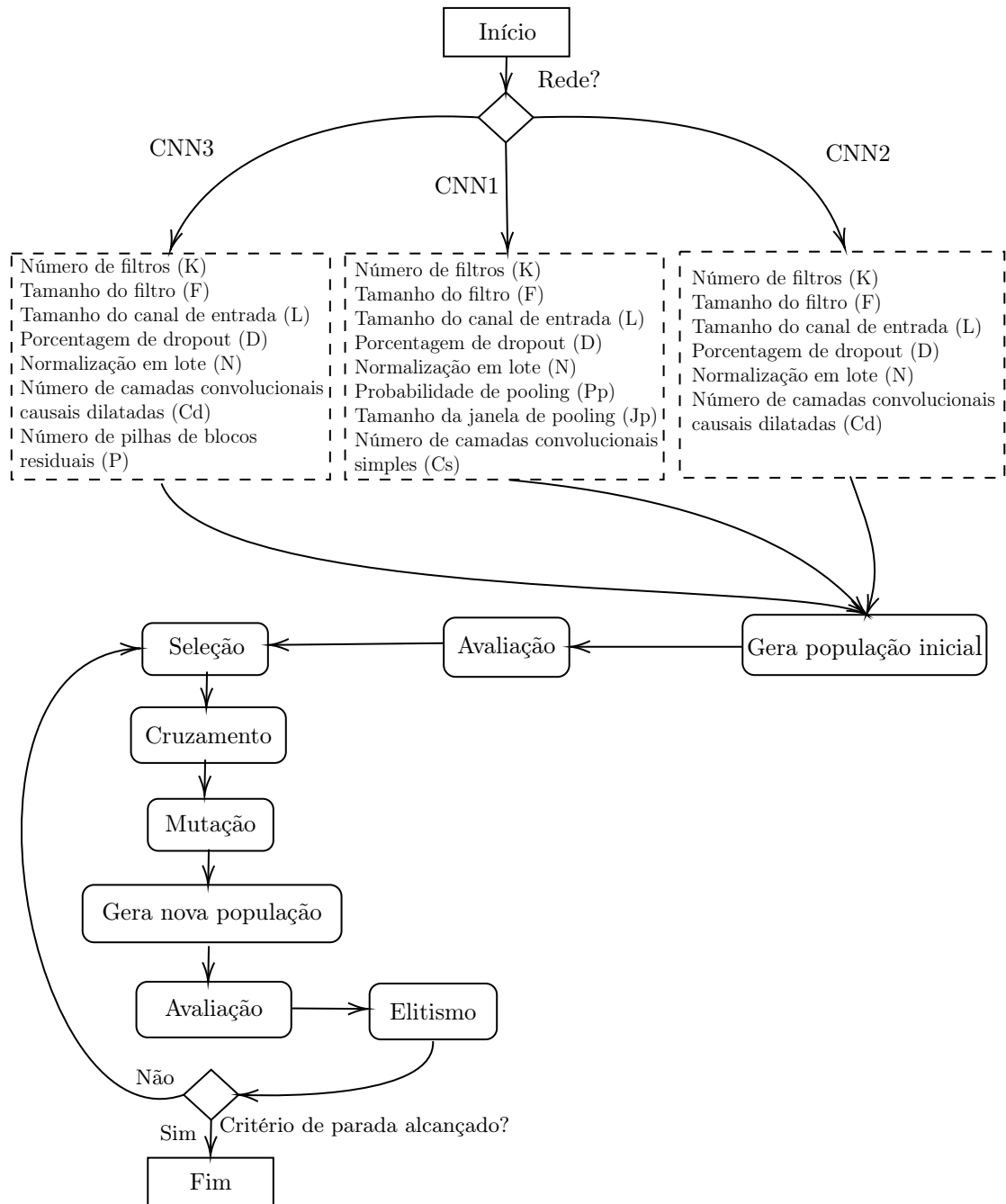


Figura 14 – Fluxograma do algoritmo genético

2. Avaliação

Na etapa de avaliação cada genótipo é transformado em um modelo treinado (o fenótipo) e então avaliado. Para isso seguiu-se o procedimento descrito em (Lucas et al. 2019) para avaliar a precisão dos modelos na série temporal de ET_t . Esse procedimento usa o método de janela deslizante (*sliding window*) ilustrado na Figura 15, onde o conjunto de dados Y de tamanho T é fatiado em sequências sobrepostas W , chamadas janelas de dados, com tamanho fixo igual a $T/2$ e cada uma deslocada da anterior por um valor constante D igual a $0.2 \cdot W$. Cada janela de dados foi dividida em dados de treino (70%) e teste (30%) e, em seguida, o método

Hiperparâmetro	Espaço de busca
K	[16,32,64]
F	[2,3,5,11]
L	[1.. 50]
D	[0.. 0.5]
N	Sim, Não
P_p	[0, 50, 100]
J_p	[2,3,4,5]
C_s	[1.. 5]

Tabela 8 – Espaços de busca dos hiperparâmetros da CNN1.

Hiperparâmetro	Espaço de busca
K	[16,32,64]
F	[2,3,5,11]
D	[0.. 0.5]
N	Sim, Não
C_d	[1.. 5]
L	$[(2^{C_d}).. (2^{C_d} \cdot F)]$

Tabela 9 – Espaços de busca dos hiperparâmetros da CNN2.

Hiperparâmetro	Espaço de busca
K	[16,32,64]
F	[2,3,5,11]
D	[0.. 0.5]
N	Sim, Não
C_d	[1.. 5]
P	[1,2]
L	$[(2^{C_d}).. (2^{C_d} \cdot F \cdot P)]$

Tabela 10 – Espaços de busca dos hiperparâmetros da CNN3.

foi treinado e testado N vezes com dados diferentes, onde N é igual a $(T - W)/D + 1$.

Para os experimentos foi utilizado um conjunto Y com T igual a 9 anos de dados da série temporal de ET_0 , obtendo assim, 6 experiências de treinamento e teste. A avaliação do fenótipo usa as métricas de acordo com a função objetivo f_1 e f_2 .

3. Seleção

O operador de seleção é responsável por escolher os indivíduos mais aptos da população para reprodução. Como o problema é multi-objetivo, uma estratégia de Torneio Duplo foi implementada para equilibrar a seleção entre os dois objetivos. No torneio duplo, a primeira rodada escolhe aleatoriamente dois pares de indivíduos na população, e cada dupla competirá entre si com base no objetivo f_1 . No segundo turno, os vencedores do primeiro turno competirão entre si com base no objetivo f_2 .

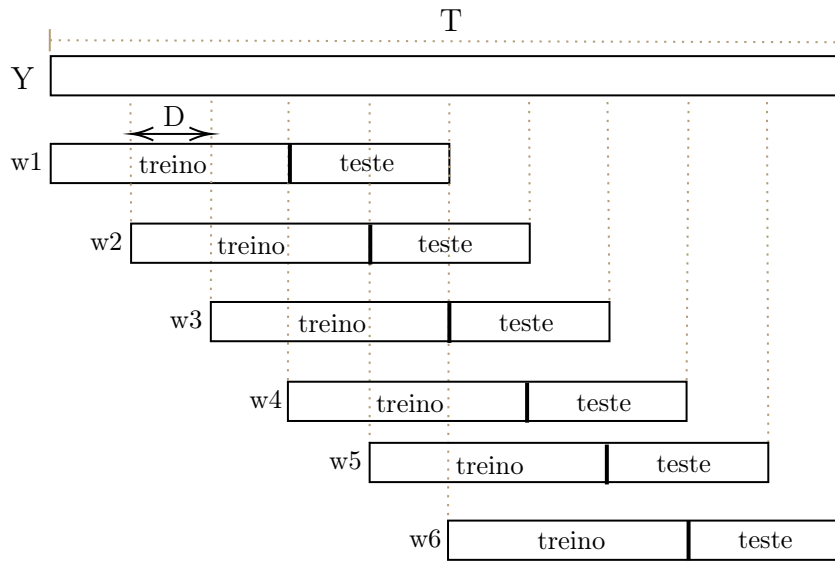


Figura 15 – Método de janela deslizante.

4. Cruzamento

O cruzamento é o operador que combina características dos pais durante a reprodução, permitindo que as próximas gerações herdem essas características. No cruzamento, os pais são ordenados como i_1 e i_2 de acordo com seus objetivos f_1 . Para todos os genes, o processo de combinação dará uma grande contribuição para o indivíduo mais bem classificado (com taxas de 0.7 e 0.3). Para os genes codificados reais, uma combinação linear como $i_N = 0.7i_1 + 0.3i_2$ será executada. Para os genes categóricos, o valor de i_N será i_1 com probabilidade 0.7 ou i_2 , com 0.3. Os genes L da CNN2 e CNN3 são gerados aleatoriamente usando os valores de C_d , F e P como mostrado nas Tabelas 9 e 10. O cruzamento é aplicado com probabilidade dada pela taxa de cruzamento igual a 0.5.

5. Mutação

O operador de mutação é necessário para a introdução e manutenção da diversidade genética da população. Para isso, o indivíduo é escolhido aleatoriamente e perturbações aleatórias, restritas aos espaços de busca de cada gene, são aplicadas em seus genes com probabilidade dada pela taxa de mutação igual a 0.3.

6. Elitismo

O elitismo é um operador que previne que os melhores indivíduos não desapareçam da população pela manipulação dos operadores genéticos, mantendo o melhor indivíduo da geração atual na próxima geração e descartando o pior.

7. Critério de parada

O AG irá parar quando for atingido um número de 7 gerações sem melhoria ou um número máximo de 30 gerações.

Hiperparâmetro	Resultado
K	32
F	11 x 11
D	0
N	Sim
C_s	1
L	4
P_p	100%
J_p	3
f_1	0.76
f_2	449

Tabela 11 – Resultado da otimização dos hiperparâmetros da CNN1.

4.3 Resultado da otimização dos hiperparâmetros das CNNs

A Figura 16 mostra a convergência dos objetivos f_1 e f_2 durante a otimização dos hiperparâmetros da CNN1, onde a linha tracejada representa a média dos valores dos objetivos da população e a linha contínua os valores dos objetivos do melhor indivíduo ao longo das gerações. Nota-se que os valores de f_1 do melhor indivíduo não variam muito ao longo das gerações, mesmo para grandes variações de f_2 . Essa observação evidencia a importância do objetivo f_2 , na otimização de hiperparâmetros de modelos complexos como uma CNN1, uma vez que funciona como um procedimento de regularização. Neste caso, verificou-se que após as 30 gerações o AG conseguiu equilibrar os dois objetivos, apresentando como resultado de 1 execução o modelo da Tabela 11. Apesar da estocasticidade do algoritmo genético, limitou-se em 1 execução do mesmo devido ao tempo disponível para desenvolvimento dos experimentos.

Nos experimentos da CNN2 e CNN3 foi identificado o mesmo comportamento do AG, mas com critério de parada no número de gerações sem melhoria. Com a CNN2 o AG parou na 23ª geração e com a CNN3 na 16ª geração. Os modelos CNN2 e CNN3 gerados pelo AG são apresentados nas Tabelas 12 e 13, respectivamente.

4.4 Resultado da otimização dos hiperparâmetros dos modelos de comparação

Long short term memory networks - LSTM

Proposta por (Hochreiter and Schmidhuber 1997), a LSTM é uma solução para o problema do gradiente de fuga das redes recorrentes. Além disso, seus módulos de memória conseguem resolver o problema da dependência de sequências a longo prazo,

Hiperparâmetro	Resultado
K	32
F	2 x 2
D	0
N	Sim
C_d	1
L	4
f_1	0.75
f_2	449

Tabela 12 – Resultado da otimização dos hiperparâmetros da CNN2.

Hiperparâmetro	Resultado
K	16
F	3 x 3
D	0.2
N	Sim
C_d	3
P	2
L	48
f_1	0.75
f_2	68257

Tabela 13 – Resultado da otimização dos hiperparâmetros da CNN3.

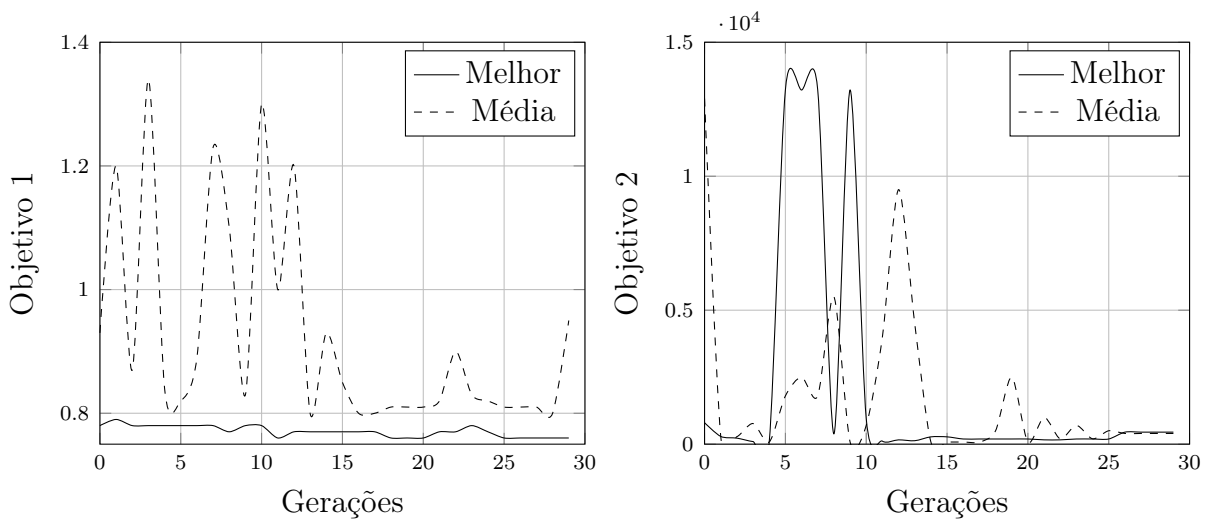


Figura 16 – Convergência dos objetivos 1 e 2 durante a otimização de hiperparâmetros da CNN1.

Hiperparâmetro	Espaço de busca	Resultado
Número de unidades	[50..100]	72
Porcentagem de dropout	[0..0.5]	0.3
Normalização em lote	sim, não	sim
Quantidade de lags	[1..50]	3
Número de camadas	[1..5]	1

Tabela 14 – Espaços de busca e resultado da otimização dos hiperparâmetros da LSTM.

Hiperparâmetro	Espaço de busca	Resultado
Número de partições	[10, 100]	16
Método de particionamento	grid, entropy	grid
Função de pertinência	triangular, trapezoidal, gaussiana	trapezoidal
Ordem	[1, 3]	2
α -cut	[0, 0.5]	0.2
Lags	[1, 50]	[2, 9]

Tabela 15 – Espaços de busca e resultado da otimização dos hiperparâmetros do WFTS.

obtendo sucesso significativo com aplicações de previsão de séries temporais, como mostrado em (Siami-Namini and Namin 2018; Yan and Ouyang 2018), com séries financeiras e em (Zhang et al. 2019; Gensler et al. 2017), com previsão de energia solar.

A otimização dos hiperparâmetros da LSTM foi realizada usando o AG descrito neste capítulo, com as devidas adaptações. A Tabela 14 apresenta os hiperparâmetros, espaços de busca e resultados da otimização.

Weighted fuzzy time series - WFTS

Os métodos de previsão do STN produzem modelos orientados a dados e não paramétricos e se tornaram atraentes devido à simplicidade, versatilidade, acurácia e desempenho computacional, além de produzir representações legíveis dos padrões de séries temporais, tornando seu conhecimento passível de ser auditado, reutilizável e atualizável. Exemplos de aplicações bem-sucedidas são mostrados em previsão de carga de energia (Sadaei et al. 2017; Severiano et al. 2017; Silva et al. 2018), previsão de preços do índice de estoque (Sadaei et al. 2016; Silva et al. 2016; Mirzaei Talarposhti et al. 2016), séries temporais (Cândido De Lima E Silva et al. 2017; Guney et al. 2018) entre outros.

Para a otimização dos hiperparâmetros do WFTS foi utilizado no AG desenvolvido por (Lucas et al. 2019), implementado na biblioteca PyFTS¹. A Tabela 15 apresenta os hiperparâmetros, espaços de busca e resultados da otimização.

Autoregressive integrated moving average - ARIMA

Para a otimização dos hiperparâmetros do ARIMA utilizou-se a função `auto_arima`

¹ <https://pyfts.github.io/pyFTS/>

da biblioteca Pyramid². O resultado encontrado foi $p = 1$, $d = 1$ e $q = 1$, onde p é a ordem do modelo, d é o grau de diferenciação e q é a ordem do modelo de média móvel.

4.5 Resumo

Neste capítulo foi descrito o algoritmo genético usado como abordagem para a otimização dos hiperparâmetros das redes CNNs descritas no capítulo anterior. Além disso, apresentou-se também as abordagens utilizadas na otimização dos modelos LSTM, WFTS e ARIMA. No próximo capítulo, os melhores modelos CNN1, CNN2, CNN3, LSTM, WFTS e ARIMA são avaliados e seus desempenhos na previsão de ET_o são comparados.

² <https://github.com/alkaline-ml/pmdarima>

Capítulo 5

Resultados

Neste capítulo são apresentados e discutidos os resultados da previsão de séries temporais diárias de ET_o . Os experimentos foram conduzidos com a série temporal diária de ET_o descrita na seção 2.3 do capítulo 2, onde 18 anos (1993 a 2011) foram usados para treinamento e 4 anos (2012, 2013, 2014 e 2015) para teste. Para cada ano de teste foram previstos 1, 3, 7 e 10 passos a frente, sendo o RMSE de cada horizonte de previsão usado para medir o desempenho dos modelos.

Os modelos CNN1, CNN2 e CNN3, descritos nas Tabelas 11, 12 e 13, foram comparados em três momentos. Na seção 5.1 esses modelos são validados a partir da comparação com os modelos LSTM, WFTS e ARIMA, descritos na seção 4.4 do capítulo anterior, na seção 5.2 são comparados a modelos *ensemble* e na seção 5.3 os modelos CNN e ensembles são comparados ao modelo Naïve. Além disso, é feita na seção 5.4 a análise dos resíduos dos modelos CNN e ensembles.

Todos os modelos utilizados nos experimentos foram implementados seguindo a interface representada pela Figura 17, onde cada horizonte previsto é usado para atualizar as entradas das próximas previsões.

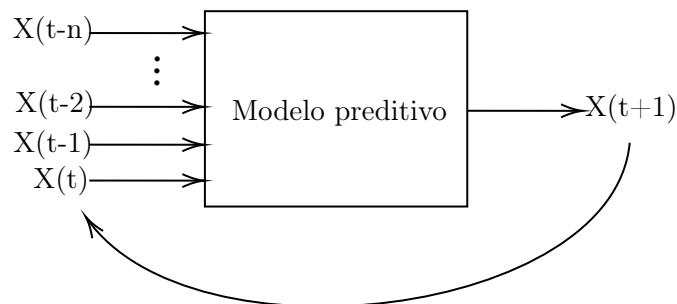


Figura 17 – Interface dos modelos preditivos

Horizonte de previsão	CNN1	CNN2	CNN3	LSTM	ARIMA	WFTS
1	0.80 \pm 0.00	0.80 \pm 0.01	0.81 \pm 0.01	0.81 \pm 0.01	0.79	0.82
3	0.88 \pm 0.04	0.98 \pm 0.05	0.97 \pm 0.05	0.99 \pm 0.05	0.92	0.95
7	1.07 \pm 0.11	1.14 \pm 0.11	1.11 \pm 0.08	1.16 \pm 0.09	1.00	1.03
10	1.19 \pm 0.14	1.21 \pm 0.15	1.19 \pm 0.10	1.22 \pm 0.11	1.03	1.07

Tabela 16 – RMSE em mmd^{-1} da previsão de 1, 3, 7 e 10 dias a frente dos modelos CNN1, CNN2, CNN3, LSTM, ARIMA e WFTS para o ano de 2012.

5.1 Comparação entre os modelos CNN, LSTM, WFTS e ARIMA

As tabelas 16, 17, 18 e 19 apresentam a média do RMSE e o desvio padrão de 30 execuções dos modelos CNN1, CNN2, CNN3 e LSTM, bem como o RMSE dos modelos WFTS e ARIMA das previsões feitas para os horizontes de previsão 1, 3, 7 e 10 dias dos anos de teste 2012, 2013, 2014 e 2015. Esses resultados também podem ser visualizados graficamente nos boxplots das Figuras 18, 20, 22 e 24.

Além da análise gráfica, foi realizado um teste de hipótese para verificar a existência de diferenças significativas na média do RMSE dos modelos CNN e LSTM. Uma vez que a distribuição dos resíduos das amostras não eram normais, realizou-se o teste de Friedman considerando um nível de confiança de 95%. A hipótese nula H_0 é a ausência de diferenças entre as médias e a hipótese alternativa H_1 é a existência de pelo menos uma diferença. Verificada a existência de diferenças estatísticas, usou-se o teste Nemenyi para visualizar quais deles possuem médias diferentes para cada ano de teste e horizonte de previsão, como mostram as Figuras 19, 21, 23 e 25.

À luz dos resultados obtidos, percebeu-se que os modelos ARIMA e WFTS apresentaram comportamento semelhante aos modelos *deep learning* nos 4 anos de teste, especialmente nos horizontes de previsão 1 e 3 onde obtiveram desempenhos muito próximos. Esses resultados podem ser explicados pelo comportamento da série temporal de ET_o utilizada que, segundo o teste de Dickey-Fuller aplicado, possui 99% de confiança de estacionariedade.

Observando apenas os métodos *deep learning*, nas Figuras 19, 21, 23 e 25, percebeu-se que não houve diferença estatística no horizonte de previsão 10. A medida que os passos de previsão diminuem, o desempenho do modelo CNN1 se destaca em relação aos demais. Quando comparados os modelos CNN2, CNN3 e LSTM, que são modelos voltados para aplicações em dados de sequência, observa-se que esses modelos obtiveram médias, na maioria das vezes, estatisticamente iguais, independente do horizonte de previsão.

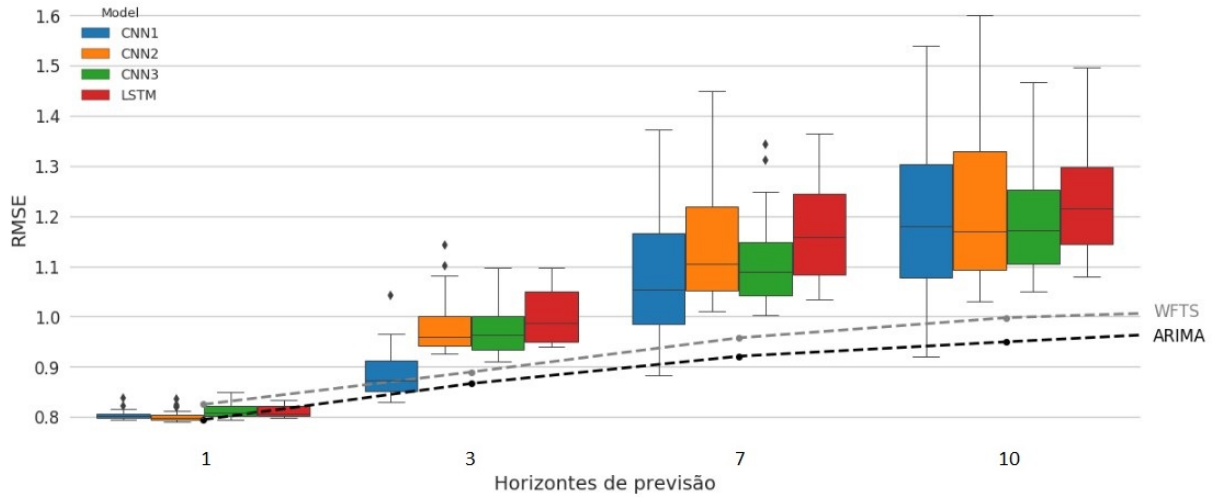


Figura 18 – Boxplot do RMSE dos modelos CNN1, CNN2, CNN3 e LSTM para os horizontes de previsão 1, 3, 7 e 10 do ano de 2012.

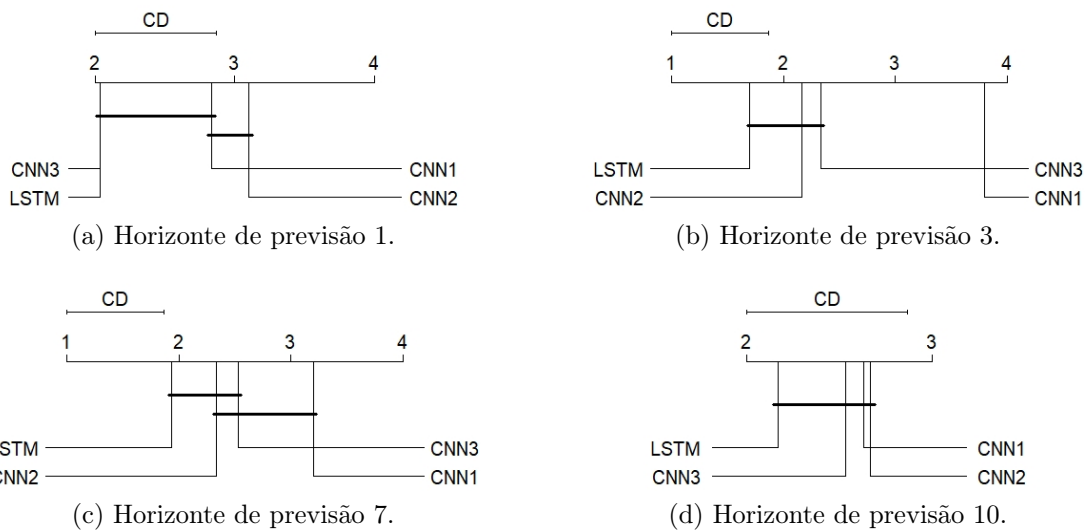


Figura 19 – Visualização para o Teste de Friedman dos resultados dos modelos CNN e LSTM do ano de 2012.

Horizonte de previsão	CNN1	CNN2	CNN3	LSTM	ARIMA	WFTS
1	0.79 ±0.00	0.80 ±0.01	0.80 ±0.01	0.80 ±0.00	0.80	0.80
3	0.85 ±0.02	1.03 ±0.03	1.00 ±0.03	1.02 ±0.02	1.00	0.98
7	0.97 ±0.07	1.09 ±0.09	1.06 ±0.05	1.07 ±0.05	1.02	0.98
10	1.06 ±0.10	1.19 ±0.12	1.13 ±0.07	1.14 ±0.06	1.07	1.02

Tabela 17 – RMSE em mmd^{-1} da previsão de 1, 3, 7 e 10 dias a frente dos modelos CNN1, CNN2, CNN3, LSTM, ARIMA e WFTS para o ano de 2013.

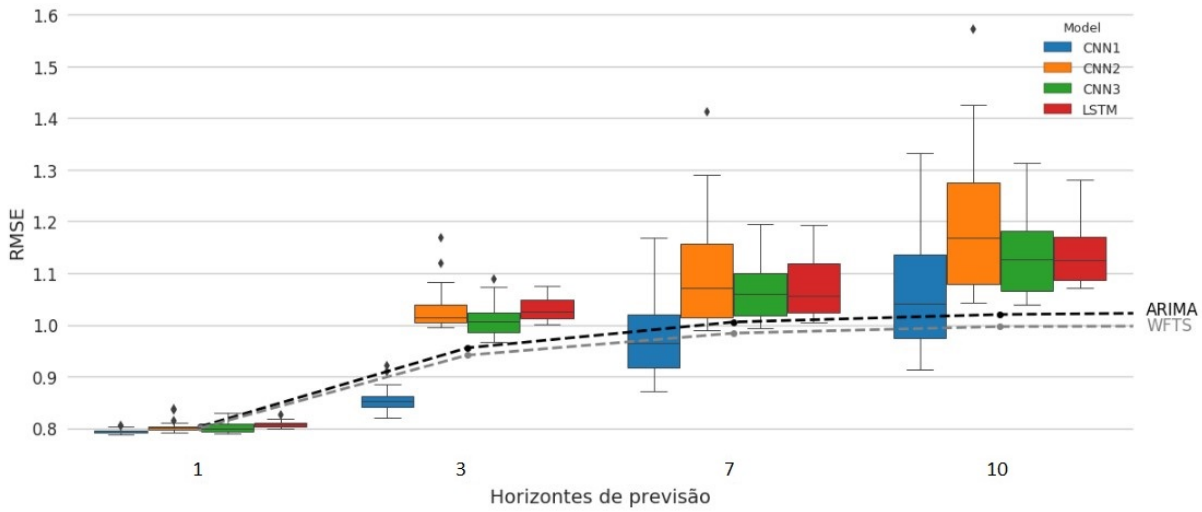


Figura 20 – Boxplot do RMSE dos modelos CNN1, CNN2, CNN3 e LSTM para os horizontes de previsão 1, 3, 7 e 10 do ano de 2013.

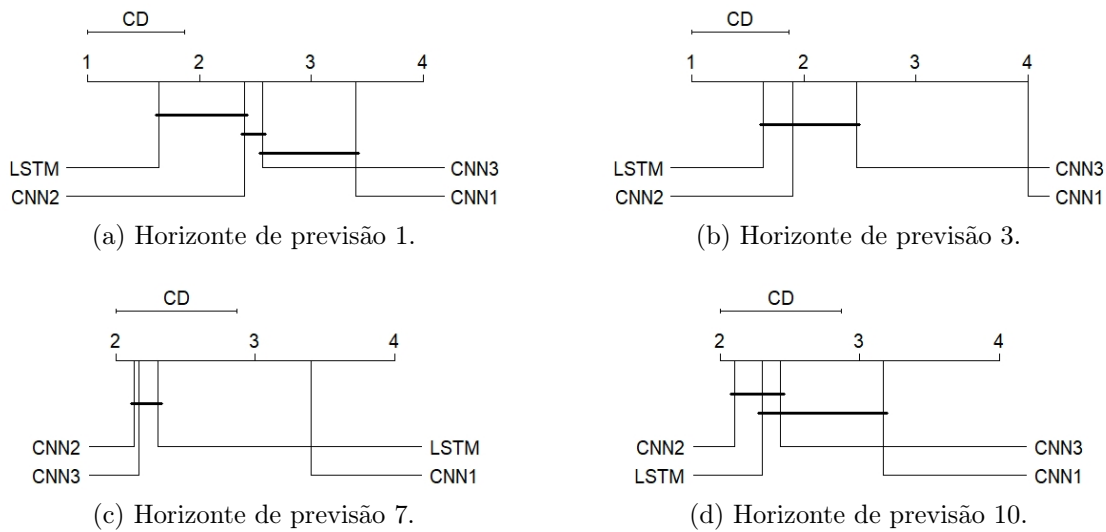


Figura 21 – Visualização para o Teste de Friedman dos resultados dos modelos CNN e LSTM do ano de 2013.

Horizonte de previsão	CNN1	CNN2	CNN3	LSTM	ARIMA	WFTS
1	0.74 ± 0.01	0.74 ± 0.01	0.72 ± 0.02	0.72 ± 0.01	0.74	0.74
3	0.87 ± 0.07	0.99 ± 0.05	0.95 ± 0.07	0.96 ± 0.05	0.92	0.97
7	1.11 ± 0.16	1.13 ± 0.12	1.07 ± 0.12	1.13 ± 0.11	0.96	1.01
10	1.25 ± 0.23	1.20 ± 0.16	1.14 ± 0.14	1.19 ± 0.13	0.98	1.05

Tabela 18 – RMSE em mmd^{-1} da previsão de 1, 3, 7 e 10 dias a frente dos modelos CNN1, CNN2, CNN3, LSTM, ARIMA e WFTS para o ano de 2014.

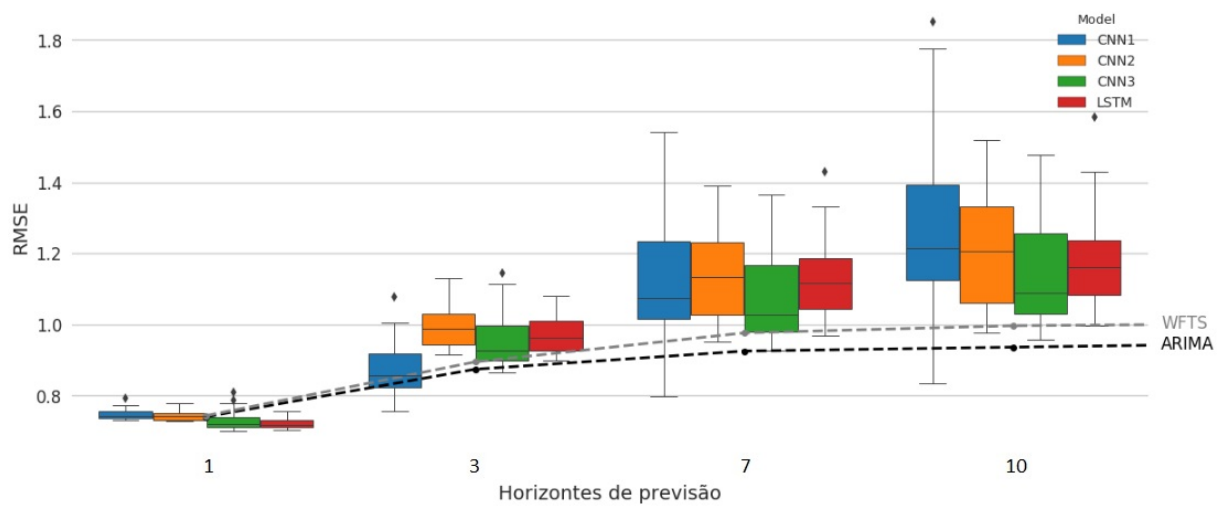


Figura 22 – Boxplot do RMSE dos modelos CNN1, CNN2, CNN3 e LSTM para os horizontes de previsão 1, 3, 7 e 10 do ano de 2014.

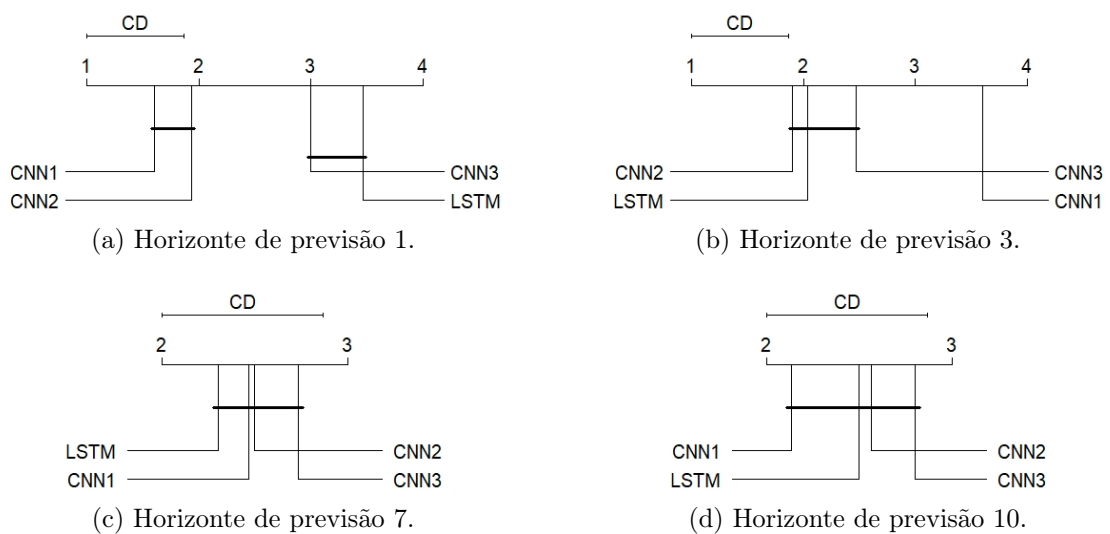


Figura 23 – Visualização para o Teste de Friedman dos resultados dos modelos CNN e LSTM do ano de 2014.

Horizonte de previsão	CNN1	CNN2	CNN3	LSTM	ARIMA	WFTS
1	0.76 ± 0.00	0.77 ± 0.01	0.77 ± 0.00	0.77 ± 0.01	0.76	0.77
3	0.84 ± 0.04	0.97 ± 0.03	0.93 ± 0.03	0.97 ± 0.03	0.91	0.93
7	0.99 ± 0.10	1.08 ± 0.09	1.04 ± 0.05	1.1 ± 0.07	0.96	1.01
10	1.09 ± 0.14	1.14 ± 0.13	1.10 ± 0.07	1.15 ± 0.09	0.98	1.04

Tabela 19 – RMSE em mmd^{-1} da previsão de 1, 3, 7 e 10 dias a frente dos modelos CNN1, CNN2, CNN3, LSTM, ARIMA e WFTS para o ano de 2015.

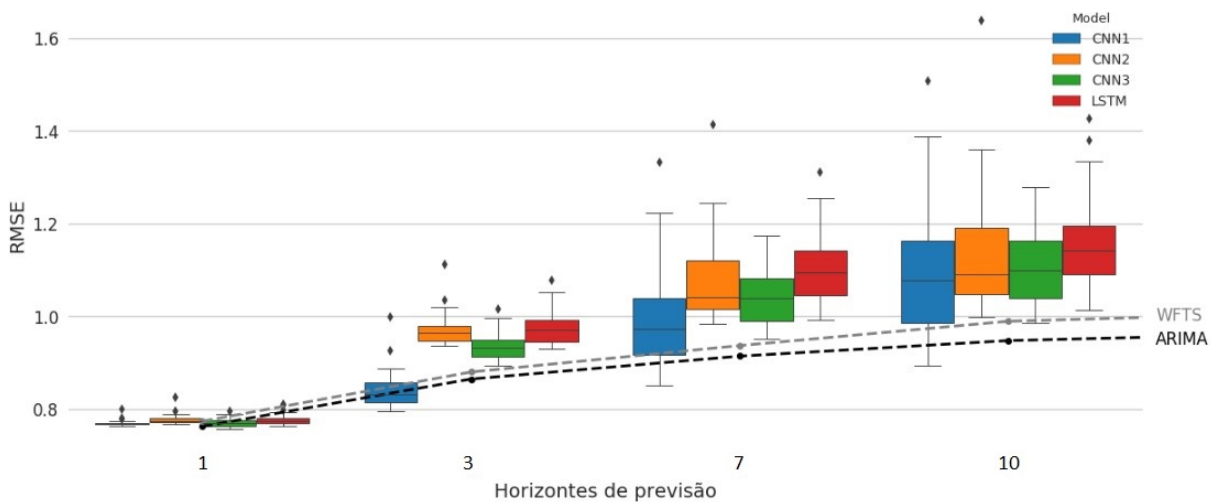


Figura 24 – Boxplot do RMSE dos modelos CNN1, CNN2, CNN3 e LSTM para os horizontes de previsão 1, 3, 7 e 10 do ano de 2015.

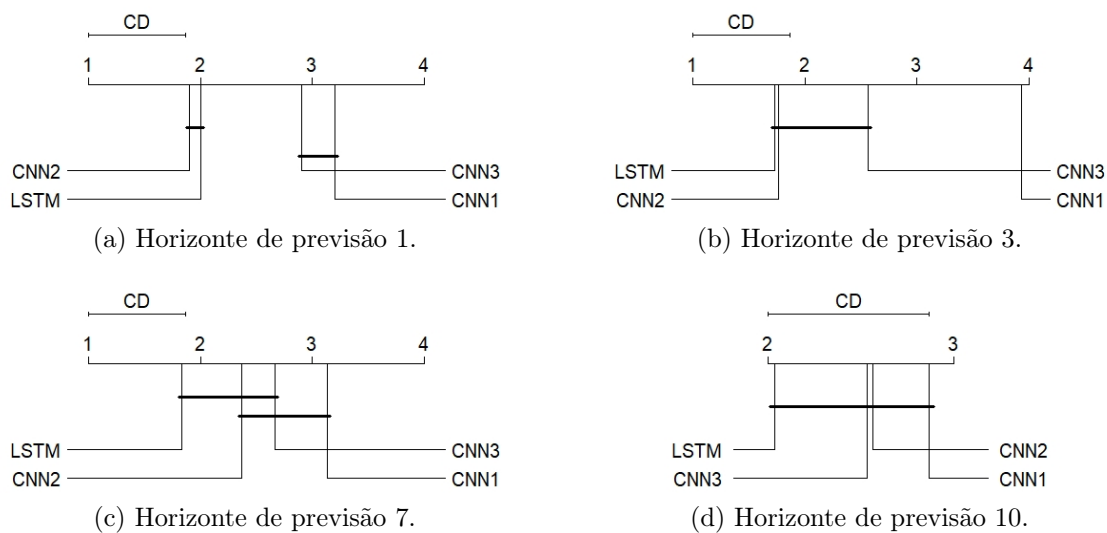


Figura 25 – Visualização para o Teste de Friedman dos resultados dos modelos CNN e LSTM do ano de 2015.

5.2 Comparação entre os modelos CNN e *Ensemble Learnings*

Para esta comparação foram criados 4 *ensembles* compostos e nomeados como na Tabela 20. A geração dos modelos que compõem os *ensembles* foi feita a partir da geração aleatória dos hiperparâmetros da Tabela 8 para modelos CNN1, da Tabela 9 para modelos CNN2 e da Tabela 10 para modelos CNN3.

O treinamento e teste dos modelos que compõem os *ensembles* foram realizados como descrito na introdução deste capítulo e suas respectivas saídas foram combinadas usando uma função de densidade de probabilidade gerada a partir do método KDE (descrito

<i>Ensemble</i>	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
<i>Ensemble</i> CNN1	CNN1	CNN1	CNN1	CNN1	CNN1	CNN1
<i>Ensemble</i> CNN2	CNN2	CNN2	CNN2	CNN2	CNN2	CNN2
<i>Ensemble</i> CNN3	CNN3	CNN3	CNN3	CNN3	CNN3	CNN3
<i>Ensemble</i> Híbrido	CNN1	CNN1	CNN2	CNN2	CNN3	CNN3

Tabela 20 – Composição dos *ensemble*.

na seção 3.3 do capítulo 3) com kernel gaussiano. O parâmetro largura de banda (h) foi otimizado por *grid search* com valores no intervalo $[0, 1]$.

As tabelas 21, 23, 24 e 25 apresentam a média do RMSE de 30 execuções dos modelos CNN1, CNN2, CNN3, *Ensemble* CNN1, *Ensemble* CNN2, *Ensemble* CNN3 e *Ensemble* Híbrido das previsões feitas para os horizontes de previsão 1, 3, 7 e 10 para o ano de 2012, 2013, 2014 e 2015, respectivamente. Esses resultados também podem ser visualizados nos boxplots das Figuras 18, 20, 22 e 24.

Assim como na comparação anterior, foi realizado o teste de Friedman com nível de confiança de 95% para verificar a existência de diferenças significativas na média do RMSE de cada modelo. Como ficou constatada a existência de diferenças estatísticas, usou-se o teste Nemenyi para visualizar quais modelos possuem médias diferentes para cada ano de teste e horizonte de previsão, como mostram as Figuras 27, 33, 39 e 45.

Para melhor organização do texto, as discussões dos resultados foram feitas separadamente com os modelos CNN e *ensembles*, de acordo com a composição destes.

CNN1 versus *Ensemble* CNN1: conforme as Figuras 27, 33, 39 e 45, o *Ensemble* CNN1 apresentou desempenho igual ou inferior em relação ao modelo individual CNN1.

Na comparação da variância desses modelos, observou-se que apenas no ano de 2012 (Figura 18) não houve diferença, já nos demais anos (Figuras 20, 22 e 24) a redução da variância no modelo *Ensemble* CNN1 fica evidente, especialmente nos horizontes de previsão 7 e 10.

CNN2 versus *Ensemble* CNN2: conforme as Figuras 27, 33, 39 e 45, o *Ensemble* CNN2 apresentou melhoria na maioria dos resultados, ficando com desempenho abaixo do modelo CNN2 apenas no horizonte de previsão 1 dos anos de teste 2012 e 2014 e 10 de 2014.

Na comparação da variância desses modelos, observou-se a variância do modelo *Ensemble* CNN2 é menor em todos os horizontes de previsão de todos os anos de teste, conforme os boxplots das Figuras 18, 20, 22 e 24.

CNN3 versus *Ensemble* CNN3: conforme as Figuras 27, 33, 39 e 45, o *Ensemble* CNN3 mostrou-se com desempenho superior ao CNN3 em todos os horizontes de

Horizonte de previsão	CNN1	Ensemble CNN1	CNN2	Ensemble CNN2	CNN3	Ensemble CNN3	Ensemble híbrido
1	0.80	0.80	0.80	0.83	0.81	0.79	0.80
3	0.88	0.91	0.98	0.93	0.97	0.85	0.89
7	1.07	1.02	1.14	0.95	1.11	0.96	1.00
10	1.19	1.10	1.21	1.07	1.19	1.02	1.05

Tabela 21 – RMSE da previsão dos modelos CNN1, Ensemble CNN1, CNN2, Ensemble CNN2, CNN3, Ensemble CNN3 e Ensemble híbrido para o ano de 2012.

previsão dos anos 2012, 2013 e 2015. No ano 2014, o *Ensemble* CNN3 apresentou melhoria apenas no horizonte de previsão 3 e manteve mesmo desempenho do CNN3 nos demais horizontes de previsão.

Na comparação da variância desses modelos, observou-se a variância do modelo *Ensemble* CNN3 é menor em todos os horizontes de previsão de todos os anos de teste, conforme os boxplots das Figuras 18, 20, 22 e 24.

CNN1, CNN2 e CNN3 versus *Ensemble* Híbrido: conforme a Figura 27, o modelo *Ensemble* Híbrido apresentou mesmo desempenho do modelo CNN1 nos horizontes de previsão 1, 3 e 7 e desempenho superior no 10 no ano de 2012. Na comparação com os modelos CNN2 e CNN3, o *Ensemble* Híbrido mostrou-se melhor nesse ano. No ano 2013 (Figura 33), o *Ensemble* Híbrido apresentou melhor desempenho que o modelo CNN2 nos horizontes de previsão 3 e 10 e melhor que o modelo CNN3 no horizonte de previsão 3. Mas em comparação ao CNN1 não obteve melhoria. Em 2014 (Figura 39), o *Ensemble* Híbrido apresentou melhoria apenas no horizonte de previsão 10, obtendo mesmo desempenho na maioria dos casos. No ano 2015 (Figura 45), o *Ensemble* Híbrido apresentou melhor desempenho que os modelos individuais em todos os horizontes de previsão, com exceção do 1 e 3, em que apresentou mesma média apenas do modelo CNN1.

Na comparação da variância desses modelos, observou-se que a variância do modelo *Ensemble* Híbrido foi menor nos horizontes de previsão 7 e 10 de todos os anos de teste, conforme os boxplots das Figuras 18, 20, 22 e 24.

De maneira geral os modelos *ensemble* apresentaram melhores resultados em termos de estabilidade e os modelos *Ensemble* CNN2 e *Ensemble* CNN3 mostraram-se também mais precisos que os modelos individuais que os compõem. O desempenho desses modelos na previsão de 1 passo a frente pode ser visualizado nas Figuras 28, 29, 30 e 31 do ano de 2012, Figuras 34, 35, 36 e 37 do ano de 2013 e Figuras 40, 41, 42 e 43 do ano de 2014. É possível observar que os intervalos de confiança gerados pelos modelos *ensemble* são satisfatórios, uma vez que os valores reais de ET_o estão, na maioria das vezes, dentro dos seus limites inferior e superior. Percebe-se também que o modelo *Ensemble* CNN3 gerou o menor intervalo de confiança, ou seja, foi o modelo que apresentou menor incerteza sobre as previsões nos 4 anos de teste.

Horizonte de previsão	Ensemble CNN1	Ensemble CNN2	Ensemble CNN3	Ensemble híbrido
1	0.89	0.93	0.88	0.89
3	0.84	0.86	0.78	0.82
7	0.85	0.79	0.80	0.83
10	0.88	0.86	0.82	0.84

Tabela 22 – Scaled RMSE da previsão de 1, 3, 7 e 10 dias a frente dos modelos Ensemble CNN1, Ensemble CNN2, Ensemble CNN3 e Ensemble híbrido para o ano de 2012.

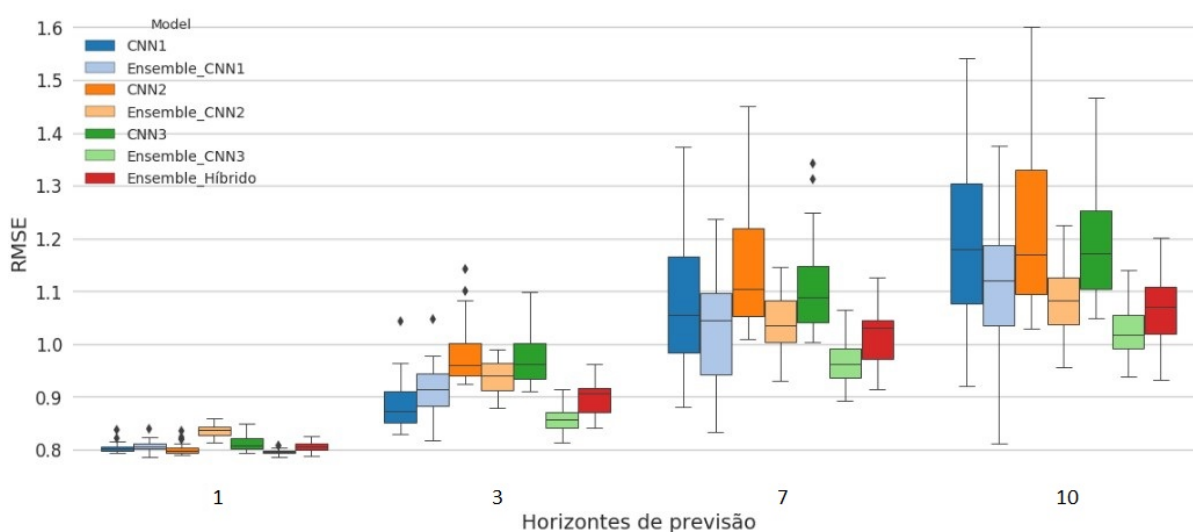


Figura 26 – Boxplot do RMSE dos modelos CNNs e Ensembles para os horizontes de previsão 1, 3, 7 e 10 do ano de 2012.

Horizonte de previsão	CNN1	Ensemble CNN1	CNN2	Ensemble CNN2	CNN3	Ensemble CNN3	Ensemble híbrido
1	0.79	0.80	0.80	0.80	0.80	0.78	0.79
3	0.85	0.87	1.03	0.89	1.00	0.85	0.90
7	0.97	0.98	1.09	0.95	1.06	0.95	1.01
10	1.06	1.01	1.019	0.97	1.13	0.98	1.05

Tabela 23 – RMSE da previsão dos modelos CNN1, Ensemble CNN1, CNN2, Ensemble CNN2, CNN3, Ensemble CNN3 e Ensemble híbrido para o ano de 2013.

Horizonte de previsão	CNN1	Ensemble CNN1	CNN2	Ensemble CNN2	CNN3	Ensemble CNN3	Ensemble híbrido
1	0.74	0.77	0.74	0.77	0.72	0.74	0.76
3	0.87	0.95	0.99	0.92	0.95	0.85	0.92
7	1.11	1.11	1.13	1.05	1.07	0.99	1.09
10	1.25	1.17	1.20	1.13	1.14	1.05	1.16

Tabela 24 – RMSE da previsão dos modelos CNN1, Ensemble CNN1, CNN2, Ensemble CNN2, CNN3, Ensemble CNN3 e Ensemble híbrido para o ano de 2014.

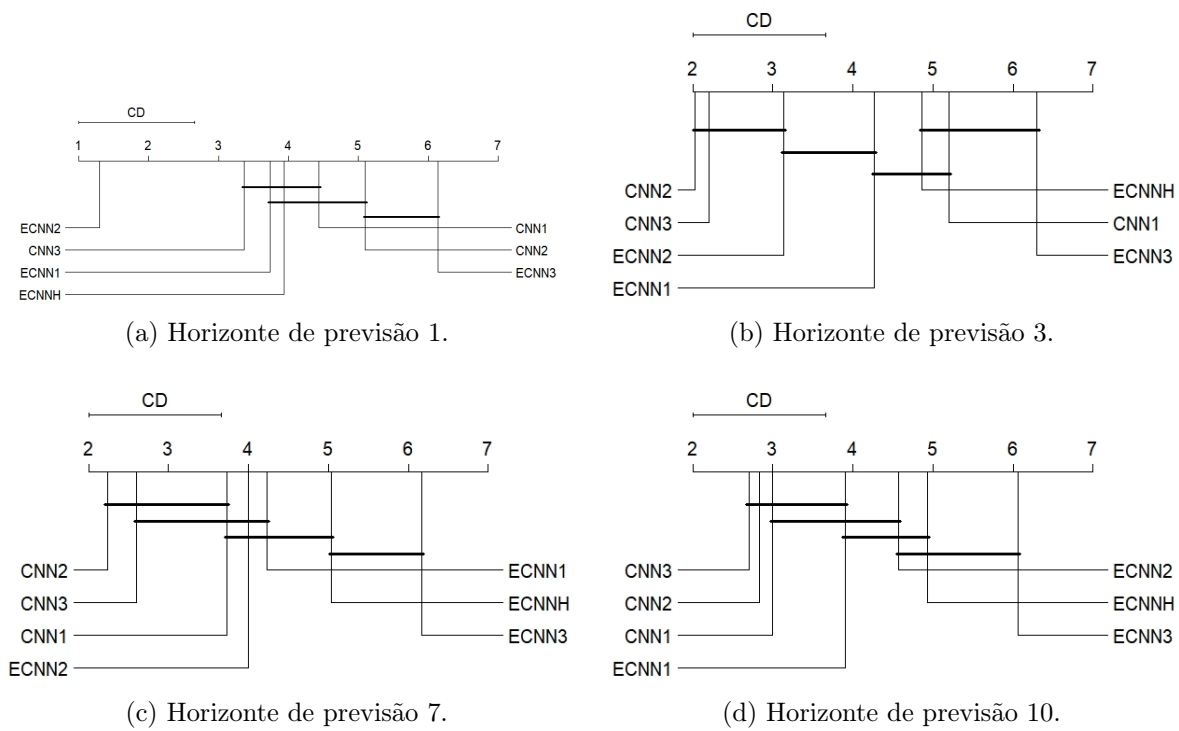


Figura 27 – Visualização para o Teste de Friedman dos resultados dos modelos CNNs e Ensembles do ano de 2012.

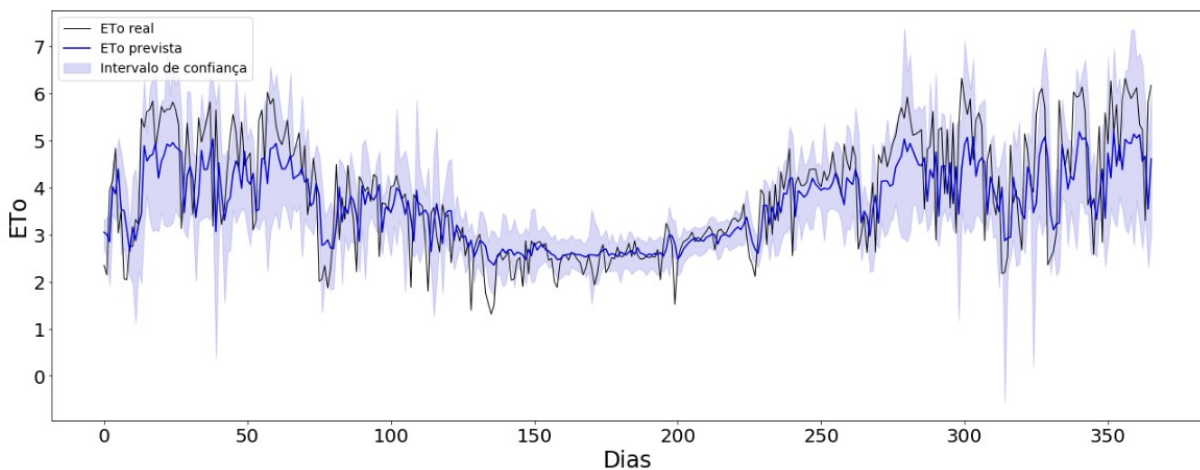


Figura 28 – Desempenho do modelo Ensemble CNN1 para horizonte de previsão 1 do ano de teste 2012.

Horizonte de previsão	CNN1	Ensemble CNN1	CNN2	Ensemble CNN2	CNN3	Ensemble CNN3	Ensemble híbrido
1	0.76	0.78	0.77	0.77	0.77	0.75	0.76
3	0.84	0.88	0.97	0.87	0.93	0.79	0.84
7	0.99	0.96	1.08	0.97	1.04	0.90	0.92
10	1.09	1.01	1.14	1.02	1.10	0.94	0.94

Tabela 25 – RMSE da previsão dos modelos CNN1, Ensemble CNN1, CNN2, Ensemble CNN2, CNN3, Ensemble CNN3 e Ensemble híbrido para o ano de 2015.

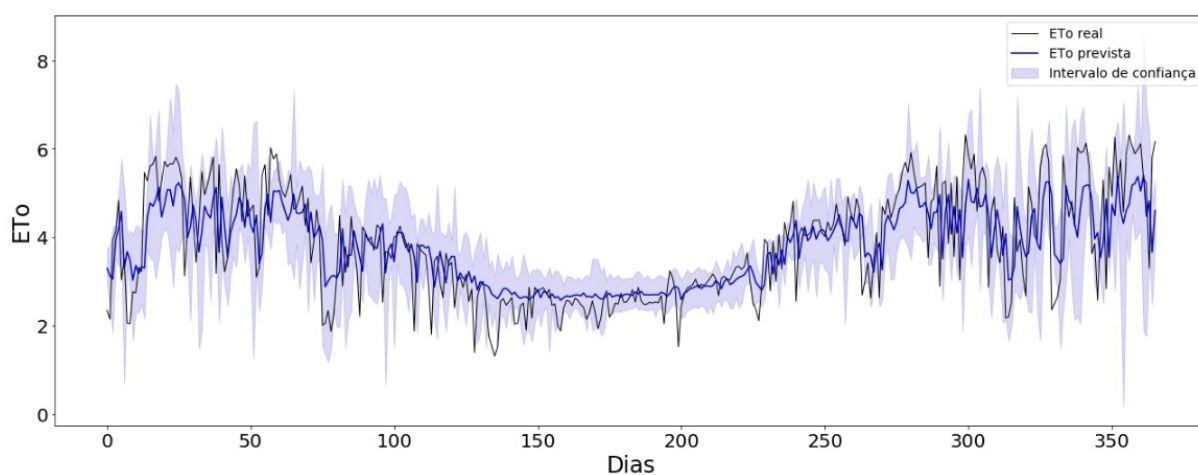


Figura 29 – Desempenho do modelo Ensemble CNN2 para horizonte de previsão 1 do ano de teste 2012.

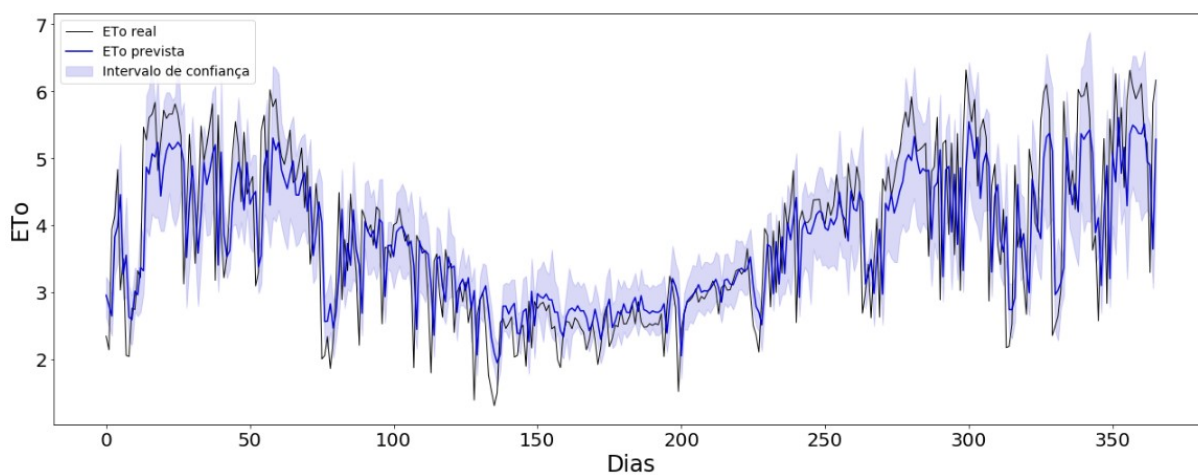


Figura 30 – Desempenho do modelo Ensemble CNN3 para horizonte de previsão 1 do ano de teste 2012.

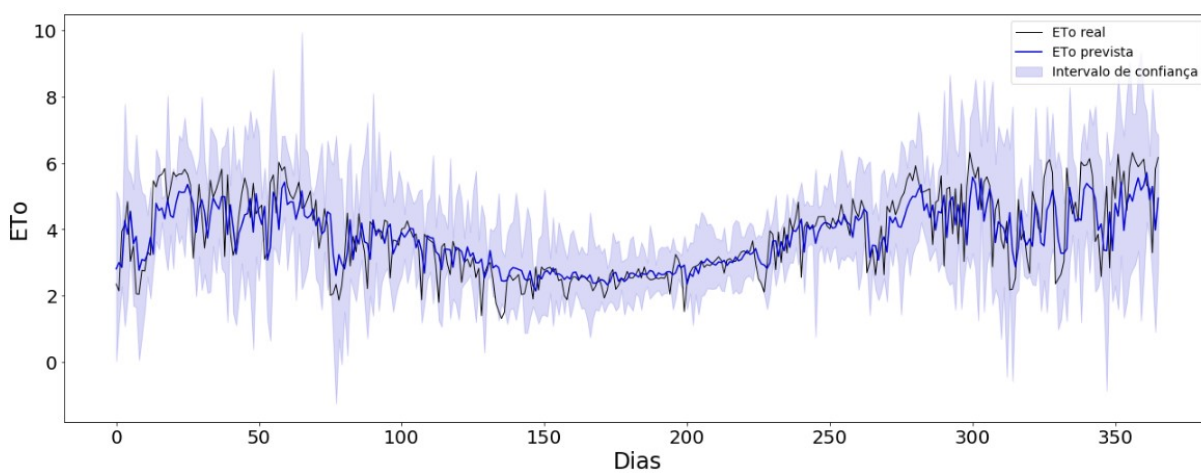


Figura 31 – Desempenho do modelo Ensemble Híbrido para horizonte de previsão 1 do ano de teste 2012.

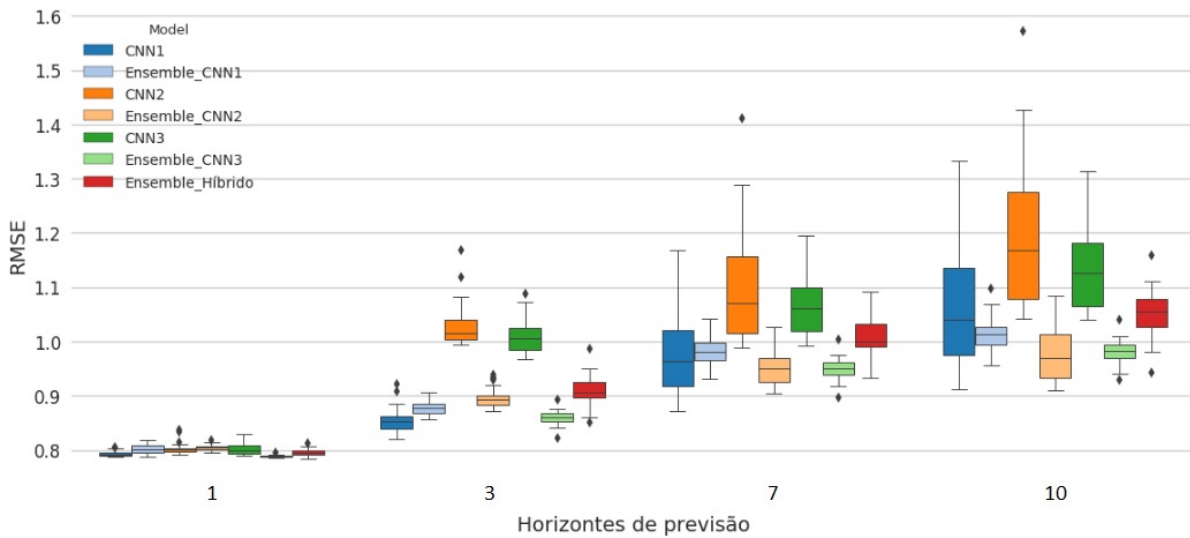


Figura 32 – Boxplot do RMSE dos modelos CNNs e Ensembles para os horizontes de previsão 1, 3, 7 e 10 do ano de 2013.

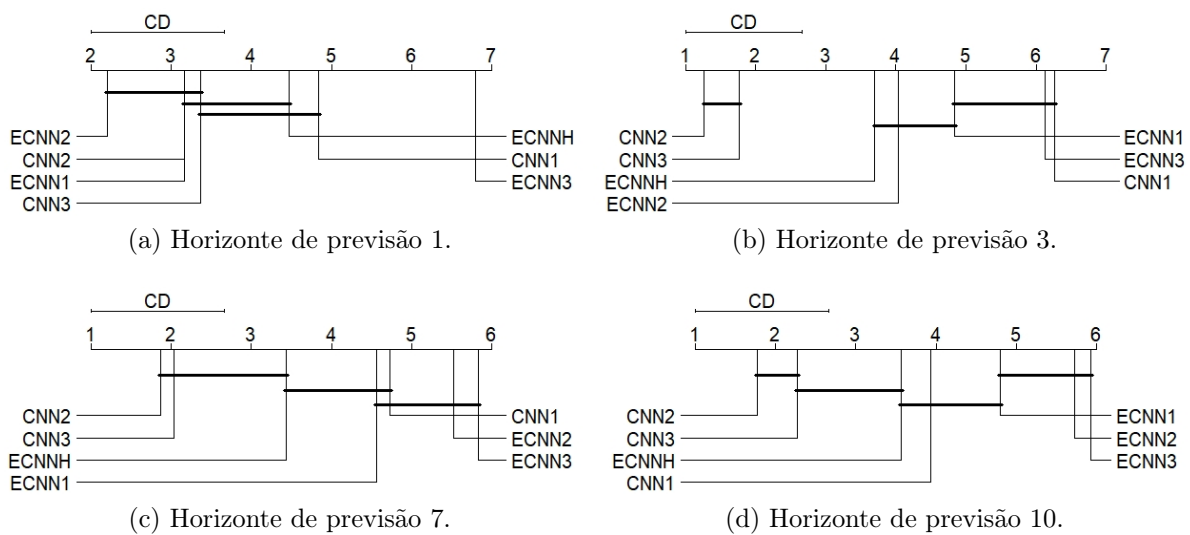


Figura 33 – Visualização para o Teste de Friedman dos resultados dos modelos CNNs e Ensembles do ano de 2013.

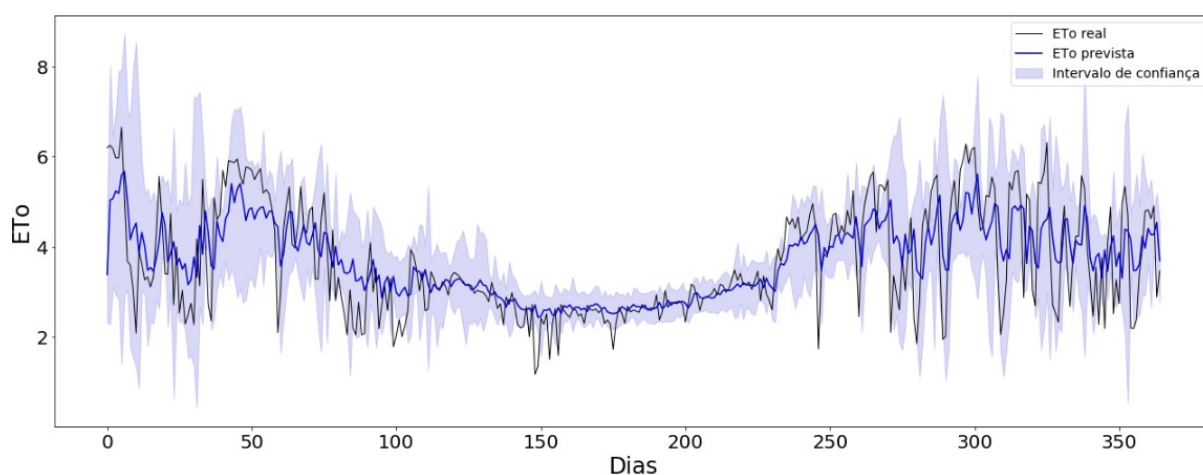


Figura 34 – Desempenho do modelo Ensemble CNN1 para no horizontes de previsão 1 do ano de teste 2013.

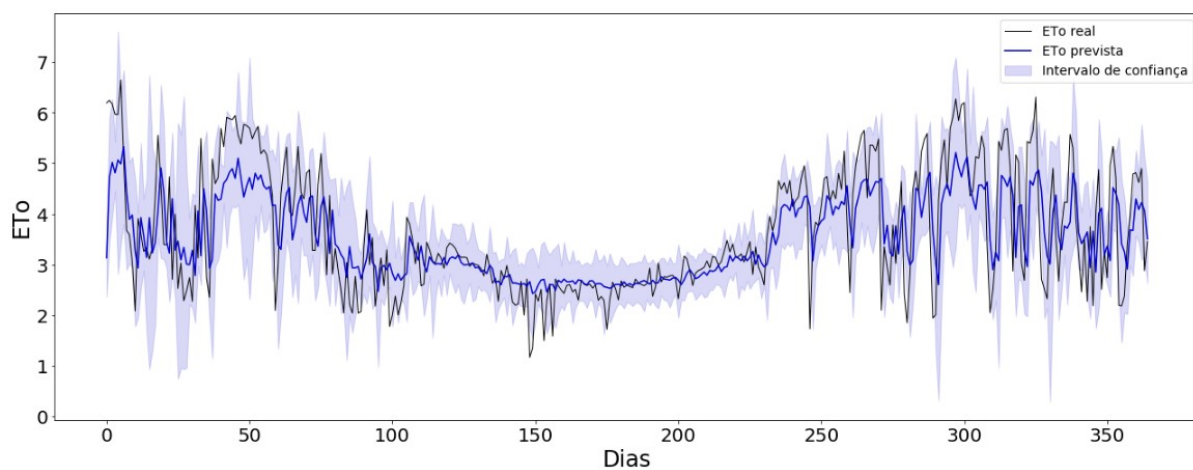


Figura 35 – Desempenho do modelo Ensemble CNN2 para no horizontes de previsão 1 do ano de teste 2013.

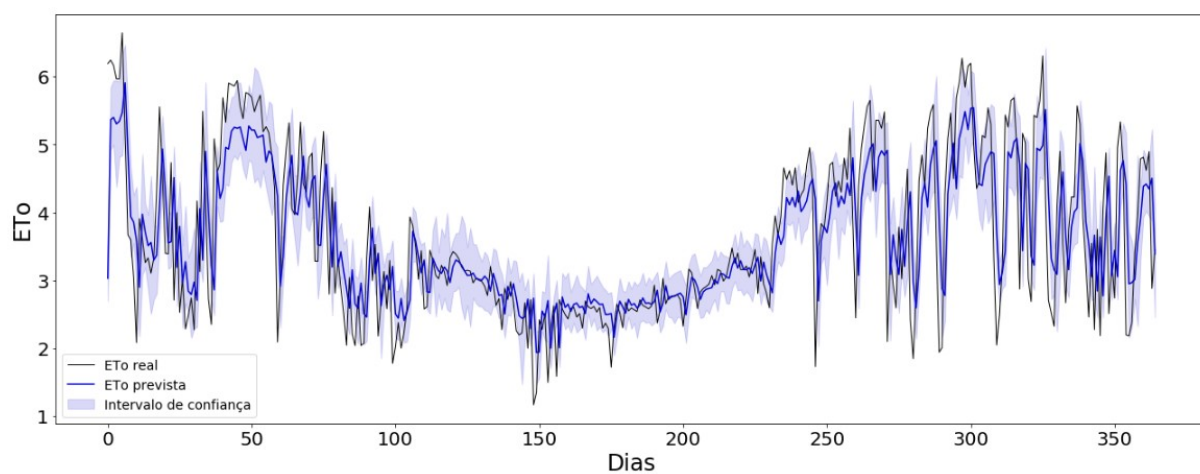


Figura 36 – Desempenho do modelo Ensemble CNN3 para no horizontes de previsão 1 do ano de teste 2013.

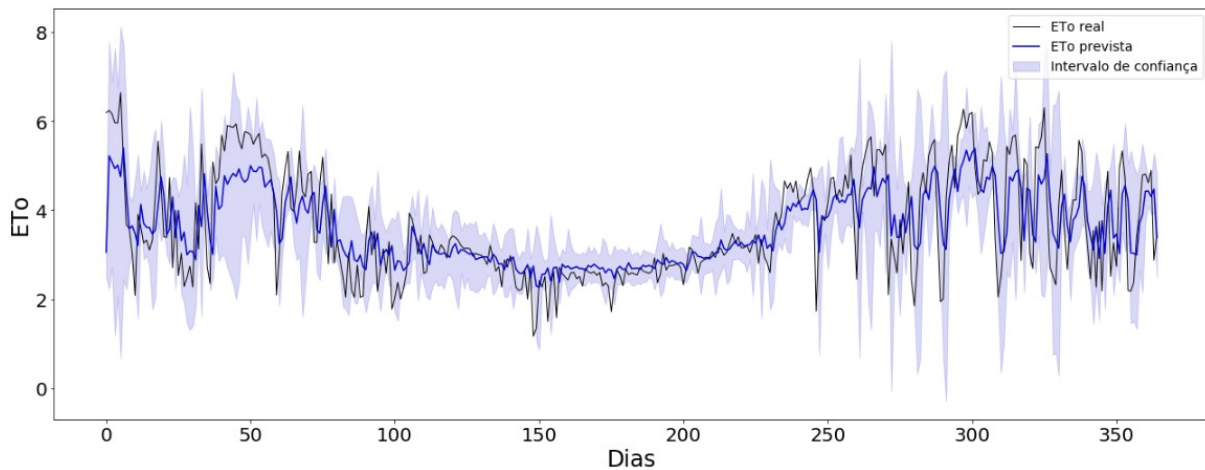


Figura 37 – Desempenho do modelo Ensemble Híbrido para no horizontes de previsão 1 do ano de teste 2013.

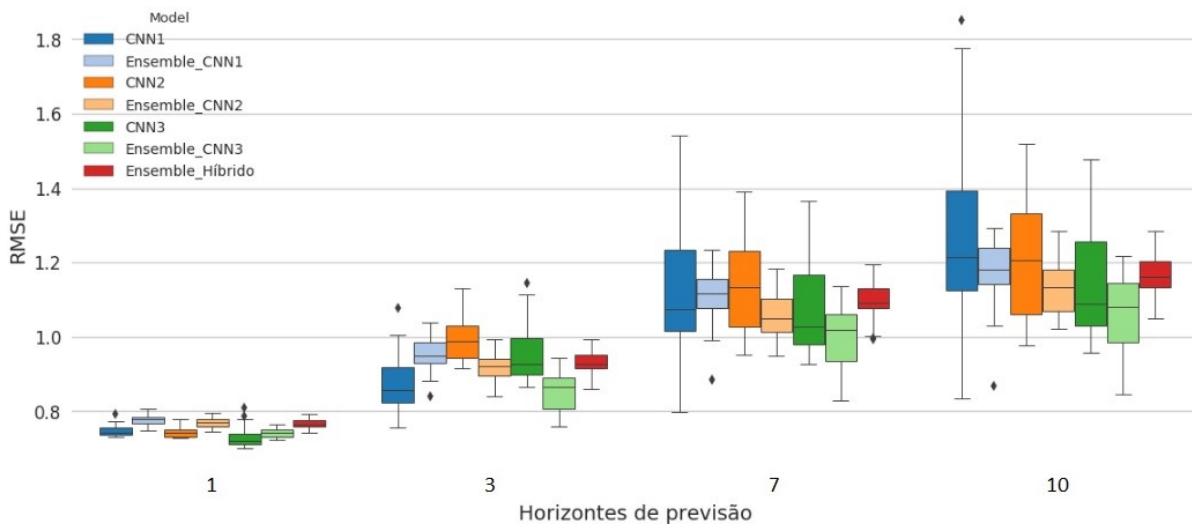


Figura 38 – Boxplot do RMSE dos modelos CNNs e Ensembles para os horizontes de previsão 1, 3, 7 e 10 do ano de 2014.

5.3 Comparação entre os modelos CNN e o método Naïve

Para essa comparação utilizou-se a métrica Scaled RMSE (SRMSE), Equação 5.1, para medir o quanto os modelos CNN e ensembles são melhores que o método Naïve. $SRMSE = 1$, significa que ambos os métodos são iguais, $SRMSE > 1$ o método proposto é pior que Naïve e $SRMSE < 1$ é melhor. Os resultados são apresentados nas Tabelas 26 e 27 para os modelos individuais e ensembles, respectivamente, e mostram que os modelos CNN e ensembles são melhores que o método Naïve em todos os horizontes de previsão de todos os anos de teste, com exceção do ano 2014 que, no horizonte 10, apresentou mesmo

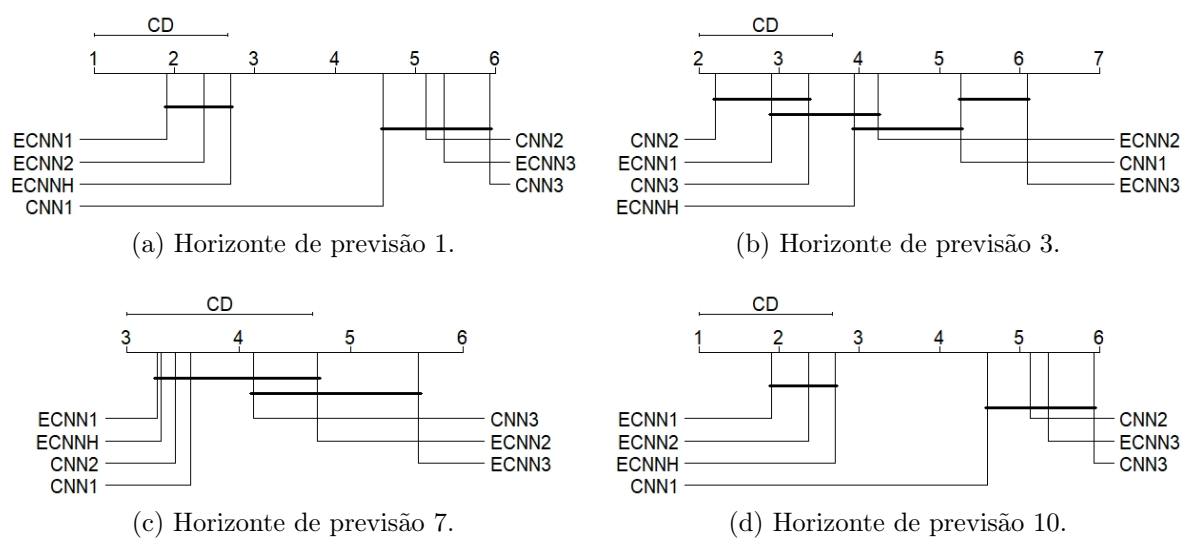


Figura 39 – Visualização para o Teste de Friedman dos resultados dos modelos CNNs e Ensembles do ano de 2014.

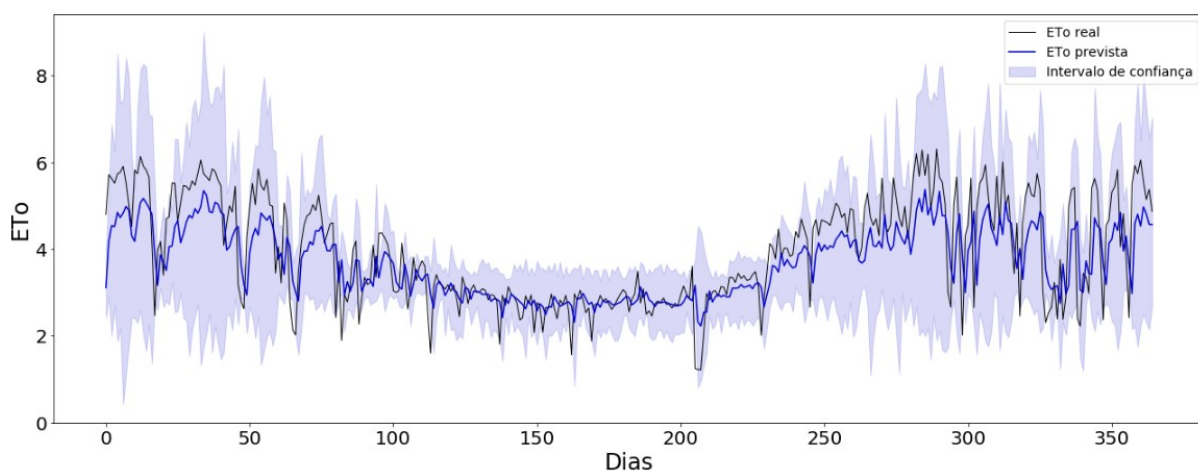


Figura 40 – Desempenho do modelo Ensemble CNN1 para no horizontes de previsão 1 do ano de teste 2014.

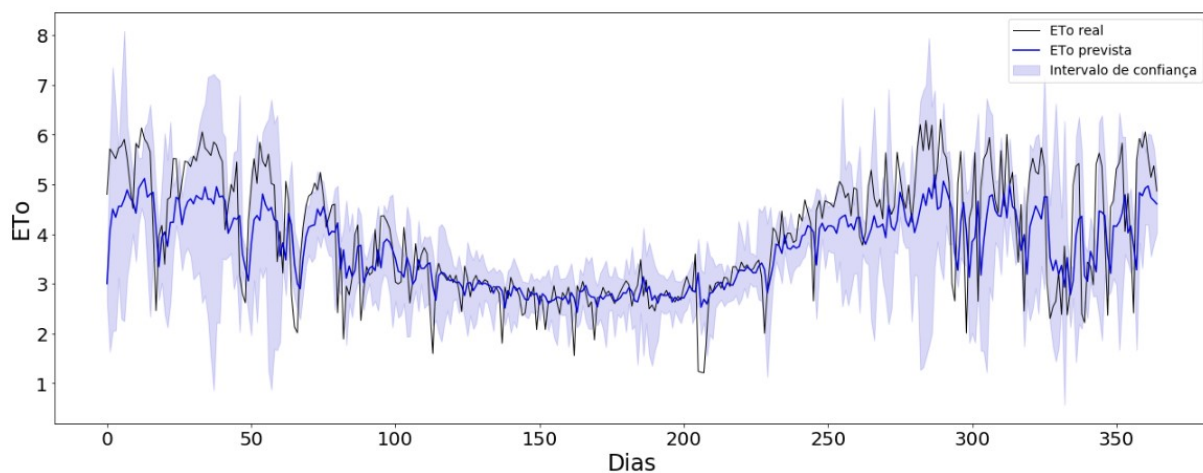


Figura 41 – Desempenho do modelo Ensemble CNN2 no horizontes de previsão 1 do ano de teste 2014.

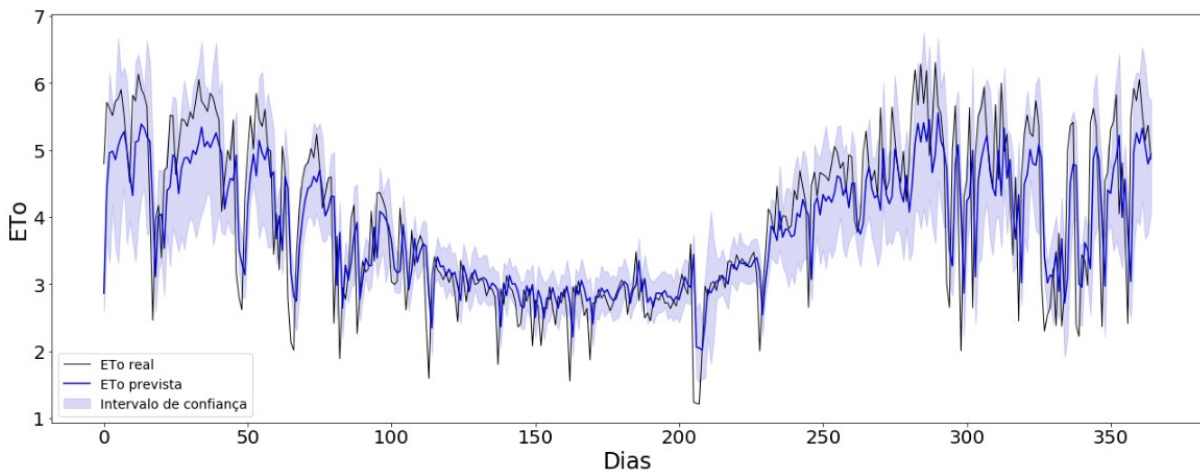


Figura 42 – Desempenho do modelo Ensemble CNN3 para no horizontes de previsão 1 do ano de teste 2014.

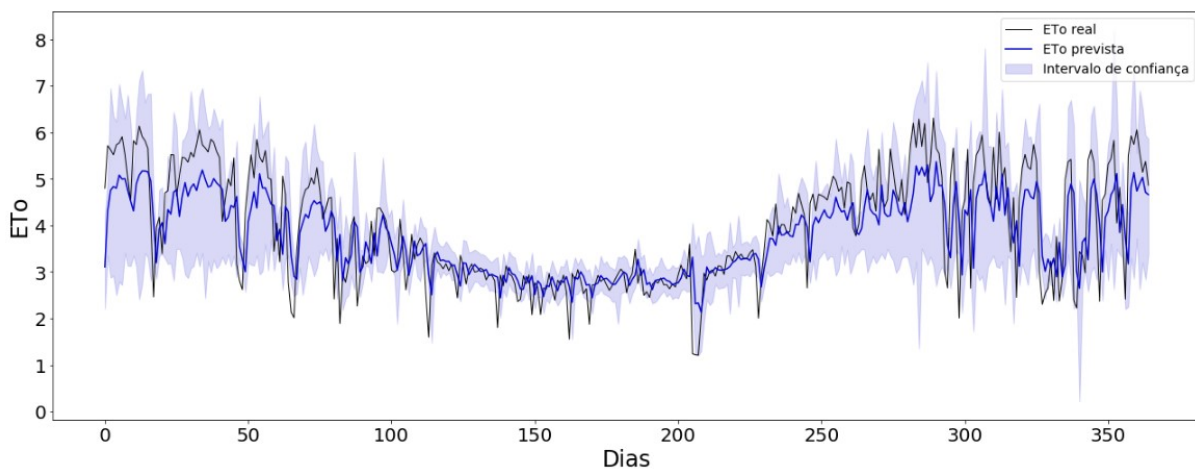


Figura 43 – Desempenho do modelo Ensemble Híbrido para no horizontes de previsão 1 do ano de teste 2014.

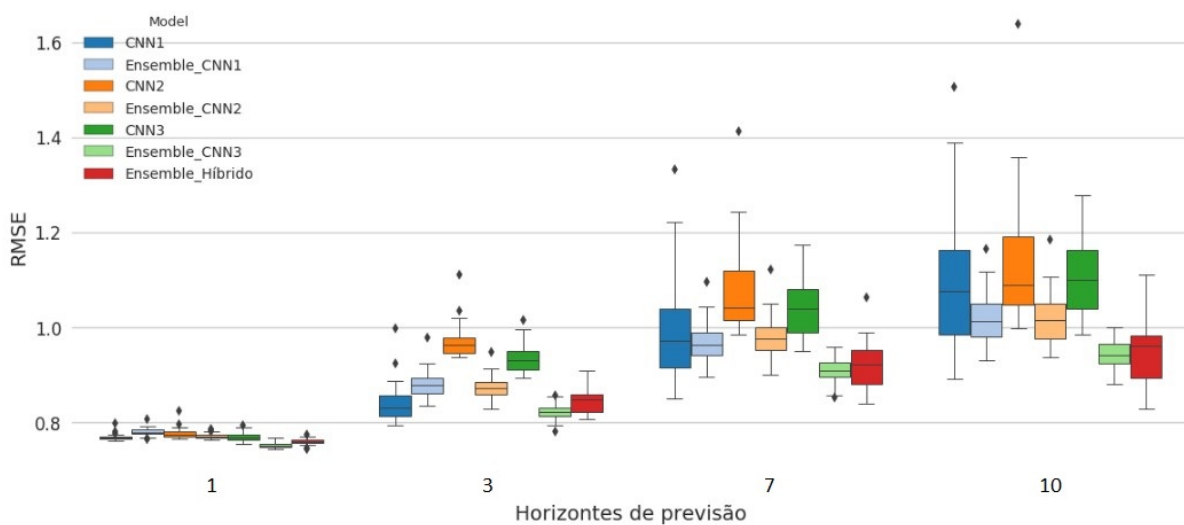


Figura 44 – Boxplot do RMSE dos modelos CNNs e Ensembles para os horizontes de previsão 1, 3, 7 e 10 do ano de 2015.

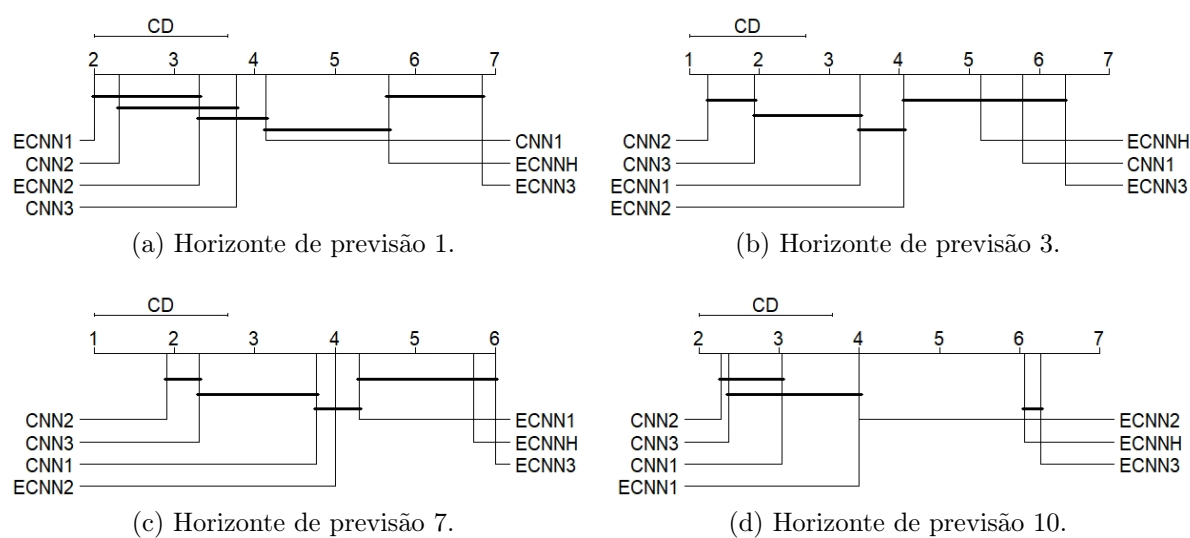


Figura 45 – Visualização para o Teste de Friedman dos resultados dos modelos CNNs e Ensembles do ano de 2015.

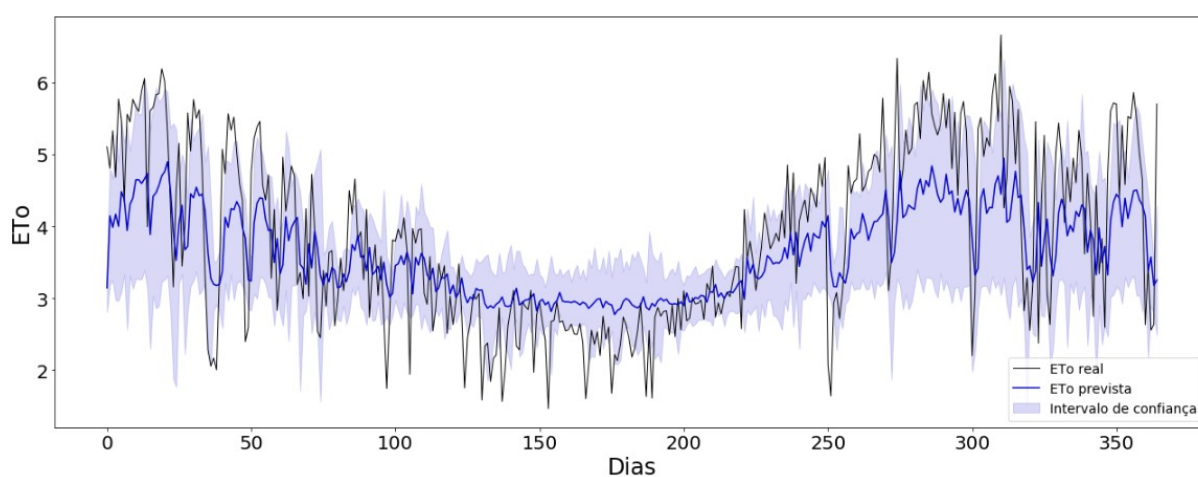


Figura 46 – Desempenho do modelo Ensemble CNN1 para o horizonte de previsão 1 do ano de teste 2015.

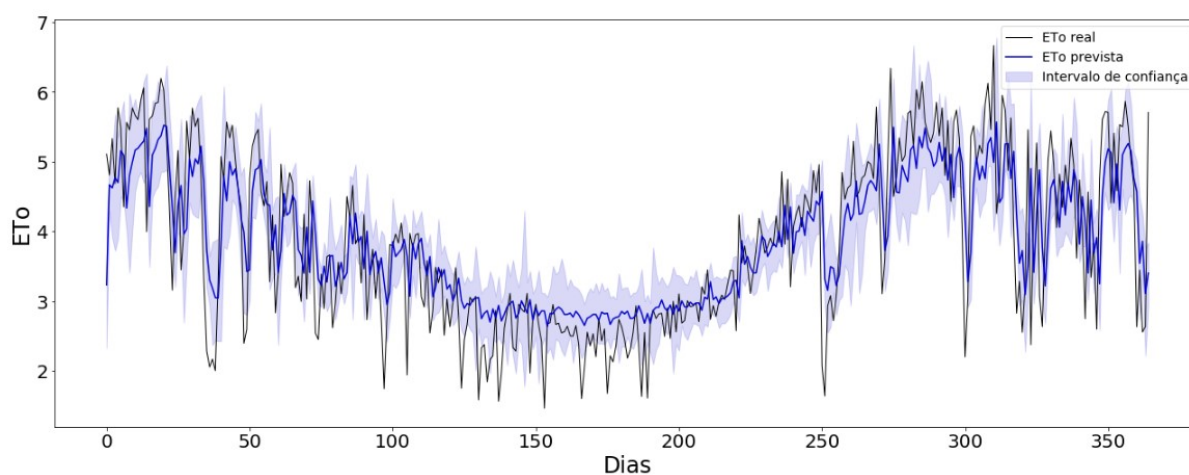


Figura 47 – Desempenho do modelo Ensemble CNN2 para o horizonte de previsão 1 do ano de teste 2015.

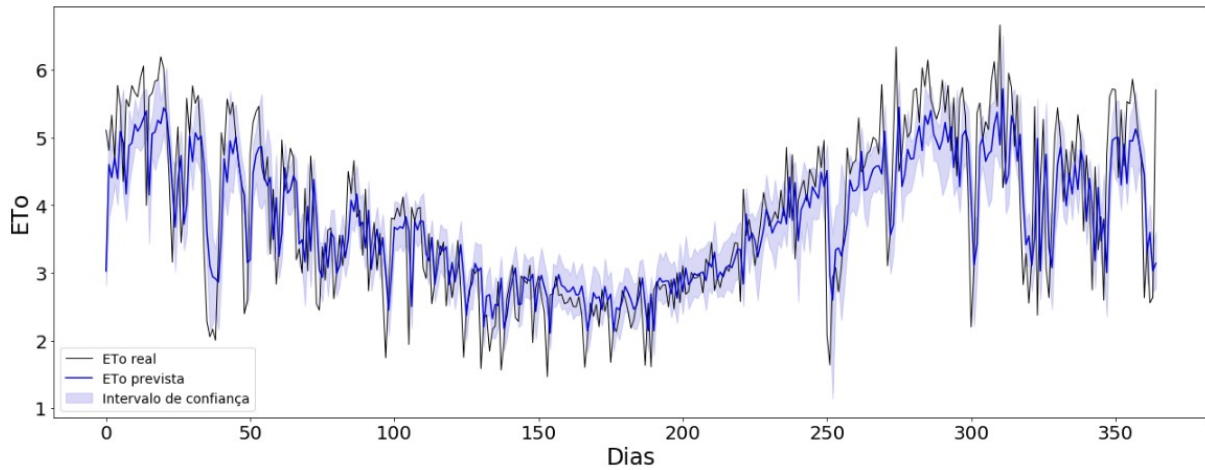


Figura 48 – Desempenho do modelo Ensemble CNN3 para o horizonte de previsão 1 do ano de teste 2015.

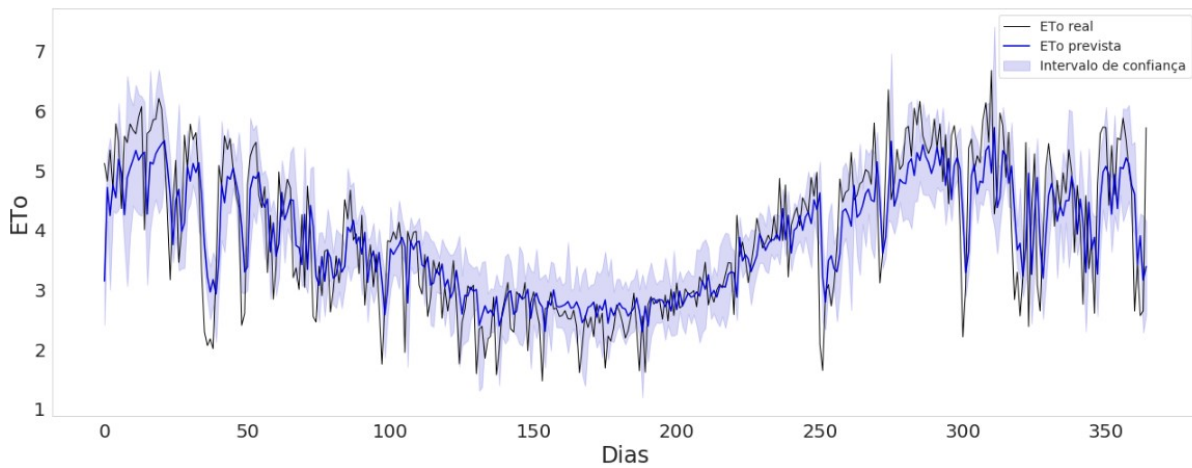


Figura 49 – Desempenho do modelo Ensemble Híbrido para o horizonte de previsão 1 do ano de teste 2015.

desempenho que o método Naïve.

$$SRMSE = \frac{RMSE \text{ do método proposto}}{RMSE \text{ do método Naïve}} \quad (5.1)$$

Para visualizar graficamente esses resultados, a Figura 50 mostra a previsão de 10 passos a frente do modelo Ensemble CNN3 e Naïve do ano de 2015.

5.4 Análise dos resíduos

Os resíduos dos modelos CNN1, CNN2, CNN3, Ensemble CNN1, Ensemble CNN2, Ensemble CNN3 e Ensemble Híbrido são apresentados nas Figuras 51, 52, 53, 54, 55, 56 e 57. As imagens mostram que os resíduos de todos os modelos apresentam média zero e normalmente distribuídos, o que supõe que sejam ruído branco.

Ano	Horizonte de previsão	CNN1	CNN2	CNN3
2012	1	0.89	0.89	0.91
	3	0.81	0.90	0.89
	7	0.89	0.95	0.92
	10	0.95	0.97	0.95
2013	1	0.92	0.94	0.94
	3	0.70	0.85	0.80
	7	0.76	0.86	0.84
	10	0.77	0.87	0.83
2014	1	0.92	0.92	0.90
	3	0.77	0.88	0.64
	7	0.94	0.96	0.91
	10	1.02	0.98	0.93
2015	1	0.91	0.92	0.92
	3	0.78	0.90	0.86
	7	0.82	0.90	0.86
	10	0.90	0.95	0.91

Tabela 26 – Scaled RMSE da previsão de 1, 3, 7 e 10 dias a frente dos modelos CNN1, CNN2 e CNN3 para os anos de 2012, 2013, 2014 e 2015.

Ano	Horizonte de previsão	Ensemble CNN1	Ensemble CNN2	Ensemble CNN3	Ensemble Híbrido
2012	1	0.89	0.93	0.88	0.89
	3	0.84	0.86	0.78	0.82
	7	0.85	0.79	0.80	0.83
	10	0.88	0.86	0.82	0.84
2013	1	0.94	0.94	0.91	0.92
	3	0.71	0.73	0.70	0.74
	7	0.77	0.75	0.75	0.80
	10	0.74	0.71	0.72	0.77
2014	1	0.96	0.96	0.92	0.95
	3	0.84	0.82	0.75	0.82
	7	0.94	0.96	0.84	0.93
	10	0.95	0.92	0.86	0.95
2015	1	0.93	0.92	0.90	0.91
	3	0.82	0.81	0.73	0.78
	7	0.80	0.80	0.75	0.76
	10	0.84	0.85	0.78	0.78

Tabela 27 – Scaled RMSE da previsão de 1, 3, 7 e 10 dias a frente dos modelos Ensemble CNN1, Ensemble CNN2, Ensemble CNN3 e Ensemble híbrido para os anos de 2012, 2013, 2014 e 2015.

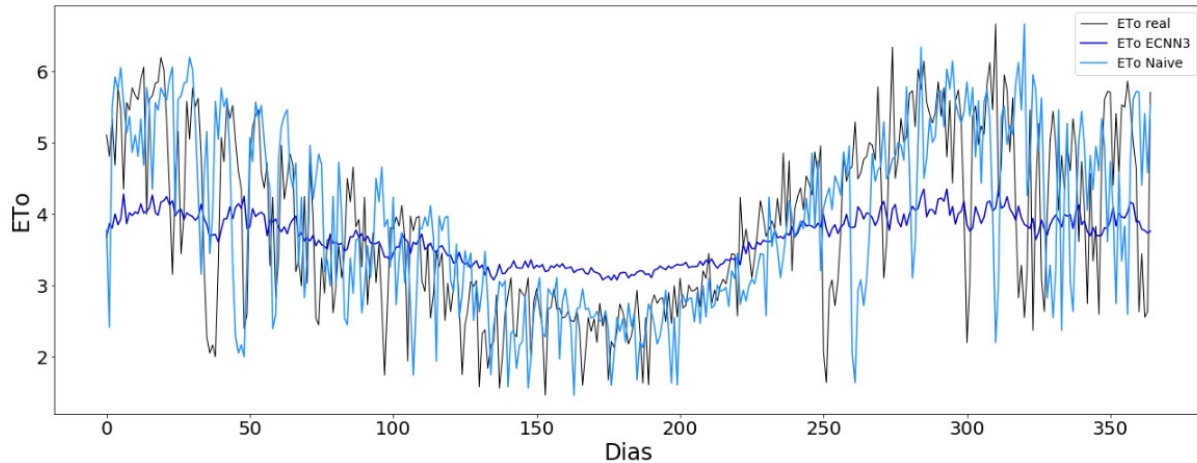


Figura 50 – Desempenho do modelo Ensemble CNN3 e do método Naïve para o horizontes de previsão 10 do ano de teste 2015.

Lag	Estatística	p-Valor	Valor Crítico	Resultado
1	336.885	0.00	2.706	Rejeita H0
2	673.134	0.00	4.605	Rejeita H0
3	999.494	0.00	6.251	Rejeita H0

Tabela 28 – Teste Ljung-Box para os resíduos do modelo CNN1.

Lag	Estatística	p-Valor	Valor Crítico	Resultado
1	337.950	0.00	2.706	Rejeita H0
2	675.429	0.00	4.605	Rejeita H0
3	1006.779	0.00	6.251	Rejeita H0

Tabela 29 – Teste Ljung-Box para os resíduos do modelo CNN2.

Também realizou-se o Teste de Ljung-Box que verifica, para cada atraso da função de autocorrelação dos resíduos, as hipóteses H_0 - os resíduos são i.i.d (independentes e identicamente distribuídos) e H_1 - os resíduos não são i.i.d. Para rejeitar H_0 , a estatística de teste Q deve atender à condição $Q > X_{1-\alpha, df}^2$ em que α é o nível de confiança e df é o número do atraso. As Tabelas 28, 29, 30, 31, 32, 33 e 34 apresentam os resultados do Teste de Ljung-Bom para os 3 primeiros atrasos dos modelos CNN1, CNN2, CNN3, Ensemble CNN1, Ensemble CNN2, Ensemble CNN3 e Ensemble Híbrido. Percebeu-se que, devido a existência de correlações entre os resíduos em todos os modelos, estes podem ser ainda aprimorados.

Lag	Estatística	p-Valor	Valor Crítico	Resultado
1	350.375	0.00	2.706	Rejeita H0
2	697.229	0.00	4.605	Rejeita H0
3	1034.990	0.00	6.251	Rejeita H0

Tabela 30 – Teste Ljung-Box para os resíduos do modelo CNN3.

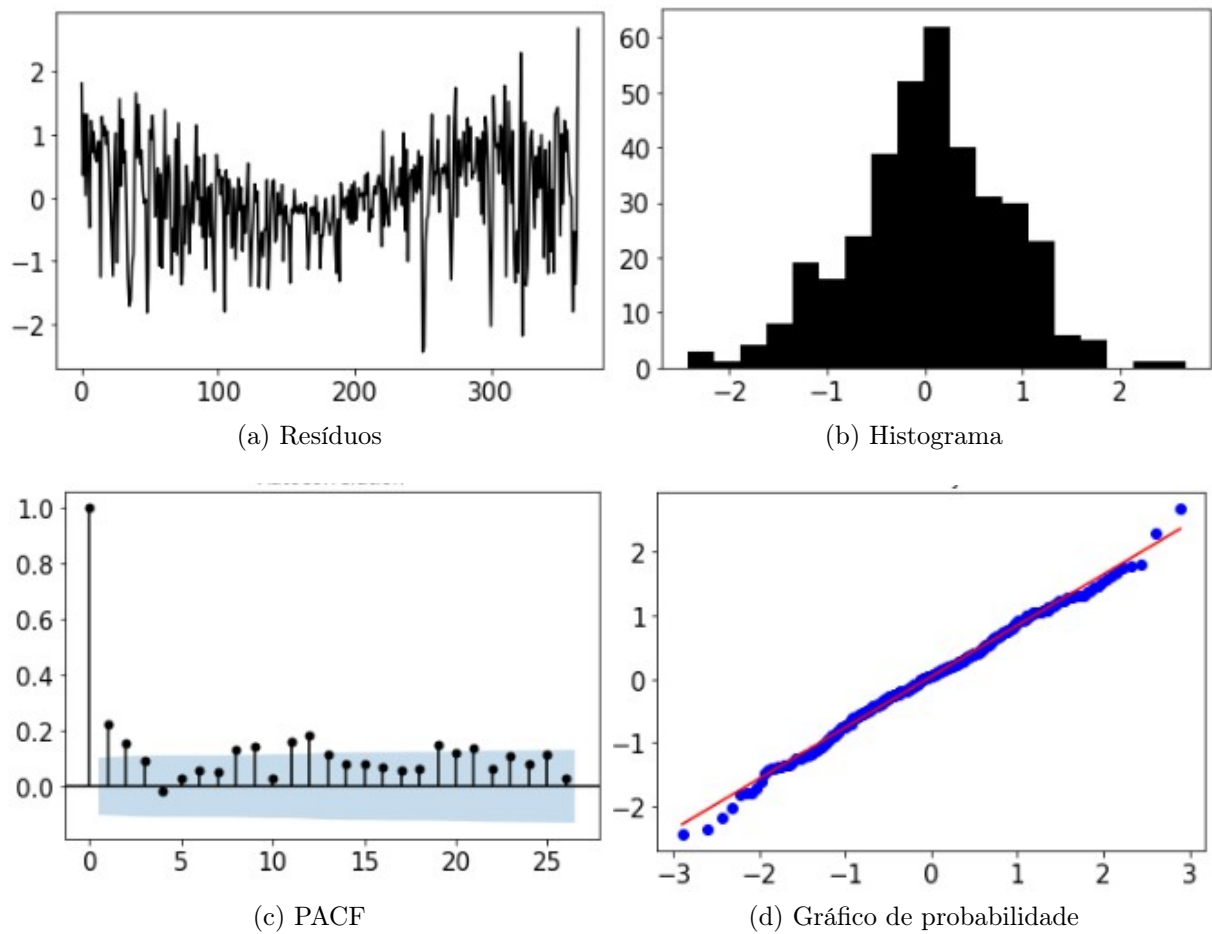


Figura 51 – Resíduos do modelo CNN1.

Lag	Estatística	p-Valor	Valor Crítico	Resultado
1	338.375	0.00	2.706	Rejeita H0
2	676.527	0.00	4.605	Rejeita H0
3	1006.651	0.00	6.251	Rejeita H0

Tabela 31 – Teste Ljung-Box para os resíduos do modelo Ensemble CNN1.

Lag	Estatística	p-Valor	Valor Crítico	Resultado
1	340.616	0.00	2.706	Rejeita H0
2	680.804	0.00	4.605	Rejeita H0
3	1013.667	0.00	6.251	Rejeita H0

Tabela 32 – Teste Ljung-Box para os resíduos do modelo Ensemble CNN2.

Lag	Estatística	p-Valor	Valor Crítico	Resultado
1	344.433	0.00	2.706	Rejeita H0
2	688.042	0.00	4.605	Rejeita H0
3	1025.161	0.00	6.251	Rejeita H0

Tabela 33 – Teste Ljung-Box para os resíduos do modelo Ensemble CNN3.

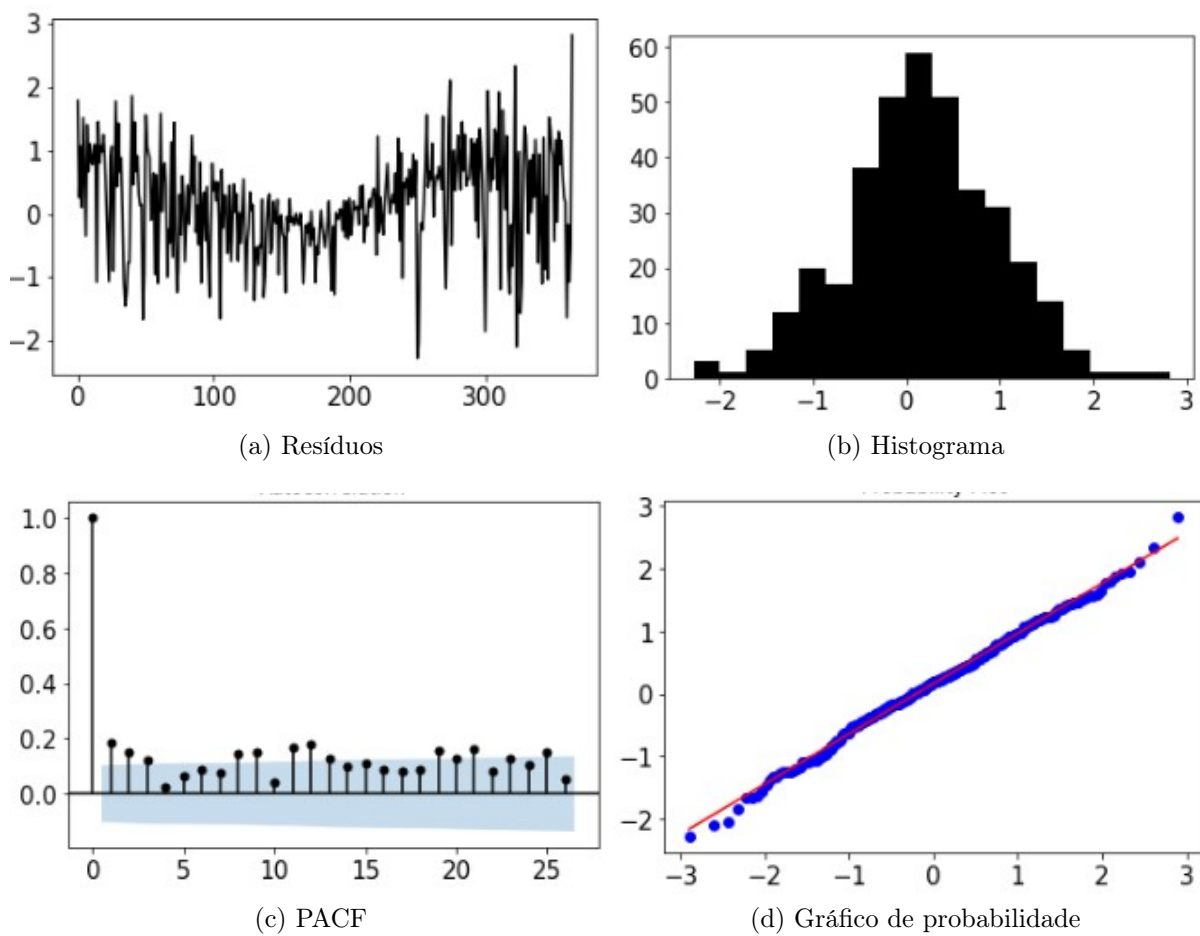


Figura 52 – Resíduos do modelo CNN2.

Lag	Estatística	p-Valor	Valor Crítico	Resultado
1	335.785	0.00	2.706	Rejeita H0
2	671,138	0.00	4.605	Rejeita H0
3	999.764	0.00	6.251	Rejeita H0

Tabela 34 – Teste Ljung-Box para os resíduos do modelo Ensemble CNN Híbrido.

5.5 Resumo

Neste capítulo foram apresentados os resultados e discussões dos experimentos realizados com os modelos CNN, LSTM, ARIMA, WFTS, Naïve e modelos *ensemble* na previsão de séries temporais de evapotranspiração de referência. No próximo capítulo, serão descritas as conclusões encontradas após desenvolvimento deste trabalho, bem como os trabalhos futuros que podem vir a ser desenvolvidos a partir dessas conclusões.

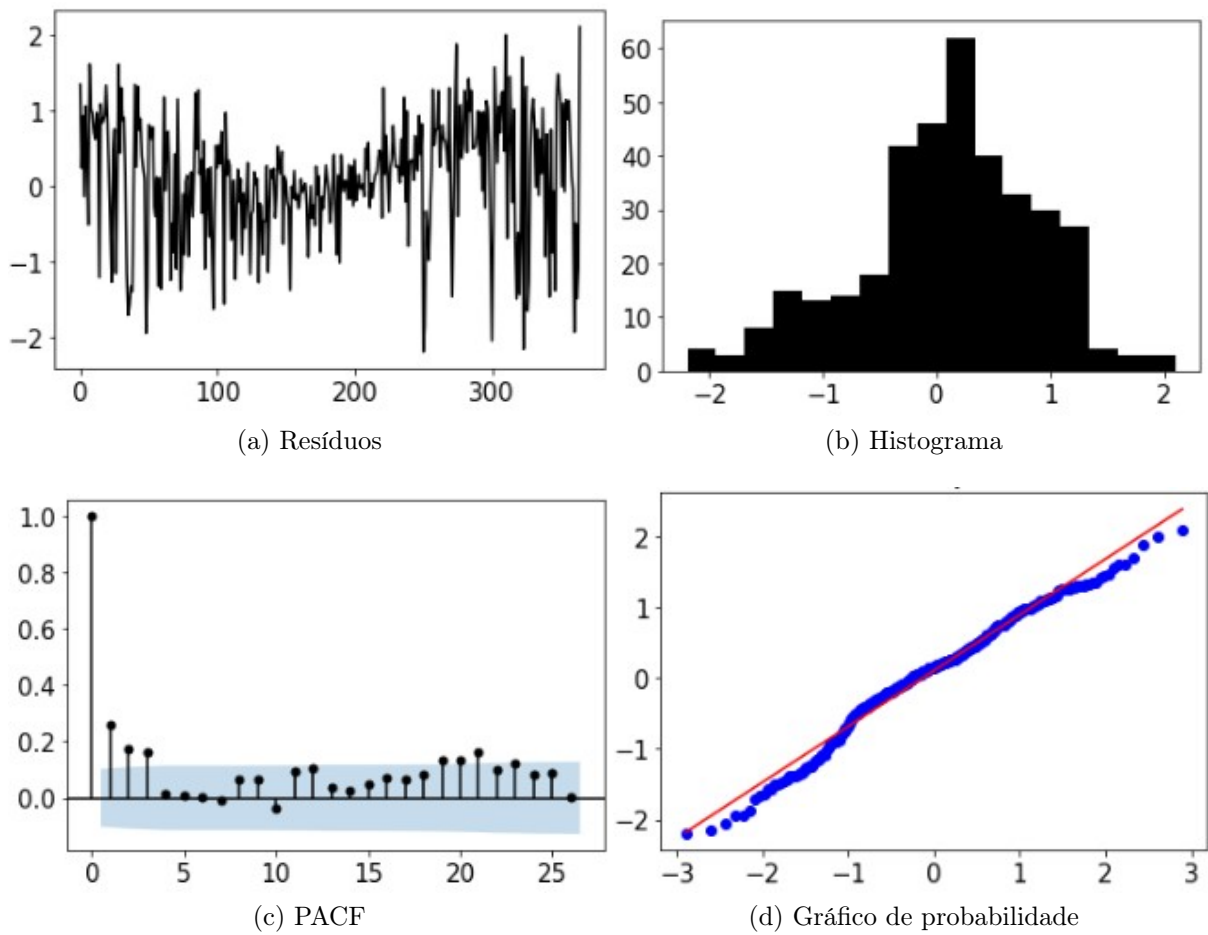


Figura 53 – Resíduos do modelo CNN3.

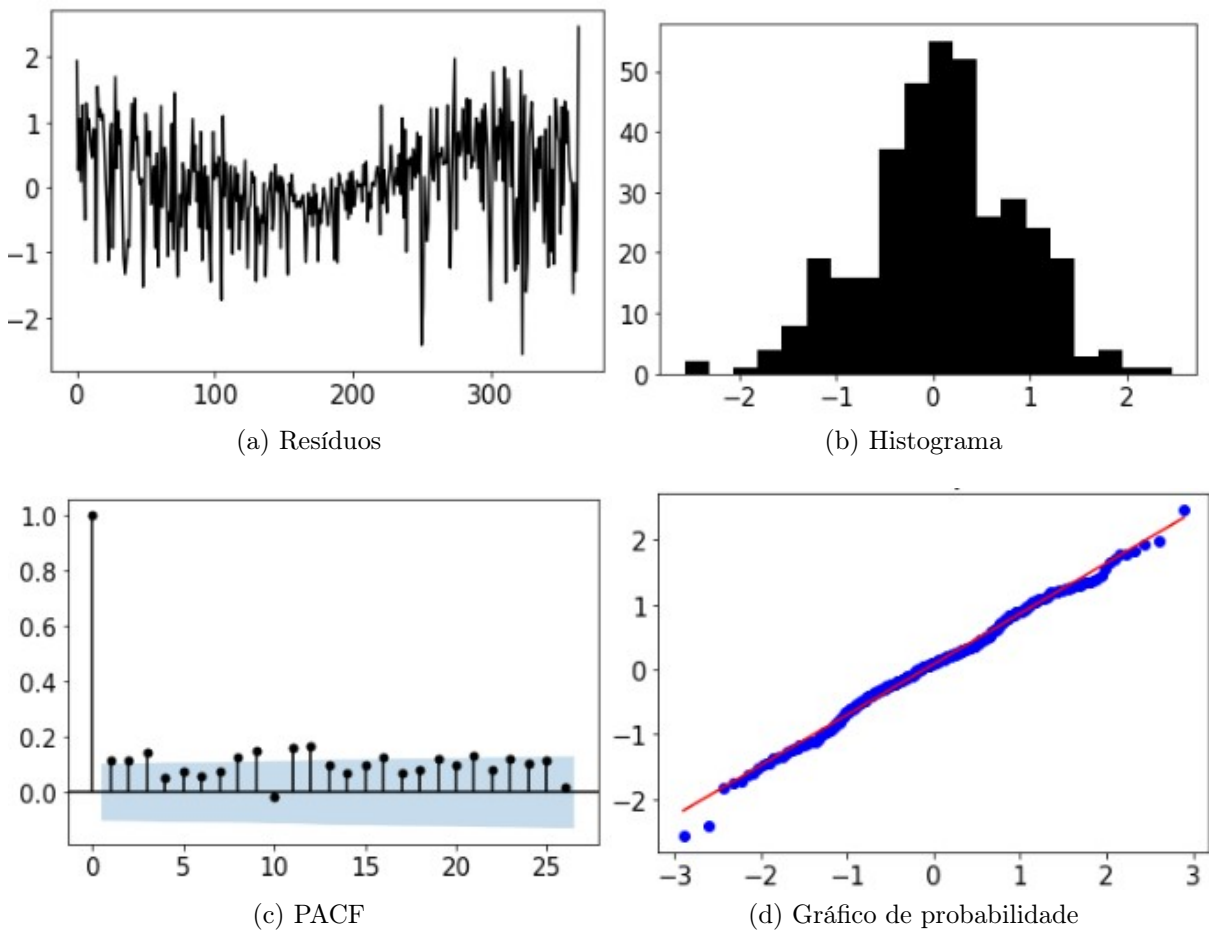


Figura 54 – Resíduos do modelo Ensemble CNN1.

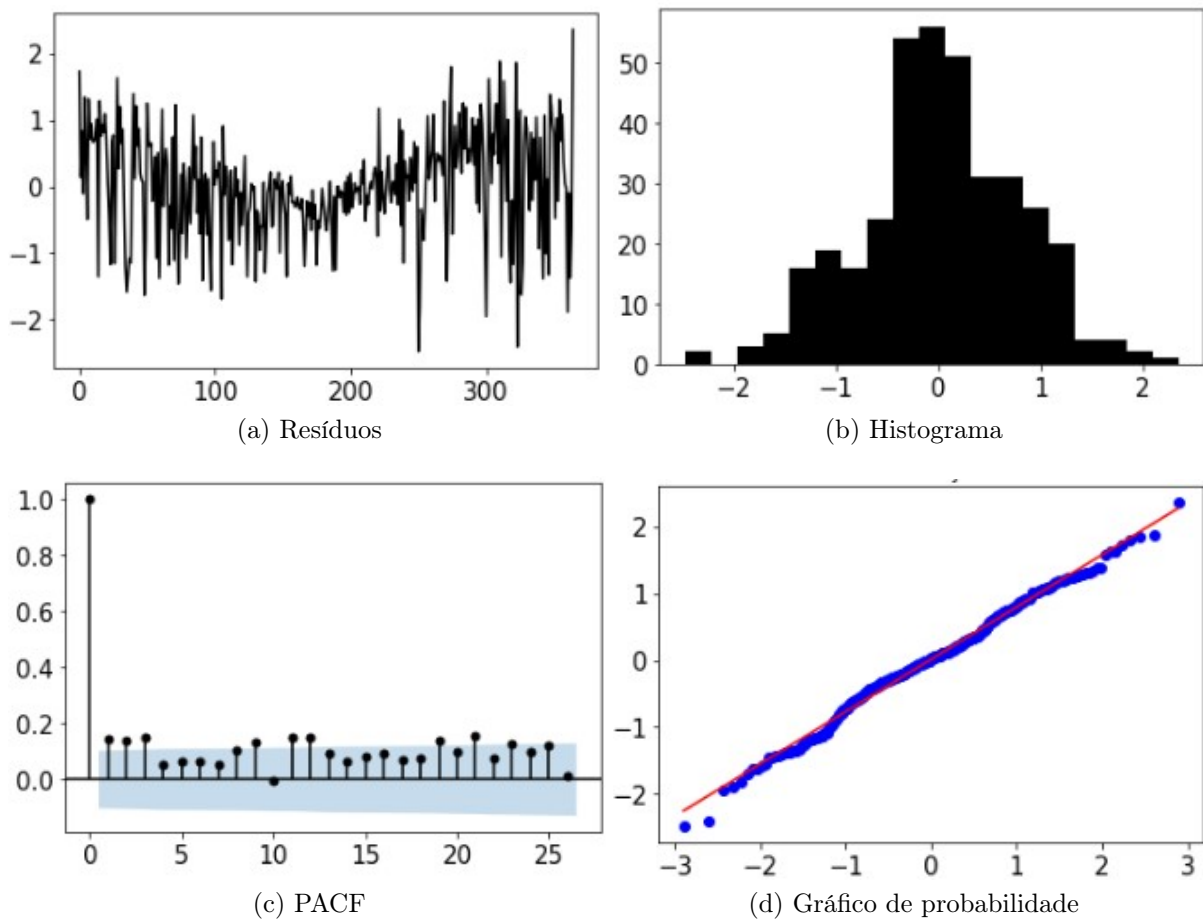


Figura 55 – Resíduos do modelo Ensemble CNN2.

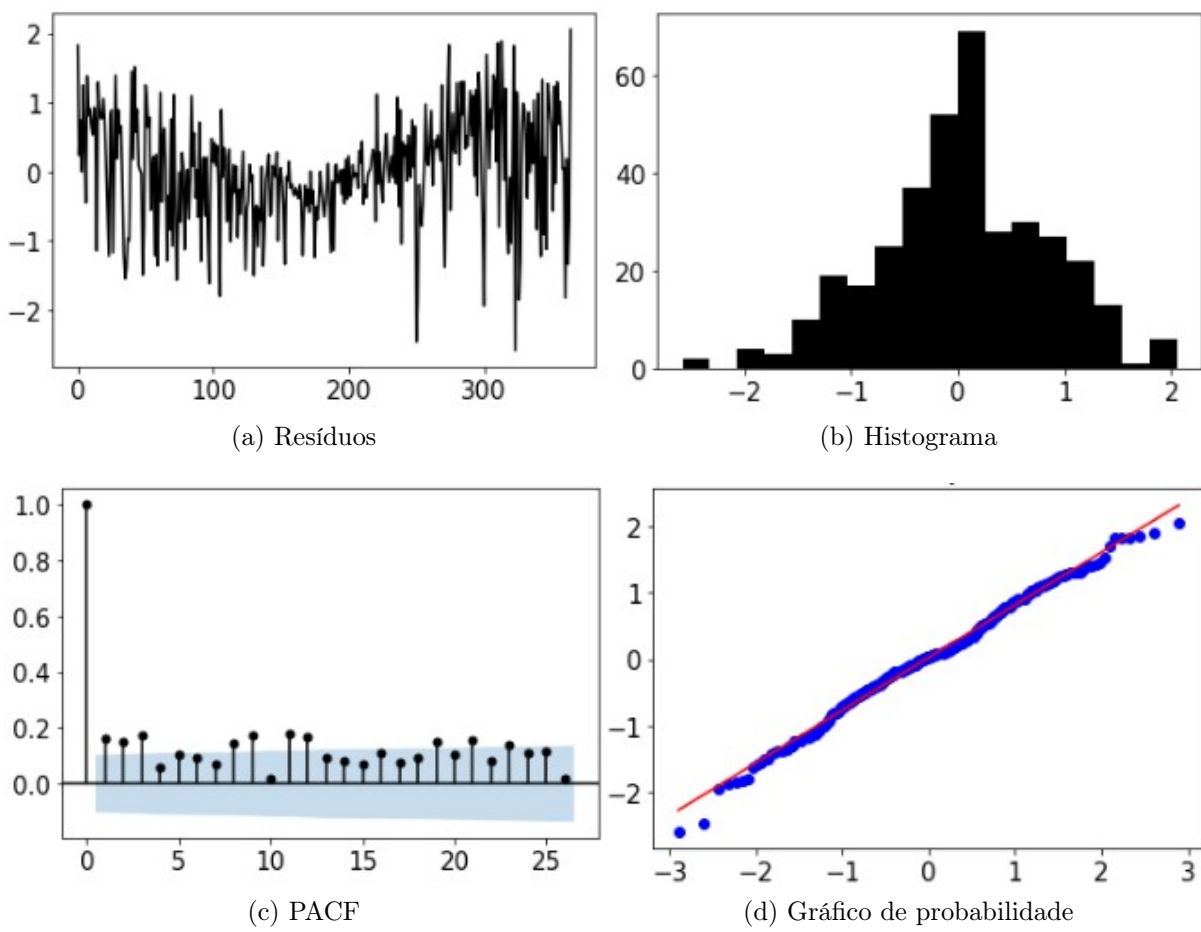


Figura 56 – Resíduos do modelo Ensemble CNN3.

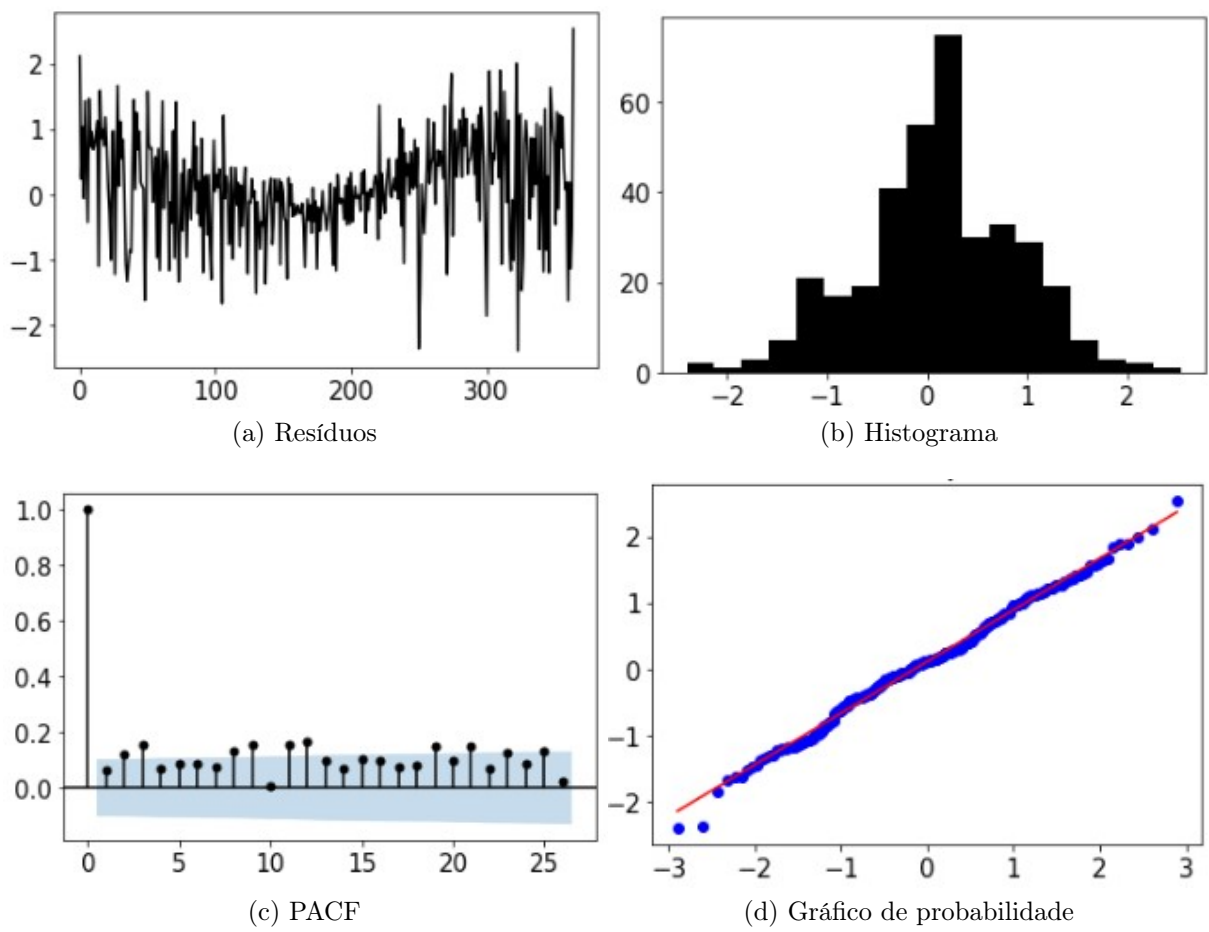


Figura 57 – Resíduos do modelo Ensemble CNN Híbrido.

Capítulo 6

Conclusão

Nesse trabalho, foi realizada uma análise do problema de previsão diária de ET_o usando redes neurais convolucionais. A previsão da ET_o é essencial para o planejamento de irrigação e gerenciamento dos recursos hídricos, especialmente na atual conjuntura onde a escassez desses recursos está se tornando uma questão importante em todo o mundo. O crescimento da população e mudanças climáticas estão fazendo com que o setor agrícola busque abordagens mais precisas e eficientes para garantir um suprimento adequado e regular de alimentos para a sociedade com menor consumo de água.

No capítulo 2, foram apresentados os conceitos que envolvem o processo de evapotranspiração de referência. Essa variável é afetada pelos parâmetros climáticos radiação solar, temperatura do ar, umidade e velocidade do vento e pode ser estimada a partir da equação Penman-Monteith parametrizada segundo as características da superfície de referência. Através dessa equação gerou-se a série temporal de ET_o para realizar os experimentos deste trabalho.

No capítulo 3, foram descritas arquiteturas de redes neurais convolucionais. Inicialmente apresentou-se a LeNet-5, rede desenvolvida para reconhecimento de caracteres manuscritos e impressos. Apesar de serem conhecidas pelos ótimos resultados na área de visão computacional, as redes CNN têm tido muito sucesso em um vasto número de domínios que dependem de dados sequenciais, como é o caso das TCNs, descritas também neste capítulo.

Todos os elementos dessas redes foram usados para construir 3 arquiteturas diferentes de forma que se pudesse avaliar o desempenho de CNNs e TCNs na previsão de séries temporais e ET_o . As CNNs implementadas possuíam as seguintes características: CNN1 – com camadas convolucionais simples e camadas de agrupamento; CNN2 – camadas convolucionais causais dilatadas e CNN3 – camadas convolucionais causais dilatadas com o uso de blocos residuais.

Outro conceito importante descrito neste capítulo foi o *ensemble learning* com

previsões probabilísticas usando o método KDE e como esses modelos podem melhorar as previsões em termos de precisão e variância. Para verificar essas melhorias, foram implementados 4 *ensembles* a partir da combinação de modelos CNN1, CNN2 e CNN3 com variação tanto dos hiperparâmetros como dos modelos.

O capítulo 4 apresentou o algoritmo genético desenvolvido para otimizar os hiperparâmetros das redes CNN1, CNN2 e CNN3, buscando minimizar dois objetivos, precisão e parcimônia. Resultados mostraram que os modelos CNNs de alta complexidade apresentaram desempenhos parecidos com modelos CNNs menos complexos, sendo possível concluir que o uso do AG com objetivo de encontrar um equilíbrio entre precisão e parcimônia é muito útil para modelos *deep learning*. Este capítulo apresentou também os modelos usados para validar os modelos CNN e como os hiperparâmetros desses modelos foram otimizados.

No capítulo 5, foram apresentados os resultados dos experimentos feitos em 4 anos de teste para horizontes de previsão 1, 3, 7 e 10. Num primeiro momento, com o objetivo de validar os modelos CNN, estes foram comparados aos modelos ARIMA, WFTS e LSTM. Os resultados mostraram que todos os modelos CNN são viáveis para previsão de séries temporais de ET_o e que obtiveram desempenho bem próximos aos modelos ARIMA e WFTS. Verificou-se também que o modelo CNN1 apresentou melhores resultados que os modelos CNN temporais (CNN2 e CNN3) e LSTM, modelos estes, desenvolvidos especialmente para aplicações que dependem de dados sequenciais.

Em outro momento, comparou-se os modelos CNNs aos modelos *ensemble* e concluiu-se que:

- Os modelos *ensemble* (*Ensemble* CNN1, *Ensemble* CNN2, *Ensemble* CNN3 e *Ensemble* Híbrido) apresentaram de forma geral menor variância que os modelos CNNs individuais.
- Em termos de precisão, nem todos os modelos *ensemble* apresentaram melhoria. Concluiu-se que, apenas os modelos CNN2 e CNN3, especializados em dados sequenciais, conseguiram diminuir o erro quando trabalharam em conjunto.
- É possível medir a incerteza dos modelos *ensemble*, já que suas saídas são distribuições de probabilidade e não uma previsão pontual, como nos modelos individuais. Concluiu-se que o desempenho dos modelos *ensemble* foi satisfatório, visto que os valores da série temporal de ET_o ficaram dentro dos limites da incerteza desses modelos, sendo o *Ensemble* CNN3, o modelo que apresentou menor incerteza.
- Além das melhorias apresentadas nos resultados dos modelos *ensemble*, pode-se concluir também que o método de aprendizagem em conjunto é computacionalmente

mais barato e mais simples de desenvolver, pois elimina a etapa de otimização de hiperparâmetros.

Por último, concluiu-se que todos os modelos CNN individuais e ensembles são melhores que o método Naïve.

6.1 Trabalhos futuros

Apresenta-se como sugestões para trabalhos futuros os seguintes pontos:

1. Utilizar os métodos aqui apresentados para fazer previsões de ET_o :
 - com dados de outras estações meteorológicas com condições climáticas diferentes.
 - de forma indireta e comparar com os resultados deste trabalho.
 - com uma série temporal multivariada e comparar com os resultados deste trabalho.
2. Utilizar o modelo *ensemble* CNN3 para prever séries temporais de ET_o multivariadas compostas pelas temperaturas máxima e mínima. Este estudo é interessante para verificar se é possível prever a ET_o satisfatoriamente em locais onde não existem estações meteorológicas e que sensores de temperatura simples podem realizar as medições.
3. Utilizar algum método de clusterização para definir as zonas homogêneas de ET_o do norte de Minas Gerais com dados das séries temporais de todas as estações meteorológicas localizadas na região.
4. Fazer a previsão de ET_o de cada zona homogênea utilizando dados das estações meteorológicas que nelas estão localizadas com o modelos *ensemble* CNN3.

Referências

- [Adhikari and Verma 2016] Adhikari, R. and Verma, G. (2016). Time Series Forecasting Through a Dynamic Weighted Ensemble Approach. In *3rd International Conference on Advanced Computing, Networking and Informatics, Smart Innovation, Systems and Technologies (ICACNI)*, pages 455–465, Orissa, India. Springer India.
- [Allen et al. 2006] Allen, R. G., Pereira, L. S., and Raes, D. (2006). Evapotranspiración del cultivo. Guías para la determinación de los requerimientos de agua de los cultivos. Technical report, FAO, Roma, Itália.
- [Allende and Valle 2017] Allende, H. and Valle, C. (2017). Ensemble Methods for Time Series Forecasting. In *Studies in Fuzziness and Soft Computing*, pages 217–232. Springer, Cham, 349 edition.
- [Anjitha 2019] Anjitha, K. P. (2019). Evapotranspiration and agriculture-A review. *Agricultural Reviews*, 40(1):1–11.
- [Bachour et al. 2016] Bachour, R., Maslova, I., Ticlavilca, A. M., Walker, W. R., and McKee, M. (2016). Wavelet-multivariate relevance vector machine hybrid model for forecasting daily evapotranspiration. *Stochastic Environmental Research and Risk Assessment*, 30(1):103–117.
- [Bai et al. 2018a] Bai, S., Kolter, J. Z., and Koltun, V. (2018a). Convolutional Sequence Modeling Revisited. In *International Conference on Learning Representations*.
- [Bai et al. 2018b] Bai, S., Zico Kolter, J., and Koltun, V. (2018b). An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *CoRR*, abs/1803.01271.
- [Bengio 2009] Bengio, Y. (2009). Learning Deep Architectures for AI. *Found. Trends Mach. Learn.*, 2(1):1–127.
- [Bernardo et al. 2019] Bernardo, S., Soares, A. A., and Mantovani, E. C. (2019). *Manual de irrigação*. Editora UFV, Viçosa - MG, 9ª edition.

- [Borovykh et al. 2017] Borovykh, A., Bohte, S., and Oosterlee, C. W. (2017). Conditional Time Series Forecasting with Convolutional Neural Networks. In *CWI, Amsterdam, The Netherlands*.
- [Cândido De Lima E Silva et al. 2017] Cândido De Lima E Silva, P., Alves, M. A., Severiano Junior, C. A., Vieira, G. L., Guimarães, F. G., and Sadaei, H. J. (2017). Probabilistic Forecasting with Seasonal Ensemble Fuzzy Time-Series. In *Anais do 13 Congresso Brasileiro de Inteligência Computacional*, page 12, Curitiba - PR. ABRICOM.
- [Chen 2017] Chen, Y.-C. (2017). A Tutorial on Kernel Density Estimation and Recent Advances. *Biostatistics & Epidemiology*, 1(1):161–187.
- [Coopersmith et al. 2014] Coopersmith, E. J., Minsker, B. S., Wenzel, C. E., and Gilmore, B. J. (2014). Machine learning assessments of soil drying for agricultural planning. *Computers and Electronics in Agriculture*, 104:93–104.
- [Du et al. 2017] Du, S., Pandey, M., and Xing, C. (2017). Modeling approaches for time series forecasting and anomaly detection. Technical report.
- [Food and Agriculture Organization of the United Nations 2017] Food and Agriculture Organization of the United Nations (2017). Water for Sustainable Food and Agriculture. A report produced for the G20 Presidency of Germany. Technical report, Roma.
- [Frizzone et al. 2012] Frizzone, J. A., Freitas, P. S. L., Rezende, R., and Faria, M. A. (2012). *Microirrigação - Gotejamento e microaspersão*. Editora Eduem, 1^a edition.
- [Gehring et al. 2017] Gehring, J., Auli, M., Grangier, D., Yarats, D., and Dauphin, Y. N. (2017). Convolutional Sequence to Sequence Learning. *CoRR*, abs/1705.03122.
- [Gensler et al. 2017] Gensler, A., Henze, J., Sick, B., and Raabe, N. (2017). Deep Learning for solar power forecasting - An approach using AutoEncoder and LSTM Neural Networks. In *2016 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2016 - Conference Proceedings*, pages 2858–2865. Institute of Electrical and Electronics Engineers Inc.
- [Girshick et al. 2014] Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA. IEEE.
- [Gu et al. 2018] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., and Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, 77:354–377.

- [Guney et al. 2018] Guney, H., Bakir, M. A., and Aladag, C. H. (2018). A Novel Stochastic Seasonal Fuzzy Time Series Forecasting Model. *International Journal of Fuzzy Systems*, 20(3):729–740.
- [Guo et al. 2016] Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., and Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, 187:27–48.
- [Hansen and Salamon 1990] Hansen, L. and Salamon, P. (1990). Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(10):993–1001.
- [He et al. 2016] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778.
- [Hochreiter and Schmidhuber 1997] Hochreiter, S. and Schmidhuber, U. J. J. (1997). Long Short-Term Memory. Technical Report 8.
- [Kalchbrenner et al. 2017] Kalchbrenner, N., Espeholt, L., Simonyan, K., Van Den Oord, A., Graves, A., Kavukcuoglu, K., Deepmind, G., and Uk, L. (2017). Neural Machine Translation in Linear Time. *ArXiv*, abs/1610.10099:9.
- [Kamilaris and Prenafeta-Boldú 2018] Kamilaris, A. and Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147:70–90.
- [Krizhevsky et al. 2012] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems*, pages 1097–1105. Curran Associates Inc., Lake Tahoe, Nevada, 1 edition.
- [Kumar and Raj 2017] Kumar, A. and Raj, B. (2017). Deep CNN Framework for Audio Event Recognition using Weakly Labeled Web Data. *ArXiv*, abs/1707.02530:5.
- [Landeras et al. 2009] Landeras, G., Ortiz-Barredo, A., and López, J. J. (2009). Forecasting Weekly Evapotranspiration with ARIMA and Artificial Neural Network Models. *Journal of Irrigation and Drainage Engineering*, 135(3):323–334.
- [Längkvist et al. 2014] Längkvist, M., Karlsson, L., and Loutfi, A. (2014). A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters*, 42:11–24.
- [Lea et al. 2016] Lea, C., Flynn, M. D., Vidal, R., Reiter, A., and Hager, G. D. (2016). Temporal Convolutional Networks for Action Segmentation and Detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1003–1012. IEEE.

- [LeCun and Bengio 1998] LeCun, Y. and Bengio, Y. (1998). *Convolutional Networks for Images, Speech, and Times-Series*. MIT Press.
- [Lecun et al. 2015] Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521:436–444.
- [LeCun et al. 1990] LeCun, Y., Boser, B. E., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. E., and Jackel, L. D. (1990). *Handwritten Digit Recognition with a Back-Propagation Network*. Morgan-Kaufmann.
- [Lecun et al. 1998] Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.
- [Liakos et al. 2018] Liakos, K., Busato, P., Moshou, D., Pearson, S., and Bochtis, D. (2018). Machine Learning in Agriculture: A Review. *Sensors*, 18(8):2674.
- [Lucas et al. 2019] Lucas, P. d. O. e., Silva, P. C. d. L., and Guimarães, F. G. (2019). Otimização Evolutiva de Hiperparâmetros para Modelos de Séries Temporais Nebulosas. In *Anais do 14º Simpósio Brasileiro de Automação Inteligente*, pages 1 – 7, Ouro Preto.
- [Manikumari N et al. 2017] Manikumari N, Murugappan A, and Vinodhini G (2017). Time Series Forecasting of Daily Reference Evapotranspiration by Neural Network Ensemble Learning for Irrigation System. *IOP Conference Series: Earth and Environmental Science*, 80(1):012069.
- [Mariño et al. 1993] Mariño, M. A., Tracy, J. C., and Taghavi, S. (1993). Forecasting of reference crop evapotranspiration. *Agricultural Water Management*, 24(3):163–187.
- [Matthews et al. 2017] Matthews, S. G., Miller, A. L., Plötz, T., and Kyriazakis, I. (2017). Automated tracking to measure behavioural changes in pigs for health and welfare monitoring. *Scientific Reports*, 7(1):17582.
- [Mirzaei Talarposhti et al. 2016] Mirzaei Talarposhti, F., Javedani Sadaei, H., Enayatifar, R., Gadelha Guimarães, F., Mahmud, M., and Eslami, T. (2016). Stock market forecasting by using a hybrid model of exponential fuzzy time series. *International Journal of Approximate Reasoning*, 70:79–98.
- [Moshou et al. 2004] Moshou, D., Bravo, C., West, J., Wahlen, S., McCartney, A., and Ramon, H. (2004). Automatic detection of ‘yellow rust’ in wheat using reflectance measurements and neural networks. *Computers and Electronics in Agriculture*, 44(3):173–188.

- [Nair et al. 2018] Nair, N., Thomas, C., and Jayagopi, D. B. (2018). Human Activity Recognition Using Temporal Convolutional Network. In *Proceedings of the 5th International Workshop on Sensor-based Activity Recognition and Interaction*, pages 17:1–17:8.
- [Opitz and Maclin 1999] Opitz, D. and Maclin, R. (1999). Popular Ensemble Methods: An Empirical Study. *Journal of Artificial Intelligence Research*, 11:169–198.
- [Ossama et al. 2014] Ossama, A.-H., Mohamed, A.-R., Jiang, H., Deng, L., Penn, G., and Yu, D. (2014). Convolutional Neural Networks for Speech Recognition. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, 22(10).
- [Pantazi et al. 2016] Pantazi, X., Moshou, D., Alexandridis, T., Whetton, R., and Mouazen, A. (2016). Wheat yield prediction using machine learning and advanced sensing techniques. *Computers and Electronics in Agriculture*, 121:57–65.
- [Patil and Deka 2016] Patil, A. P. and Deka, P. C. (2016). An extreme learning machine approach for modeling evapotranspiration using extrinsic inputs. *Computers and Electronics in Agriculture*, 121:385–392.
- [Perea et al. 2015] Perea, R. G., Poyato, E. C., Montesinos, P., and Díaz, J. A. R. (2015). Irrigation Demand Forecasting Using Artificial Neuro-Genetic Networks. *Water Resources Management*, 29(15):5551–5567.
- [Pereira et al. 2013] Pereira, A. R. N., Seduyama, G. C., and Nova, N. A. V. (2013). *Evapotranspiração*. Fundag.
- [Polikar 2012] Polikar, R. (2012). *Ensemble Machine Learning*. Springer New York Dordrecht Heidelberg London.
- [Qu et al. 2016] Qu, S., Li, J., Dai, W., and Das, S. (2016). Understanding Audio Pattern Using Convolutional Neural Network from Raw Waveforms. *CoRR*, abs/1611.09524.
- [Radwan and Burgard 2018] Radwan, N. and Burgard, W. (2018). Effective Interaction-aware Trajectory Prediction using Temporal Convolutional Neural Networks. In *Workshop on Crowd Navigation: Current Challenges and New Paradigms for Safe Robot Navigation in Dense Crowds at IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.
- [Ren et al. 2016] Ren, S., He, K., Girshick, R., and Sun, J. (2016). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6).
- [Russakovsky et al. 2015] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., Fei-Fei, L., Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z.,

- Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3):211–252.
- [Sadaei et al. 2016] Sadaei, H. J., Enayatifar, R., Lee, M. H., and Mahmud, M. (2016). A hybrid model based on differential fuzzy logic relationships and imperialist competitive algorithm for stock market forecasting. *Applied Soft Computing Journal*, 40:132–149.
- [Sadaei et al. 2017] Sadaei, H. J., Guimarães, F. G., José da Silva, C., Lee, M. H., and Eslami, T. (2017). Short-term load forecasting method based on fuzzy time series, seasonality and long memory process. *International Journal of Approximate Reasoning*, 83:196–217.
- [Severiano et al. 2017] Severiano, C. A., Silva, P. C., Sadaei, H. J., and Guimarães, F. G. (2017). Very short-term solar forecasting using fuzzy time series. In *IEEE International Conference on Fuzzy Systems*. Institute of Electrical and Electronics Engineers Inc.
- [Shen et al. 2018] Shen, Z., Zhang, Y., Lu, J., Xu, J., and Xiao, G. (2018). SeriesNet:A Generative Time Series Forecasting Model. In *Proceedings of the International Joint Conference on Neural Networks*. IEEE.
- [Siami-Namini and Namin 2018] Siami-Namini, S. and Namin, A. S. (2018). Forecasting Economics and Financial Time Series: ARIMA vs. LSTM.
- [Silva et al. 2018] Silva, G. C., Silva, J. L., Lisboa, A. C., Vieira, D. A., and Saldanha, R. R. (2018). Advanced fuzzy time series applied to short term load forecasting. In *2017 IEEE Latin American Conference on Computational Intelligence, LA-CCI 2017 - Proceedings*, volume 2017-November, pages 1–6. Institute of Electrical and Electronics Engineers Inc.
- [Silva et al. 2016] Silva, P. C., Sadaei, H. J., and Guimaraes, F. G. (2016). Interval forecasting with Fuzzy Time Series. In *2016 IEEE Symposium Series on Computational Intelligence, SSCI 2016*. Institute of Electrical and Electronics Engineers Inc.
- [Simonyan and Zisserman 2015] Simonyan, K. and Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. In *Pattern Recognition (ACPR), Asian Conference on*. IEEE.
- [Szegedy et al. 2015] Szegedy, C., Liu, W., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA. IEEE.

- [Tompson et al. 2015] Tompson, J., Goroshin, R., Jain, A., LeCun, Y., and Bregler, C. (2015). Efficient object localization using Convolutional Networks. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 648–656. IEEE.
- [Torres et al. 2011] Torres, A. F., Walker, W. R., and McKee, M. (2011). Forecasting daily potential evapotranspiration using machine learning and limited climatic data. *Agricultural Water Management*, 98(4):553–562.
- [Tumer and Ghosh 1995] Tumer, K. and Ghosh, J. (1995). Order Statistics Combiners For Neural Classifiers. In *Proceedings of the World Congress on Neural Networks*, pages 31–34.
- [United Nations 2015] United Nations (2015). Transforming our World: The 2030 Agenda for Sustainable Development. Technical report.
- [Van Den Oord et al. 2016] Van Den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., and Kavukcuoglu, K. (2016). WAVENET: A Generative Model for Raw Audio. *Arxiv*.
- [Waller et al. 2015] Waller, P., Yitayew, M., Waller, P., and Yitayew, M. (2015). Crop Evapotranspiration. In *Irrigation and Drainage Engineering*, pages 89–104. Springer International Publishing.
- [Xie and Hong 2016] Xie, J. and Hong, T. (2016). GEFCom2014 probabilistic electric load forecasting: An integrated solution with forecast combination and residual simulation. *International Journal of Forecasting*, 32(3):1012–1016.
- [Xiong et al. 2016] Xiong, Y., Luo, Y., Wang, Y., Traore, S., Xu, J., Jiao, X., and Fipps, G. (2016). Forecasting daily reference evapotranspiration using the Blaney–Criddle model and temperature forecasts. *Archives of Agronomy and Soil Science*, 62(6):790–805.
- [Yan and Ouyang 2018] Yan, H. and Ouyang, H. (2018). Financial Time Series Prediction Based on Deep Learning. *Wireless Personal Communications*, 102(2):683–700.
- [Yang et al. 2016] Yang, Y., Cui, Y., Luo, Y., Lyu, X., Traore, S., Khan, S., and Wang, W. (2016). Short-term forecasting of daily reference evapotranspiration using the Penman-Monteith model and public weather forecasts. *Agricultural Water Management*, 177:329–339.
- [Yu and Koltun 2016] Yu, F. and Koltun, V. (2016). Multi-Scale Context Aggregation By Dilated Convolutions. *CoRR*, abs/1511.07122:1 – 13.
- [Zarco-Tejada et al. 2014] Zarco-Tejada, P. J., Hubbard, N., and Loudjani, P. (2014). Precision Agriculture: An Opportunity for EU Farmers - Potential Support with the CAP 2014 - 2020. *European Parliament’s Committee on Agriculture and Rural Development*.

- [Zhang et al. 2019] Zhang, J., Chi, Y., and Xiao, L. (2019). Solar Power Generation Forecast Based on LSTM. In *Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS*, volume 2018-November, pages 869–872. IEEE Computer Society.
- [Zhao et al. 2019] Zhao, T., Wang, Q. J., Schepen, A., and Griffiths, M. (2019). Ensemble forecasting of monthly and seasonal reference crop evapotranspiration based on global climate model outputs. *Agricultural and Forest Meteorology*, 264:114–124.
- [Zhi-Hua 2012] Zhi-Hua, Z. (2012). *Ensemble Methods Foundations and Algorithms*. Chapman and Hall/CRC, 1 edition.