



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO DE JOINVILLE
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE SISTEMAS
ELETRÔNICOS

LUIZ HENRIQUE BUZZI

**PREDIÇÃO DE GERAÇÃO DE ENERGIA FOTOVOLTAICA UTILIZANDO IMAGENS
DE SATÉLITE E APRENDIZADO DE MÁQUINA**

DISSERTAÇÃO DE MESTRADO

Joinville
2023

Luiz Henrique Buzzi

**PREDIÇÃO DE GERAÇÃO DE ENERGIA FOTOVOLTAICA UTILIZANDO IMAGENS
DE SATÉLITE E APRENDIZADO DE MÁQUINA**

Dissertação submetida ao Programa de Pós-Graduação em Engenharia de Sistemas Eletrônicos da Universidade Federal de Santa Catarina para a obtenção do título de Mestre em Engenharia de Sistemas Eletrônicos.
Orientador: Prof. Pablo Andretta Jaskowiak, Dr.

Joinville
2023

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Buzzi, Luiz Henrique

Predição de geração de energia fotovoltaica utilizando
imagens de satélite e aprendizado de máquina / Luiz
Henrique Buzzi ; orientador, Pablo Andretta Jaskowiak,
2023.

150 p.

Dissertação (mestrado) - Universidade Federal de Santa
Catarina, Campus Joinville, Programa de Pós-Graduação em
Engenharia de Sistemas Eletrônicos, Joinville, 2023.

Inclui referências.

1. Engenharia de Sistemas Eletrônicos. 2. Energia
fotovoltaica. 3. Imagens de satélite. 4. Aprendizado de
máquina. 5. GOES-16. I. Jaskowiak, Pablo Andretta. II.
Universidade Federal de Santa Catarina. Programa de Pós
Graduação em Engenharia de Sistemas Eletrônicos. III. Título.

Luiz Henrique Buzzi

**Predição de geração de energia fotovoltaica utilizando imagens de satélite e
aprendizado de máquina**

O presente trabalho em nível de Mestrado foi avaliado e aprovado por banca
examinadora composta pelos seguintes membros:

Luiz Fernando Sommaggio Coletta, Dr.

Dell

Prof. Lucas Weihmann, Dr.

UFSC

Prof. Moises Ferber, Dr.

UFSC

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi
julgado adequado para obtenção do título de Mestre em Engenharia de Sistemas
Eletrônicos.

Prof. Lucas Weihmann, Dr.

Coordenador do Programa

Prof. Pablo Andretta Jaskowiak, Dr.

Orientador

Joinville, 11 de abril de 2023.

Dedico este trabalho a todos aqueles que me auxiliaram para a concretização desta etapa.

AGRADECIMENTOS

Gostaria de agradecer minha família, especialmente, meu pai que sempre me deu suporte e minha mãe que sempre acreditou em mim. A minha companheira, uma pessoa que foi compreensiva do início ao fim, mesmo com os desafios desta jornada. Ao meu orientador pelo conhecimento e sabedoria compartilhados, foi sua motivação que me trouxe até aqui. A todos os professores que de alguma forma contribuíram para a construção do conhecimento que foi base para este trabalho. A todos os amigos que durante a execução deste trabalho me apoiaram e engajaram para a sua conclusão. Ao laboratório Fotovoltaica-UFSC pela disponibilização dos dados da estação solarimétrica e ao NOAA que permite o acesso de sua base de dados ao público e nos direcionou as fontes corretas.

"A dúvida é o princípio da sabedoria."
(Aristóteles)

RESUMO

A busca crescente por energias renováveis cria um desafio de gestão de fornecimento energético. Fontes de energia fotovoltaica possuem intermitência de geração ocasionada por fatores meteorológicos e a imprevisibilidade destas variações deixa as redes suscetíveis a problemas de instabilidade, qualidade e equilíbrio. A produção de energia fotovoltaica possui uma forte relação com a irradiação solar, porém esta irradiação depende de diversos fatores meteorológicos que combinados determinam sua capacidade de geração. Neste aspecto, a previsão da geração de energia em painéis fotovoltaicos pode melhorar o gerenciamento das diversas fontes energética através do planejamento de geração. Esta previsão permite equilibrar diferentes fontes de energia, o que é vital para alcançar uma maior penetração da utilização de energia fotovoltaica. Para contornar o problema de previsibilidade da geração, este trabalho se propõe a investigar a combinação de dados de satélite com dados meteorológicos para prever o potencial de geração de energia em painéis fotovoltaicos em horizontes de 30, 60, 120 e 180 minutos. Para tal são utilizadas imagens do satélite GOES-16 em combinação com dados de uma estação meteorológica terrestre na ilha de Florianópolis. Os dados são utilizados para alimentar uma rede neural convolucional. As convoluções são utilizadas na construção de uma camada para extração de características das imagens de satélite com intuito de analisar a informação para relacionar com a irradiação solar. A saída da rede convolucional é utilizada como entrada de uma rede *multilayer perceptron*, que utiliza os dados para prever a irradiação na estação. Como resultado do estudo foi observado que os modelos com imagens de satélite fornecem previsões na ordem de 41,14% melhor no horizonte de 30min e 21,24% melhor no horizonte de 180min quando comparados com os modelos sem imagens de satélite.

Palavras-chave: Energia fotovoltaica. Imagens de satélite. Aprendizado de máquina. GOES-16. Redes neurais convolucionais. Redes neurais artificiais.

ABSTRACT

The increasing demand for renewable energy presents a challenge in energy supply management. Photovoltaic energy sources have intermittent generation due to weather factors, and the unpredictability of these variations leaves the power grids susceptible to instability, quality and balance issues. The production of photovoltaic energy is closely related to solar irradiation, which, depends on various meteorological factors that collectively determine its generation capacity. In this regard, forecasting energy generation in photovoltaic panels can enhance the management of diverse energy sources through generation planning. This forecast enables the balancing of different energy sources, which is crucial for achieving higher penetration of photovoltaic energy utilization. To address the predictability issue in generation, this study aims to investigate the combination of satellite data with meteorological data to predict the energy generation potential in photovoltaic panels within 30, 60, 120, and 180-minute horizons. For this purpose, satellite images from GOES-16 are used in combination with data from a ground-based weather station on Florianopolis Island. The data is fed into a convolutional neural network, where convolutions are employed to extract features from satellite images, aiming to analyze the information and establish a relationship with solar irradiation. The output of the convolutional network serves as input for a multilayer perceptron network, which utilizes the data to predict irradiation at the station. As a result of the study, it was observed that models incorporating satellite images provide predictions that are approximately 41.14% better for the 30-minute horizon and 21.24% better for the 180-minute horizon when compared to models without satellite images.

Keywords: Photovoltaic. Satellite images. Machine learning. GOES-16. Convolutional neural network. Artificial neural network

LISTA DE FIGURAS

Figura 2.1 – Propagação da irradiação solar na superfície da terra.	34
Figura 2.2 – Metodologias aplicadas para previsão de geração de energia em painéis fotovoltaicos em diferentes horizontes de previsão.	38
Figura 3.1 – Primeira visualização completa do clima da terra.	44
Figura 3.2 – Imagem gerada pelo satélite Terra no dia 13 de maio de 2018.	44
Figura 3.3 – Imagem gerada pelo satélite GOES-16 no dia 15 de janeiro de 2017.	45
Figura 3.4 – Composição das camadas externas da terra.	47
Figura 3.5 – Posição espacial e cobertura do GOES-East e GOES-West.	49
Figura 3.6 – Posições das regiões de escaneamento do GOES-16 e efeito da curvatura sobre a resolução espacial.	51
Figura 3.7 – Primeira imagem coletada do CONUS pelo sensor ABI em seus 16 canais espectrais. Começando pelos canais 1 até a 16 da esquerda para direita, de cima para baixo.	52
Figura 4.1 – Modelo do Perceptron.	57
Figura 4.2 – Exemplo de MLP, onde x_1 a x_5 são as entradas e y a saída.	58
Figura 4.3 – Detecção de bordas, utilizando um <i>kernel</i> de 2×2	61
Figura 4.4 – Exemplo de rede convolucional para processamento de imagem.	62
Figura 4.5 – Aplicação do MaxPooling.	62
Figura 4.6 – Arquitetura de predição utilizando imagens de satélite, CNN e MLP proposta por Si et al. (2021).	64
Figura 4.7 – Arquitetura de predição utilizando imagens de satélite, proposta por Eissa et al. (2013).	65
Figura 5.1 – Distribuição média do GHI ao longo de 24h nos anos de 2018 e 2019.	69
Figura 5.2 – Distribuição média do GHI ao longo dos meses nos anos de 2018 e 2019.	70
Figura 5.3 – Distribuição média do GHI Hora x Mês para os anos de 2018 e 2019.	70
Figura 5.4 – Recorte de 10° de latitude e 10° de longitude do ponto de interesse, com zoom na estação Fotovoltaica, sendo esta a área utilizada como entrada da rede.	72
Figura 5.5 – A imagem (a) é uma representação do local obtida do Google (2022). As imagens (b) a (g) foram obtidas do satélite GOES-16 canal 2 em diferentes intervalos de tempo no mesmo local, para representar a refletância em diferentes momentos do dia com os valores mínimos e máximos de GHI.	73
Figura 5.6 – Distribuição média da refletância ao longo de 24h nos anos de 2018 e 2019.	74

Figura 5.7 – Distribuição média da refletância ao longo dos meses nos anos de 2018 e 2019.	75
Figura 5.8 – Distribuição média da refletância Hora x Mês para os anos de 2018 e 2019.	76
Figura 5.9 – Relação do GHI da estação terrestre com os dados de satélite no intervalo das 8h às 17:30h GMT-3.	77
Figura 5.10 – Correlação da média de 4 pontos da imagem de satélite com a estação terrestre a cada 30min, no intervalo das 8h às 17:30h. . .	77
Figura 5.11 – Matriz de correlação de cada ponto da imagem de satélite com a estação terrestre a cada 1h, no intervalo das 8h às 17:30h.	78
Figura 5.12 – Correlação da média de 4 pontos da imagem de satélite com a estação terrestre mensalmente, no intervalo das 8h às 17:30h. . .	79
Figura 5.13 – Correlação da média de 4 pontos da imagem de satélite com a estação terrestre mês x hora.	79
Figura 6.1 – Sequência temporal para cada predição.	81
Figura 6.2 – Topologia da ANN proposta.	82
Figura 6.3 – Topologia da camada convolucional.	83
Figura 6.4 – Topologia da camada de combinação dos dados.	84
Figura 6.5 – Topologia da rede considerando apenas imagens de satélite. . . .	87
Figura 6.6 – Topologia da rede considerando apenas dados da estação terrestre.	88
Figura 7.1 – Distribuição dos resultados de RMSE para todos os modelos treinados nas configurações apresentadas, no ano de 2019.	93
Figura 7.2 – Densidade do erro para os melhores modelos no ano de 2019. . .	94
Figura 7.3 – RMSE horário para o ano de 2019, onde o tempo é momento da predição e o resultado se refere a predição mais o horizonte. . . .	94
Figura 7.4 – RMSE mensal para o ano de 2019.	95
Figura 7.5 – Distribuição dos resultados de RMSE para todos os modelos treinados nas configurações apresentadas, no ano de 2019.	96
Figura 7.6 – Variação do RMSE dos modelos em relação ao Conjunto Img+t. .	97
Figura 7.7 – RMSE horário para o ano de 2019, onde o tempo é momento da predição e o resultado se refere a predição mais o horizonte. . . .	97
Figura 7.8 – RMSE mensal para o ano de 2019.	98
Figura 7.9 – Distribuição dos resultados de RMSE para todos os modelos treinados nas configurações apresentadas, no ano de 2019.	99
Figura 7.10 – Comparação do ganho de desempenho dos modelos em relação ao Conjunto Img+t, sem dados da estação terrestre.	99
Figura 7.11 – RMSE horário para o ano de 2019, onde o tempo é momento da predição e o resultado se refere a predição mais o horizonte. . . .	100
Figura 7.12 – RMSE mensal para o ano de 2019.	101

Figura 7.13 – Comparação do ganho de desempenho dos modelos em relação ao Conjunto $Img+t$, sem dados da estação terrestre.	102
Figura 7.14 – RMSE horário para o ano de 2019, onde o tempo é momento da predição e o resultado se refere a predição mais o horizonte.	102
Figura 7.15 – RMSE mensal para o ano de 2019.	103
Figura 7.16 – RMSE dos 3 melhores conjuntos por categoria para os dados de validação em W/m^2	104

LISTA DE TABELAS

Tabela 2.1 – Correlação entre a geração de energia PV e os parâmetros meteorológicos de uma usina PV na China.	35
Tabela 2.2 – Classificação dos horizontes de previsão.	36
Tabela 3.1 – Tempo de captura em cada uma das regiões e de acordo com o modo de operação.	50
Tabela 3.2 – Canais espectrais disponíveis no imageador ABI e suas aplicações típicas.	53
Tabela 4.1 – Funções de ativação.	59
Tabela 5.1 – Dados do laboratório e usina fotovoltaica UFSC.	68
Tabela 5.2 – Correlação da potência do gerador com as variáveis de irradiação.	68
Tabela 5.3 – Correlação do GHI com as variáveis da estação para os anos de 2018 e 2019 das 8h às 18h.	69
Tabela 6.1 – Exemplo de representação das bases de dados considerando diferentes horizontes de previsão, $s = 30min$	85
Tabela 7.1 – Configuração e RMSE dos melhores resultados para cada conjunto em um horizonte de 30min.	91
Tabela 7.2 – Configuração e RMSE dos melhores resultados para cada conjunto em um horizonte de 60min.	92
Tabela 7.3 – Configuração e RMSE dos melhores resultados para cada conjunto em um horizonte de 120min.	92
Tabela 7.4 – Configuração e RMSE dos melhores resultados para cada conjunto em um horizonte de 180min.	92
Tabela 7.5 – Comparação dos resultados com o estudo de Marquez, Pedro e Coimbra (2013). Os valores representam o RMSE dos modelos e expressos em W/m^2	105
Tabela 7.6 – Comparação dos resultados com o estudo de Si et al. (2021). Os valores representam o RMSE dos modelos e expressos em W/m^2	106

LISTA DE ABREVIATURAS E SIGLAS

NZE	<i>Net Zero Emissions by 2050 Scenario</i>	29
IEA	<i>International Energy Agency</i>	29
PV	Fotovoltaica	29
ANN	<i>Artificial Neural Network</i>	30
SVM	<i>Support Vector Machine</i>	30
CNN	<i>Convolutional Neural Network</i>	30
NWP	<i>Numerical Weather Prediction</i>	30
GOES	<i>Geostationary Operational Environmental Satellite</i>	30
DNI	<i>Direct Normal Irradiance</i>	33
SOM	<i>Self-Organizing Map</i>	36
IR	Infravermelho	36
kNN	<i>k-Nearest Neighbor</i>	37
MLR	<i>Multivariate Linear Regression</i>	37
TSI	<i>Total Sky Imager</i>	37
SVR	<i>Support Vector Regression</i>	39
ARX	<i>Autoregressive Model With Exogenous</i>	39
ARMA	<i>Autoregressive Moving Average</i>	39
ARIMA	<i>Autoregressive Integrated Moving Average</i>	39
MAE	<i>Mean Absolute Error</i>	40
MAPE	<i>Mean Absolute Percentage Error</i>	40
RMSE	<i>Root Mean Squared Error</i>	40
TIROS	<i>Television and Infrared Observation Satellite</i>	43
NOAA	<i>National Oceanic and Atmospheric Administration</i>	46
NASA	<i>National Aeronautics and Space Administration</i>	46
EUA	Estados Unidos da América	47
EXIS	<i>Extreme Ultraviolet and X-ray Irradiance Sensors</i>	49
SUVI	<i>Solar Ultraviolet Imager</i>	49
MAG	<i>Magnetometer</i>	49
SEISS	<i>Space Environment In-Situ Suite</i>	49
ABI	<i>Advanced Baseline Imager</i>	49
GLM	<i>Geostationary Lightning Mapper</i>	49
CONUS	<i>Continental US</i>	50
VNIR	<i>Visible and Near-Infrared</i>	51
MWIR	<i>Mid-Wavelength Infrared</i>	51
LWIR	<i>Long-Wavelength Infrared</i>	52
MLP	<i>Multilayer Perceptron</i>	55
DT	<i>Decision Trees</i>	62

LR	<i>Linear Regression</i>	62
PIV	<i>Particle Image Velocimetry</i>	63
MSE	<i>Mean Squared Error</i>	85

LISTA DE SÍMBOLOS

DHI	Direct Horizontal Irradiance	33
GHI	Global Horizontal Irradiance	33
θ	Ângulo Zenital	33
I_0	Constante Solar	33
ε	Correção de Excentricidade	33
ν	Massa do Ar	33
δ	Profundidade Óptica	33
δ_R	Profundidade Óptica de Rayleigh	34
δ_{gas}	Profundidade Óptica Devido as Moléculas de Ar e Dióxido de Car- bono	34
δ_{ozonio}	Profundidade Óptica do Ozônio	34
$\delta_{aerossol}$	Profundidade Óptica dos Aerossóis	34
δ_{agua}	Profundidade Óptica dos Vapores de Água	34
R^2	Coefficient of Determination	40
g_i	Potência gerada	41
\hat{g}_i	Potência estimada	41
N	Número de pontos estimados	41
g_0	Capacidade da usina	41
I	Entrada normalizada	56
I_{max}	Valor máximo de entrada	56
I_{min}	Valor mínimo de entrada	56
D	Dado a ser normalizado	56
D_{max}	Valor máximo do conjunto de dados de entrada	56
D_{min}	Valor mínimo do conjunto de dados de entrada	56
x	Entrada da ANN	57
y	Saída da ANN	57
w	Peso do da conexão da ANN	57
b	Bias	57
t	Instante atual	81
h	Horizonte de predição	81
s	Granularidade dos dados	81

SUMÁRIO

1	INTRODUÇÃO	29
1.1	OBJETIVOS	31
1.1.1	Objetivo Geral	31
1.1.2	Objetivos Específicos	31
1.2	ORGANIZAÇÃO DO TRABALHO	31
2	PREVISÃO DE GERAÇÃO DE ENERGIA PV	33
2.1	IRRADIAÇÃO SOLAR	33
2.2	VARIÁVEIS QUE POSSUEM CORRELAÇÃO COM A GERAÇÃO PV	34
2.3	HORIZONTE DE PREVISÃO	36
2.4	TÉCNICAS DE PREVISÃO DE GERAÇÃO DE ENERGIA PV	37
2.4.1	Métodos estatísticos de séries temporais	37
2.4.2	Métodos físicos	39
2.4.3	Métodos híbridos	40
2.5	ANÁLISE DE PERFORMANCE DO MODELO	40
2.6	CONSIDERAÇÕES FINAIS	41
3	IMAGENS DE SATÉLITE	43
3.1	INTRODUÇÃO A IMAGENS DE SATÉLITE	43
3.2	GOES-16	47
3.2.1	Advanced Baseline Imager - ABI	50
3.2.2	Produtos	52
3.3	CONSIDERAÇÕES FINAIS	54
4	MACHINE LEARNING	55
4.1	PREPARAÇÃO DOS DADOS	55
4.2	ARTIFICIAL NEURAL NETWORKS - ANN	57
4.2.1	Multilayer Perceptron - MLP	58
4.2.2	Convolutional Neural Network - CNN	60
4.3	TRABALHOS RELACIONADOS	62
4.4	CONSIDERAÇÕES FINAIS	64
5	ESTUDO DE CASO	67
5.1	ESTAÇÃO TERRESTRE	67
5.2	SATÉLITE	70
5.3	HORIZONTE E GRANULARIDADE	74
5.4	ANÁLISE EXPLORATÓRIA	75
5.5	CONSIDERAÇÕES FINAIS	80
6	MÉTODOS	81
6.1	TOPOLOGIA	81
6.1.1	Camada convolucional	82

6.1.2	Estratégia de combinação dos dados	83
6.1.3	Camada MLP	83
6.1.4	Organização dos dados	84
6.2	PARÂMETROS DE TREINAMENTO	85
6.3	EXPERIMENTOS	86
6.3.1	Previsões baseadas em dados de satélite	86
6.3.2	Previsões baseadas em dados da estação terrestres	87
6.3.3	Previsões baseadas na combinação dos dados de satélite com a estação terrestre	88
6.4	CONSIDERAÇÕES FINAIS	89
7	RESULTADOS	91
7.1	SELEÇÃO DE MODELOS PARA ANÁLISE DOS RESULTADOS	91
7.2	INFLUÊNCIA DE PARÂMETROS TEMPORAIS	93
7.3	INFLUÊNCIA DOS DADOS DA ESTAÇÃO TERRESTRE	95
7.4	INFLUÊNCIA DO UV	98
7.5	MELHORES RESULTADOS POR CATEGORIA E ANÁLISE GERAL	102
8	CONSIDERAÇÕES FINAIS	107
8.1	TRABALHOS FUTUROS	108
	REFERÊNCIAS	111
	APÊNDICE A – SATÉLITES GOES	117
	APÊNDICE B – ESTUDO PRELIMINAR	123
	APÊNDICE C – APRESENTAÇÃO DOS RESULTADOS	137

1 INTRODUÇÃO

A energia renovável tem se tornado cada vez mais presente no nosso cotidiano, sendo utilizada em diversas atividades. No *Net Zero Emissions by 2050 Scenario* (NZE) a eletricidade é responsável por aproximadamente 50% da soma de todas as fontes de energia do mundo até 2050 (COZZI; GOULD, 2021).

Um estudo realizado pela *International Energy Agency* (IEA), prevê que entre os anos de 2021 e 2026 devemos esperar uma aceleração anual de 60% na geração de energia por fontes renováveis, se compararmos com a expansão dos últimos 5 anos. O crescimento da capacidade de geração de energias renováveis utilizando a tecnologia fotovoltaica (PV) dos próximos 5 anos deve ser quase o dobro, se comparado com os anos anteriores, correspondendo sozinha a 60% das adições de energia renováveis (IEA, 2021).

Para América Latina é esperado um crescimento de 34% na geração de energia renovável. A participação do Brasil neste cenário é de 45GW, sendo que a expectativa para energia PV é acima de 70% do crescimento da geração de energia renovável no país (IEA, 2021).

Um dos grandes desafios na transição do modelo convencional de geração de energia para os modelos renováveis está na flexibilidade do sistema de distribuição. O sistema precisa ser flexível às variações de carga, variações climáticas e interrupção de geração, o que fomenta a pesquisa de como integrar e gerenciar estas diferentes fontes de energia (IMPRAM; NESE; ORAL, 2020)

Se considerarmos especificamente a geração de energia PV, a produção possui uma forte relação com a irradiação solar (AHMED et al., 2020). A irradiação depende de diversos fatores meteorológicos que combinados determinam o potencial de geração da estação (ANTONANZAS et al., 2016). A implementação em grande escala do modelo PV de geração cria um desafio de estabilidade, confiabilidade e equilíbrio, sendo que a intermitência da irradiação pode afetar diretamente a estabilidade do sistema (SOBRI; KOOHI-KAMALI; RAHIM, 2018).

Neste aspecto, a previsão da geração de energia em painéis fotovoltaicos pode melhorar o gerenciamento através do planejamento de geração. Esta previsão permite equilibrar diferentes fontes de energia, o que é vital para alcançar uma maior penetração da utilização de energia PV (RAZA; NADARAJAH; EKANAYAKE, 2016).

Observando este cenário, modelos vêm sendo amplamente estudados buscando melhorar o resultado em diferentes horizontes de previsão. Segundo Ahmed et al. (2020) várias abordagens de modelagem podem ser utilizadas, podendo citar os modelos físicos, estatísticos, de aprendizado de máquina, de previsão baseados em conjuntos e híbridos.

Segundo os estudos conduzidos por Voyant et al. (2017), Das et al. (2018) e So-

bri, Koochi-Kamali e Rahim (2018), os modelos que utilizam *Artificial Neural Networks* (ANN) e *Support Vector Machines* (SVM) têm sido muito empregados devido à sua habilidade de solucionar problemas complexos e não lineares. Enquanto para Ahmed et al. (2020) não só as ANNs como as *Convolutional Neural Networks* (CNN) sozinhas ou em modelos híbridos são as mais promissoras, principalmente para previsão em curto intervalo de tempo.

Além dos modelos, outras variáveis influenciam nos resultados de previsão observadas. No levantamento realizado por Sobri, Koochi-Kamali e Rahim (2018) os principais fatores que influenciam a previsão de geração de energia PV são a localização, o horizonte de previsão e a seleção dos parâmetros de entrada.

Atualmente diversos sistemas de medição instalados em estações terrestres coletam grandes bases de dados que podem ser utilizadas para a preparação e otimização dos modelos. Dentre os principais parâmetros estão dados meteorológicos como irradiação, temperatura, umidade, velocidade do vento, pressão atmosférica, ângulo zenital e cobertura de nuvens, sendo este último normalmente estimado utilizando imagens de satélite ou câmeras terrestres que capturam imagens do céu.

Em um estudo realizado por Miller et al. (2018), foram utilizadas imagens de satélite para descrever a distribuição das nuvens e suas propriedades ópticas. Os dados combinados com método *Numerical Weather Prediction* (NWP) são utilizados para a previsão da irradiância em horizonte de até 3h. Em seu estudo, Miller et al. (2018) evidencia a nova geração de satélites *Geostationary Operational Environmental Satellite* (GOES) 16 e 17 com maior resolução espacial que podem beneficiar a previsão utilizando satélites.

Este trabalho investiga a combinação de modelos de aprendizado de máquina e imagens de satélite para a previsão de geração de energia fotovoltaica em horizontes diários. Tal investigação é feita com base em um estudo de caso em uma planta PV localizada na cidade de Florianópolis - SC. No que diz respeito às imagens de satélite propriamente ditas, é utilizado como fonte o imageador ABI do satélite GOES-16, um radiômetro de 16 bandas espectrais de última geração. Além do uso de imagens de satélite isoladas, é investigada ainda sua combinação com dados provenientes de estações solarimétricas terrestres.

Para o desenvolvimento deste trabalho foram utilizados os dados da estação meteorológica e usina PV localizada no Laboratório de Sistemas Solares Fotovoltaica UFSC. O laboratório conta com uma extensa base de dados, com parâmetros meteorológicos e de geração. Para relacionar os dados da estação terrestre, será utilizada a base de imagens do satélite GOES-16.

1.1 OBJETIVOS

1.1.1 Objetivo Geral

O objetivo geral deste trabalho é empregar e avaliar modelos capazes de prever a geração de energia PV. Para tal, são utilizadas imagens do satélite GOES-16 em combinação com dados da estação meteorológica Fotovoltaica UFSC. Os resultados são utilizados para alimentar métodos de previsão com a capacidade de prever um horizonte *intra-day*, que compreende 30 minutos, 60 minutos, 120 minutos e 180 minutos.

1.1.2 Objetivos Específicos

- Obter e tratar os dados da estação meteorológica e usina Fotovoltaica UFSC dos anos de 2018 a 2019;
- Identificar os parâmetros meteorológicos obtidos da estação que possuem maior correlação com geração de energia PV na usina Fotovoltaica-UFSC;
- Estudar os produtos entregues pelo satélite GOES-16 para identificar quais possuem maior relação com o estudo;
- Obter imagens do satélite GOES-16 dos produtos selecionados nos anos 2018 a 2019. Com uma janela de 10° de latitude e longitude, centrada nas coordenadas da estação e planta Fotovoltaica-UFSC;
- Pré processar a base de dados e gerar conjuntos de treinamento;
- Modelar redes neurais artificiais para a tarefa;
- Otimizar os parâmetros das redes neurais para horizontes de 30 minutos, 60 minutos, 120 minutos e 180 minutos;
- Comparar os dados obtidos com resultados da literatura.

1.2 ORGANIZAÇÃO DO TRABALHO

Este trabalho se organiza da seguinte maneira. O Capítulo 2 traz uma introdução sobre os principais conceitos para previsão de energia PV. O Capítulo 3 apresenta um estudo sobre satélites, em especial o satélite GOES-16, se aprofundando nos imageadores e os produtos que a tecnologia entrega. O Capítulo 4 aprofunda o estudo em aprendizado de máquina, especificamente nas redes neurais e redes convolucionais que são a base para o desenvolvimento deste trabalho. O Capítulo 5 apresenta um estudo das bases de dados que são utilizadas no desenvolvimento do trabalho, passado pelos dados da estação terrestre, os dados de satélite e a combinação dos dois. O

Capítulo 6 apresenta a metodologia utilizada para a avaliação do estudo, descrevendo os modelos estudados e a composição dos parâmetros. O Capítulo 7 aborda os resultados alcançados, fazendo uma análise dos melhores modelos para cada conjunto de treinamento. Ao final o Capítulo 8 traz as considerações finais.

2 PREVISÃO DE GERAÇÃO DE ENERGIA PV

A busca crescente por energias renováveis cria um desafio de gestão de fornecimento. Fontes de energia fotovoltaica (PV) possuem intermitência de geração ocasionada por fatores meteorológicos e a imprevisibilidade destas variações deixa as redes suscetíveis a problemas de instabilidade, qualidade e equilíbrio. Os modelos de previsão, sendo eles numéricos ou estatísticos, precisam ser nutridos com dados para serem capazes de entregar o que se propõe. Este capítulo vai se aprofundar no estudo da previsão de geração de energia PV, abordando os principais conceitos, metodologias e fatores que influenciam nesta abordagem. A Seção 2.1 trata dos componentes de irradiação solar. Na Seção 2.2 são abordadas algumas variáveis físicas que influenciam a geração de energia PV com ênfase em dados de satélite. A Seção 2.3 apresenta os diferentes horizontes de previsão utilizados para esta abordagem, correlacionando com a granularidade dos dados e aplicação. Na Seção 2.4 são apresentadas algumas das metodologias utilizadas para a previsão de geração PV. O capítulo finaliza com a Seção 2.5, onde são apresentadas as principais métricas de avaliação de modelos de previsão. Ao final, na Seção 2.6, são apresentadas as notas conclusivas.

2.1 IRRADIAÇÃO SOLAR

A energia PV é obtida a partir da irradiação solar no plano das células PV (LAW et al., 2014). Esta irradiação que incide sobre a superfície da terra é responsável pela geração de energia PV e pode ser subdividida em duas componentes, a *Direct Normal Irradiance* (DNI) é a parcela que incide diretamente do sol, e a *Direct Horizontal Irradiance* (DHI) é a parcela afetada pelas reflexões e reemissões causadas na atmosfera (KUMAR et al., 2020). A Figura 2.1 ilustra a influência do ângulo zenital sobre as componentes de irradiação e a absorção causada pelos elementos da atmosfera.

A soma geométrica destas duas componentes resulta na *Global Horizontal Irradiance* (GHI), objeto de estudo deste trabalho, que corresponde ao total de irradiância incidente em uma superfície horizontal (ALBAIDI et al., 2014), dada por

$$GHI = DHI + DNI \cos(\theta)$$

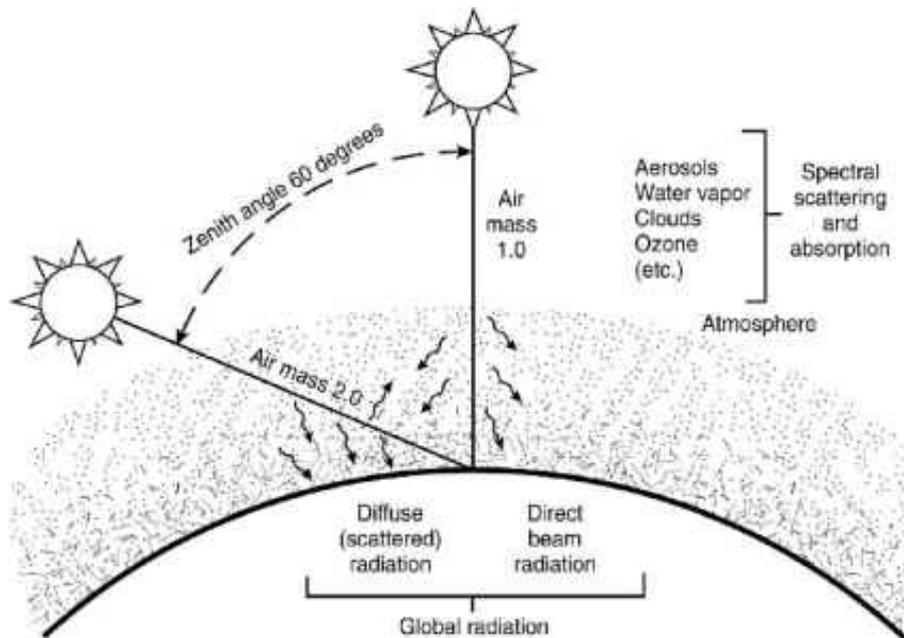
onde θ é o ângulo zenital solar.

O DNI pode ser estimado utilizado a lei de Bouguer-Lambert (KUMAR et al., 2020), dada por

$$DNI = I_0 \times \varepsilon \times e^{-(v \times \delta)}$$

onde I_0 é a constante solar, 1367 W/m^2 , ε é a correção de excentricidade, v é a massa do ar e δ é a profundidade óptica causada pela atenuação da atmosfera.

Figura 2.1 – Propagação da irradiação solar na superfície da terra.



Fonte: Williams (2023).

A profundidade óptica em uma atmosfera livre de nuvens pode ser calculada como sendo (EISSA et al., 2013),

$$\delta = \delta_R + \delta_{gas} + \delta_{ozonio} + \delta_{aerossol} + \delta_{agua}$$

onde δ_R é a profundidade óptica de Rayleigh, ou seja, atmosfera limpa e seca, δ_{gas} é a profundidade óptica devido as moléculas de ar e dióxido de carbono, δ_{ozonio} é a profundidade óptica do ozônio, $\delta_{aerossol}$ é a profundidade óptica dos aerossóis e δ_{agua} é a profundidade óptica dos vapores de água.

O GHI normalmente é medido utilizando piranômetro horizontal ou células de referência de silício, enquanto o DNI é medido utilizando pireliômetro e irradiômetro de banda de sombra giratória. O DHI é medido com piranômetro instalado em um rastreador solar (KUMAR et al., 2020).

2.2 VARIÁVEIS QUE POSSUEM CORRELAÇÃO COM A GERAÇÃO PV

A irradiância solar GHI é o parâmetro que possui maior relação com a geração de energia PV (AHMED et al., 2020). Porém a maior parte das usinas não contam com a medida de irradiância em suas estações meteorológicas (SOBRI; KOOHI-KAMALI; RAHIM, 2018). Apesar disso, a geração de energia PV e a irradiância sofrem influência de diversos fatores meteorológicos que afetam a capacidade de geração. Podendo citar a duração da luz do sol, temperatura ambiente máxima, umidade relativa, latitude,

longitude, dia do ano, radiação global diária de céu limpo, cobertura total de nuvens, temperatura, índice de clareza, altitude, temperatura média, nebulosidade média, velocidade média do vento, pressão atmosférica, índice de claridade de referência, radiação difusa média, radiação média do feixe, mês, radiação extraterrestre, evaporação e temperatura do solo (SOBRI; KOOHI-KAMALI; RAHIM, 2018). Além da composição da atmosfera, conforme tratado na Seção 2.1.

Para Ahmed et al. (2020) os parâmetros podem ter influência positiva ou negativa no modelo, reduzindo o resultado da previsão, gerando atraso e aumentando o custo computacional devido à complexidade. É preciso fazer uma análise entre a produção PV e os parâmetros meteorológicos para identificar quais possuem maior correlação. Na Tabela 2.1 são apresentadas correlações entre alguns parâmetros meteorológicos e a geração de energia PV de uma usina PV na China.

Tabela 2.1 – Correlação entre a geração de energia PV e os parâmetros meteorológicos de uma usina PV na China.

Parâmetro Meteorológico	Coefficiente de Correlação
Irradiância Solar	0,9840
Temperatura do Ar	0,7615
Tipo de Nuvem	-0,4847
Ponto de Condensação da Água	0,6386
Umidade Relativa	-0,4918
Precipitação	0,3409
Direção do Vento	0,1263
Velocidade do Vento	0,1970
Pressão	0,0815

Fonte: Adaptado de Liu et al. (2018).

Podemos identificar variáveis que possuem maior influência, como exemplo a irradiância solar e a temperatura do ar. Mas o tipo da nuvem possui influência, isso porque a nebulosidade é um dos fatores mais significativos na variação da irradiação solar (AHMED et al., 2020). É importante salientar que as nuvens não são constantes, é preciso analisar o tipo, a quantidade de formações, o movimento, a dissipação e a deformação, pois estes fatores influenciam na fração de energia que atravessam as nuvens.

Segundo Ahmed et al. (2020) a maioria dos pesquisadores utiliza satélites para a análise de nuvens, porém não tem precisão para nuvens regionais ou baixas, devido a sua baixa resolução espacial e temporal, o que afeta as previsões de curto prazo. Para estas, são utilizadas técnicas de captura de imagens do céu com câmeras terrestres.

Neste sentido algumas pesquisas vêm sendo desenvolvidas utilizando imagens de satélites para extrair informações de nuvens e combinar com dados meteorológicos. Em Dong et al. (2014) as imagens de satélite são utilizadas para classificar a cobertura

de nuvem do céu de Singapura em 20 níveis, este dado é utilizado em uma rede *Self-Organizing Map* (SOM) que estima o índice de cobertura de nuvem.

Já a pesquisa realizada por Eissa et al. (2013), utiliza as imagens de satélite em diferentes espectros para criar uma máscara de nuvem. Os dados então são separados em dois conjuntos, com nuvem e sem nuvem. Cada conjunto de dados é treinado em uma rede neural separada que busca estimar a profundidade óptica considerando as variáveis de atenuação da atmosfera.

Na pesquisa de Marquez, Pedro e Coimbra (2013) as imagens de satélite são utilizadas no espectro visível e infravermelho (IR). Na primeira, as imagens são utilizadas para estimar o índice de cobertura de nuvem. Enquanto as imagens IR são utilizadas para estimar o mapa vetorial que representa a velocidade das nuvens. Os dois dados são combinados na entrada de uma rede neural para prever a geração de energia PV.

Na maior parte dos casos os dados de satélite são utilizados para gerar informações das nuvens. Estes dados combinados a dados meteorológicos são então utilizados como entrada de um modelo para realizar a predição da geração de energia PV.

2.3 HORIZONTE DE PREVISÃO

Antonanzas et al. (2016) classifica o horizonte de previsão de geração de energia PV de acordo com sua aplicação. Sendo ele subdividido em 3 categorias, apresentadas na Tabela 2.2.

Tabela 2.2 – Classificação dos horizontes de previsão.

	Horizonte de previsão	Aplicação	Granularidade dos dados
<i>Intra-hour</i>	15min a 2h	Variações relacionadas a operação	30s a 5min
<i>Intra-day</i>	30min a 6h	Previsão de carga	30min a 1h
<i>Day-ahead</i>	1 dia a 3 dias	Programação de transmissão	1h

Fonte: Adaptado de Voyant et al. (2017).

A previsão *intra-hour* é utilizada para garantir a qualidade e estabilidade da rede. Porém, usinas PV distribuídas não sofrem tanto com a variação de curto prazo, uma vez que a produção é composta de grandes áreas e tendem a equilibrar estas variações (ANTONANZAS et al., 2016). São aplicados para predições até 2h.

A previsão *intra-day* é utilizada para balanceamento de carga e é fundamental para os operadores de rede controlarem as diferentes zonas de carga (AHMED et al., 2020). Atuam normalmente em horizontes de 1h a 6h.

Por fim, a previsão *day-ahead* é utilizada para o planejamento de utilização da estação em conjunto com a rede (ANTONANZAS et al., 2016). São utilizadas para

predições maiores que 1 dia.

O horizonte de previsão influencia diretamente na granularidade dos dados que são necessários para realizar a previsão. Em Urquhart et al. (2013) é utilizada uma granularidade de 30s e são feitas predições de 30s a 15min utilizando imagens de nuvens capturadas por uma câmera terrestre. A metodologia consiste em detectar as nuvens, sua posição e velocidade para estimar a posição futura.

Já em Almonacid et al. (2014) é feita a previsão com horizonte de 1h utilizando redes neurais dinâmicas, onde a saída depende não só dos dados de entrada atual da rede, como também os dados anteriores. Os dados de entrada são a irradiância global e a temperatura do ar com granularidade de 2 min.

Em um estudo feito por Long, Zhang e Su (2014), é apresentado um comparativo de desempenho de quatro técnicas, sendo elas *Artificial Neural Networks* (ANN), *Support Vector Machines* (SVM), *k-Nearest Neighbor* (kNN), e *Multivariate Linear Regression* (MLR), para previsão de geração diária. Utilizando como parâmetros a potência gerada, a irradiância global, dentre outras variáveis meteorológicas com granularidade de 1 dia.

Conforme definido nos objetivos, este trabalho se concentra em predições de 30, 60, 120 e 180 minutos, com objetivo de lidar com as variações relacionadas ao controle de carga. A granularidade será objeto de estudo na análise dos dados, Capítulo 5, porém limitada a 30 minutos, pela frequência de captura do satélite GOES-16, informação que será tratada no Capítulo 3.

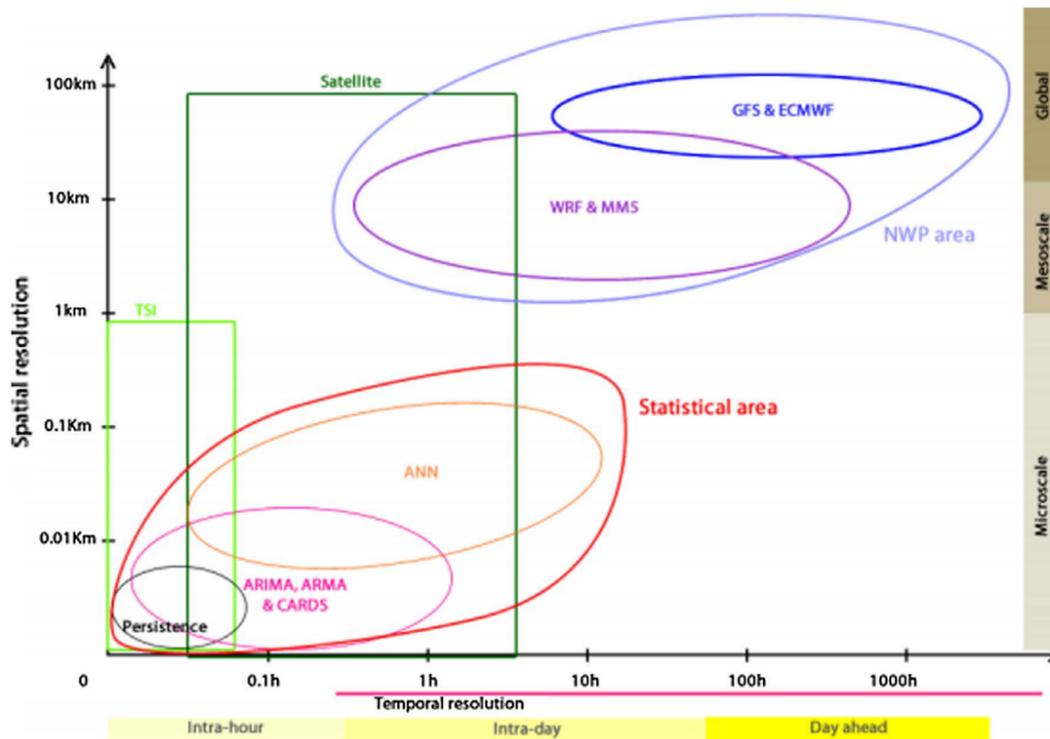
2.4 TÉCNICAS DE PREVISÃO DE GERAÇÃO DE ENERGIA PV

Foi apresentado por Sobri, Koohi-Kamali e Rahim (2018) a divisão dos métodos de previsão em 3 grandes grupos: métodos estatísticos de série temporal, métodos físicos e métodos híbridos. A Figura 2.2 correlaciona os diferentes métodos com os horizontes de predição, resolução espacial e granularidade dos dados, os retângulos representam o espaço das aplicações com imagens, satélite ou *Total Sky Imager* (TSI), o conjunto vermelho representa o grupo dos métodos estatísticos e o conjunto lilás representa o grupo dos métodos *Numerical Weather Prediction* (NWP). Podemos perceber que para aplicações que demandem uma maior resolução espacial e uma predição para um horizonte mais baixo fica a cargo das imagens de satélite, enquanto que os horizontes mais longos ficam para os modelos NWP. Nas seções seguintes serão tratados cada um dos métodos.

2.4.1 Métodos estatísticos de séries temporais

Os métodos estatísticos dependem de dados históricos e buscam encontrar uma relação entre as variáveis de entrada para formar um modelo de previsão, não re-

Figura 2.2 – Metodologias aplicadas para previsão de geração de energia em painéis fotovoltaicos em diferentes horizontes de previsão.



Fonte: Sobri, Koohi-Kamali e Rahim (2018).

querendo informações de estado interno do sistema (VOYANT et al., 2017). Segundo Sobri, Koohi-Kamali e Rahim (2018) os métodos estatísticos e de série temporal costumam utilizar das seguintes técnicas: ANN, SVM, Cadeia de Markov, Autorregressão e Regressão.

No estudo realizado por Voyant et al. (2017), foi feita uma análise da aparição dos termos ANN, aprendizado de máquina e SVM nos 5 principais periódicos relacionados a previsão de energia solar (*Solar Energy, Energy, Applied Energy, Renewable Energy e Energy Conversion and Management*), como resultado, o termo ANN tem ganhado cada vez mais presença e se destaca como principal método para previsão de irradiância global.

As ANNs, tema de estudo deste trabalho, são comumente utilizadas para a previsão de geração de energia PV devido a sua capacidade de lidar com problemas complexos e não lineares (SOBRI; KOOHI-KAMALI; RAHIM, 2018). Isto ocorre devido a não linearidade nos dados meteorológicos (DAS et al., 2018). Aprofundaremos o estudo das ANNs no Capítulo 4.

O SVM é um método de aprendizado de máquina supervisionado que busca encontrar um hiperplano de separação que maximiza a margem da base de treina-

mento. Desenvolvido inicialmente para solução dos problemas de classificação, mas estendido para problemas de regressão com *Support Vector Regression* (SVR) (DAS et al., 2018). Este método é caracterizado pela capacidade de generalização e lidar com problemas não lineares (ANTONANZAS et al., 2016). Vários estudos podem ser citados utilizando SVM para esta aplicação, como Kromer et al. (2014), Jang et al. (2016) e Long, Zhang e Su (2014).

Já a Cadeia de Markov é definida como cadeia de estados conectados, onde o estado do sistema depende apenas do estado atual (SOBRI; KOOHI-KAMALI; RAHIM, 2018). Isto significa, que o os estados futuros são independentes dos estados passados, e cada estado precisa de todas as informações que podem influenciar na saída (PAOLI et al., 2010). Os estados futuros são ativados através de um processo probabilístico, utilizando uma matriz de transição. Alguns estudos nesta área utilizam desta metodologia como Paoli et al. (2010) e Li, Luan e Niu (2008).

Segundo Sobri, Koochi-Kamali e Rahim (2018), a autorregressão e a regressão são métodos que medem a correlação entre os parâmetros dependentes e independentes, sendo subdividido em modelos de regressão linear e não linear. Alguns modelos utilizam deste método como *Autoregressive Model With Exogenous* (ARX), *Autoregressive Moving Average* (ARMA) e *Autoregressive Integrated Moving Average* (ARIMA) (KUMAR et al., 2020). Podemos citar nesta categoria os trabalhos de Yang e Xie (2012), Almonacid et al. (2014) e Yang et al. (2015).

2.4.2 Métodos físicos

Os métodos físicos buscam correlacionar os componentes meteorológicos locais e atmosféricos para encontrar a irradiância solar na superfície, estes métodos eliminam a necessidade de dados de irradiância solar na superfície, porém os modelos precisam ser calibrados (SOBRI; KOOHI-KAMALI; RAHIM, 2018). Segundo Sobri, Koochi-Kamali e Rahim (2018) os métodos físicos são subdivididos em NWP, TSI e imagens de satélite.

O NWP é modelo que permitem determinar as condições climáticas locais e globais. Utiliza integração numérica de equações diferenciais acopladas que explicam os mecanismos de transporte de radiação e sua dinâmica na atmosfera (SOBRI; KOOHI-KAMALI; RAHIM, 2018). A grande vantagem deste método é ser baseado em modelos determinísticos. Vários autores aplicam a técnica para a previsão de geração de energia PV, dentre eles Almeida, Perpiñán e Narvarte (2015), Rodríguez-Benítez et al. (2020) e Miller et al. (2018).

O método TSI, consiste em aplicar câmeras terrestres para detecção de nuvens, medição da sua altura e determinação de seu movimento (SOBRI; KOOHI-KAMALI; RAHIM, 2018). Através das imagens o modelo busca estimar a posição futura das nuvens utilizando de modelo probabilístico, podendo citar os trabalhos de Miller et al.

(2018), Peng et al. (2013) e Chu, Pedro e Coimbra (2013).

As imagens de satélite, tema de estudo deste trabalho, buscam determinar a influência das nuvens através de imagens, muito similar com o modelo TSI. A vantagem deste, é que devido a sua área de captura, permite a previsão em horizontes mais longos (SOBRI; KOOHI-KAMALI; RAHIM, 2018). Esta metodologia tem ganhado cada vez mais espaço devido a evolução dos satélites e o aumento da sua resolução espacial. Trataremos das imagens de satélite a fundo no Capítulo 3.

2.4.3 Métodos híbridos

Os métodos híbridos, são utilizados para resolver fraquezas dos métodos individuais, podendo combinar um ou mais dos métodos anteriores (SOBRI; KOOHI-KAMALI; RAHIM, 2018). A combinação pode ser feita de forma linear, não linear ou ambas e é muito utilizado com os métodos estatísticos e de aprendizado de máquina com o objetivo de aumentar a precisão do modelo.

2.5 ANÁLISE DE PERFORMANCE DO MODELO

Para determinar a qualidade do modelo proposto, várias métricas podem ser utilizadas na avaliação e cada métrica foca em um determinado aspecto da geração (ANTONANZAS et al., 2016), além de permitir a comparação com outros modelos. Geralmente o erro calculado para a irradiância solar é medido em W/m^2 . Enquanto o erro de geração de energia é expresso em kW representando o erro de potência de saída da planta (SOBRI; KOOHI-KAMALI; RAHIM, 2018).

Neste contexto, Ahmed et al. (2020) sugere que as principais métricas incluem: *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE) e o *Root Mean Squared Error* (RMSE). Já Srivastava e Lessmann (2018), utilizam em seu estudo o RMSE e MAE. Outros autores como Zhang et al. (2015) e (SOBRI; KOOHI-KAMALI; RAHIM, 2018) sugerem que mesmo com a grande adesão dos pesquisadores pelos indicadores estatísticos clássicos, eles não conseguem suprir todos os requisitos necessários para aplicação, isto porque muitas das métricas de previsão desenvolvidas não levam em consideração os tipos de erros que têm maior impacto nas operações do sistema. Sendo que grandes erros de previsão podem afetar a estabilidade e o retorno econômico da operação. Desta forma é preciso uma análise completa, considerando o horizonte e a localização da estação. Neste trabalho iremos utilizar a abordagem clássica devido a facilidade de encontrar estas métricas em outros trabalhos para a comparação de resultados, deixando para um trabalho futuro a análise completa. Abaixo são descritas as métricas RMSE, MAE, MAPE e R^2 (*Coefficient of Determination*), principais métricas clássicas utilizadas para avaliação e comparação dos modelos.

O RMSE fornece uma média do erro global durante todo o período de previsão (ZHANG et al., 2015), é dado por:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{g}_i - g_i)^2}$$

onde, g_i é a potência de geração no passo i , \hat{g}_i é a potência estimada pelo modelo de previsão no passo i , e N é o número de pontos estimados pelo modelo.

O MAE é uma métrica muito utilizada para problemas de regressão. Assim como o RMSE, é uma métrica de erro global, mas que não penaliza os valores extremos (ZHANG et al., 2015). É dada pela equação:

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{g}_i - g_i|$$

Já o MAPE pode ser utilizado para comparar resultados com diferentes escalas temporais e espaciais de erro de previsão. É dado pela equação:

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{\hat{g}_i - g_i}{g_0} \right|$$

onde g_0 é a capacidade da usina.

Por fim o R^2 , representa a correlação entre as previsões e os valores reais (ANTONANZAS et al., 2016). É dado pela equação:

$$R^2 = 1 - \frac{Var(\hat{g}_i - g_i)}{Var(g_i)}$$

As métricas serão utilizadas como objeto de comparação para o modelo proposto com resultados encontrados na literatura.

2.6 CONSIDERAÇÕES FINAIS

Este capítulo apresentou os principais fatores que influenciam a geração de energia PV. Como esperado a irradiância solar possui a maior correlação, porém é afetada por diversos fatores como a composição atmosférica, o ângulo zenital, a nebulosidade e a cobertura de nuvens. Nas metodologias para a predição de geração de energia PV descritas, é válido destacar os métodos combinados que utilizam do melhor de cada técnica para aumentar a robustez dos modelos. Por fim, foram apresentadas metodologias para avaliação dos modelos levando em consideração as métricas clássicas e métricas específicas, porém para uma comparação dos resultados com outros autores as métricas clássicas se destacam em especial o RMSE pois é utilizado pela maiorias dos trabalhos analisados.

3 IMAGENS DE SATÉLITE

O estudo de imagens de satélite, um dos focos deste trabalho, é uma área de conhecimentos muito específicos e densos, desta forma, este capítulo aborda os principais conceitos sobre satélites para o desenvolvimento deste trabalho e aprofunda o estudo no satélite GOES-16 e seus produtos, que será o provedor das imagens utilizadas para as predições de geração de energia fotovoltaica (PV).

Na Seção 3.1, é apresentada a evolução dos satélites imageadores, suas aplicações e classificações, também é abordado de forma rápida alguns conceitos sobre nuvens. Na Seção 3.2 é aprofundado o estudo nos satélites da série *Geostationary Operational Environmental Satellite* (GOES), em especial o GOES-16, tratando dos sensores embarcados neste satélite, os produtos entregues e suas principais aplicações. Ao final na Seção 3.3, são apresentadas as considerações sobre o estudo para a aplicação deste trabalho.

3.1 INTRODUÇÃO A IMAGENS DE SATÉLITE

Os estudos relacionados aos satélites orbitais tiveram início no final do século XIX e início do século XX. Seu grande marco no início na década de 50, mais precisamente em 4 de outubro de 1957, foi o lançamento e operação do primeiro dispositivo criado pelo homem a orbitar a terra, o Sputnik. O primeiro satélite do mundo, do tamanho de uma bola de basquete lançado pelos russos, que deu início a corrida espacial (EMERY; CAMPS, 2017).

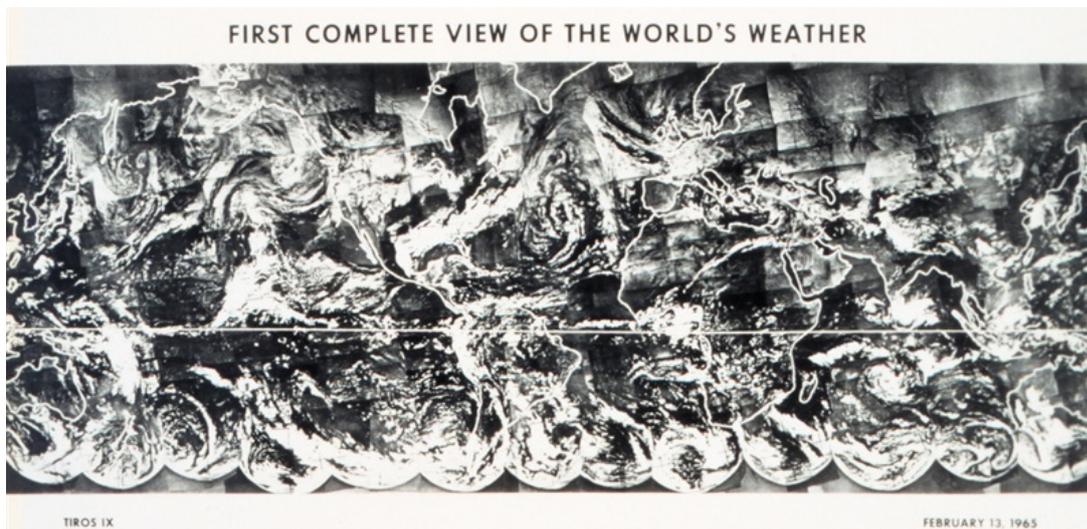
O satélite russo não contava com nenhum instrumento capaz de observar a terra, isto só foi possível em 1960 com o lançamento do *Television and Infrared Observation Satellite* (TIROS). A tecnologia de estabilização adotada pelo TIROS permitia a orientação do sensor para a terra, possibilitando coletar imagens orbitais a uma altura de 750km (EMERY; CAMPS, 2017).

Em 1965, uma iniciativa reuniu 450 fotografias para formar a primeira imagem completa do clima mundial, apresentada na Figura 3.1 (NOAA, 2018). Utilizando a evolução do satélite TIROS, o TIROS IX. Do TIROS até os dias de hoje ocorreram grandes evoluções em toda a tecnologia de satélites, que vêm sendo empregados em diversas aplicações.

Os satélites podem ser classificados de acordo com a sua órbita espacial. A órbita determina os parâmetros de operação e atuação quanto à altura, inclinação, excentricidade, direção e sincronicidade. Muitas das aplicações de satélites imageadores estão dentro de duas categorias, os geoestacionários e os polares (CHUVIECO; HUETE, 2009).

Os satélites polares orbitam a terra tipicamente a uma altura de 700 a 900 km, está em constante movimento quando referenciado a terra. São denominados polares

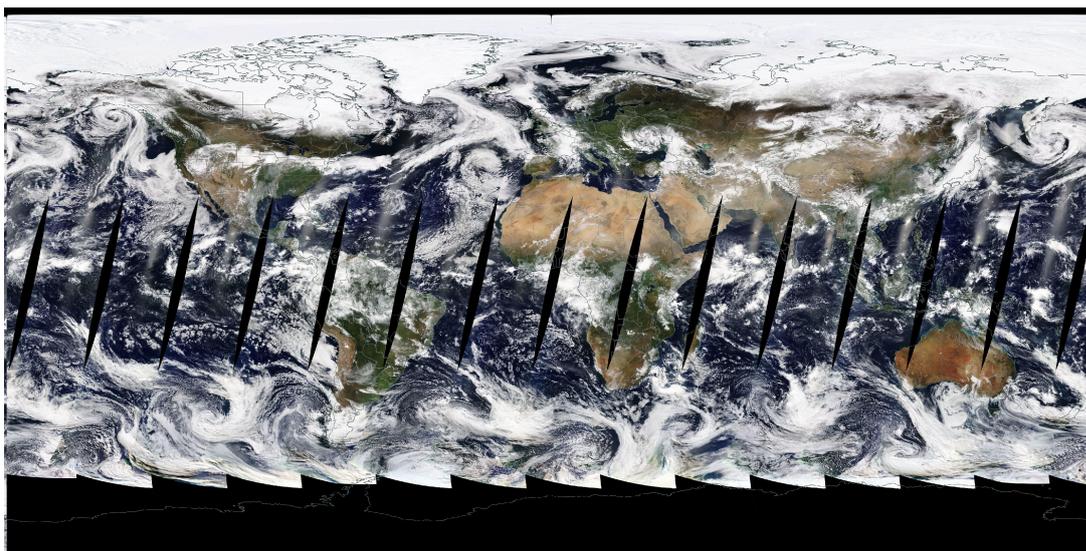
Figura 3.1 – Primeira visualização completa do clima da terra.



Fonte: Acervo de fotos NOAA (2018).

pois passam próximos dos polos. Sua periodicidade¹ varia de horas a dias, dependendo da sua altura, permitindo gerar imagens de todo o globo (Terra) (CHUVIECO; HUETE, 2009). Um exemplo da imagem gerada por este satélite pode ser vista na Figura 3.2, note que a imagem é gerada através de um ciclo contínuo de movimento orbital ao redor da Terra.

Figura 3.2 – Imagem gerada pelo satélite Terra no dia 13 de maio de 2018.



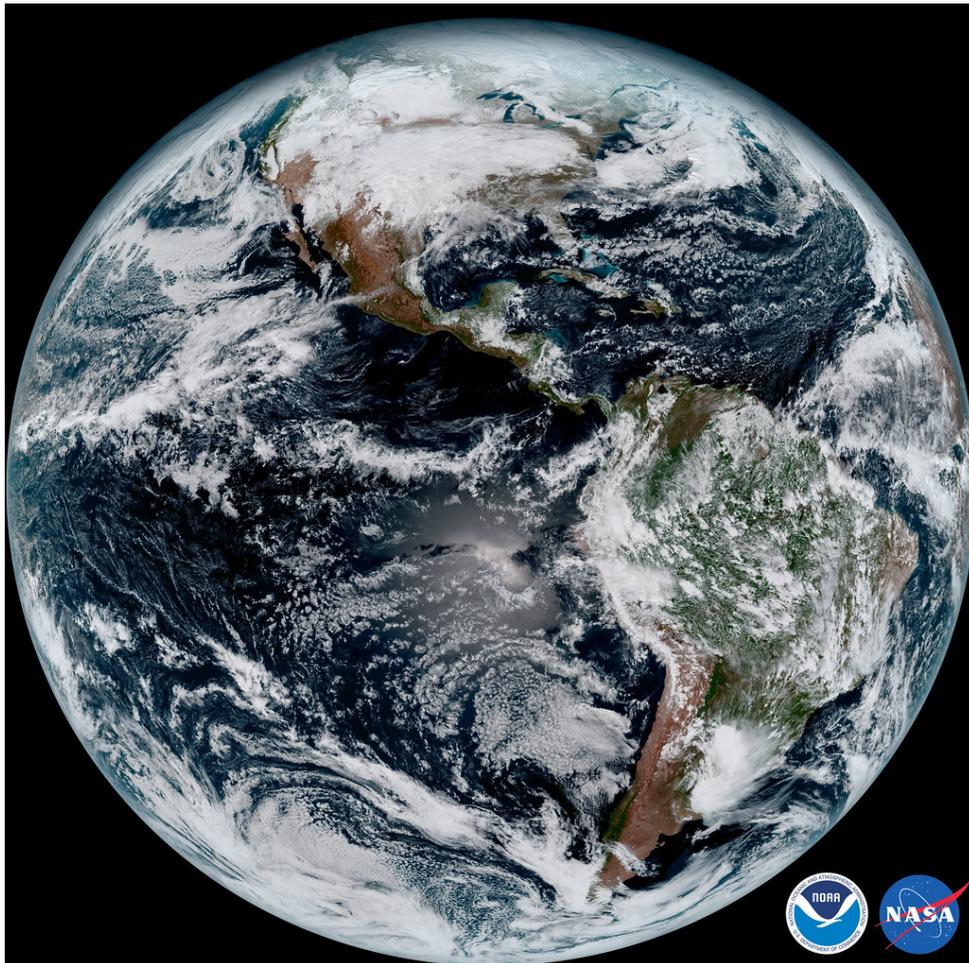
Fonte: Acervo de fotos NASA (2022).

Diferente dos satélites orbitais, os satélites geoestacionários orbitam a uma al-

¹ Intervalo de tempo que leva para o satélite fazer uma volta completa ao redor da terra.

tura de 36.000 km da terra, ele acompanha o seu movimento de rotação na linha do equador. Portanto, o período de rotação do satélite é o mesmo que o da Terra (CHUVIECO; HUETE, 2009). Satélites desta categoria estão sempre parados em relação a um ponto da superfície, ou seja, seus sensores capturam informações de um mesmo ponto geográfico. Um exemplo pode ser visto na Figura 3.3.

Figura 3.3 – Imagem gerada pelo satélite GOES-16 no dia 15 de janeiro de 2017.



Fonte: Acervo de fotos NOAA (2021a).

Uma das aplicações mais comuns de satélites é a observação da terra, para isso são equipados com sensores denominados imageadores, capazes de gerar fotografias espaciais. Os sensores que são utilizados para a aquisição de imagens atuam através da coleta de dados físicos sem contato, por isso sua operação é denominada de sensoriamento remoto. Seu modo de operação é baseado no princípio de detecção de energia, no caso dos imageadores eles atuam na leitura da energia solar refletida pela superfície da terra e na leitura da energia emitida pela própria terra (CHUVIECO; HUETE, 2009).

Uma das categorias de satélite imageador geostacionário é o satélite mete-

orológico. Estes dispositivos têm o objetivo de monitorar a superfície do globo para a geração de dados relacionados a temperatura da superfície do mar, velocidade do vento, vapor de água, índice de vegetação, cobertura de nuvens, neve e gelo, dentre outros (KISHTAWAL, 2004). Atualmente vários satélites geoestacionários cobrem a superfície da terra, podendo citar GOES-16 E 17², Himawari 8³, Meteosat 8 a 11⁴, dentre outros.

Devido a importância para este trabalho, será feito um estudo aprofundado sobre os satélites da série GOES, em especial o GOES-16 que será apresentado na Seção 3.2. A série GOES é um conjunto de satélites meteorológicos operados pelo *National Oceanic and Atmospheric Administration* (NOAA) em parceria com a *National Aeronautics and Space Administration* (NASA) e controlados pelos Estados Unidos. O imageador do GOES-16 captura imagens de quase todas as Américas, conforme apresentado na Figura 3.3.

Dentre as características dos satélites meteorológicos, uma que se destaca para este estudo é o monitoramento de nuvens. Esta tarefa é feita através da combinação dos diversos espectros de irradiação, lidos pelo imageador, refletidos pelas diferentes camadas das nuvens e pelo solo. As diferentes faixas espectrais do imageador são denominadas canais, e para esta aplicação são combinadas através de algoritmos para estimar algumas características das nuvens como a posição (HEIDINGER; STRAKA, 2012), a altura (HEIDINGER, 2012) e a composição das nuvens (WALTHER; STRAKA; HEIDINGER, 2013).

Para facilitar o entendimento destes componentes é apresentada a Figura 3.4 e discutida abaixo a composição das camadas da Terra que reflete nas diferentes formas de nuvens que serão tratadas, com base em NEDIS (2016).

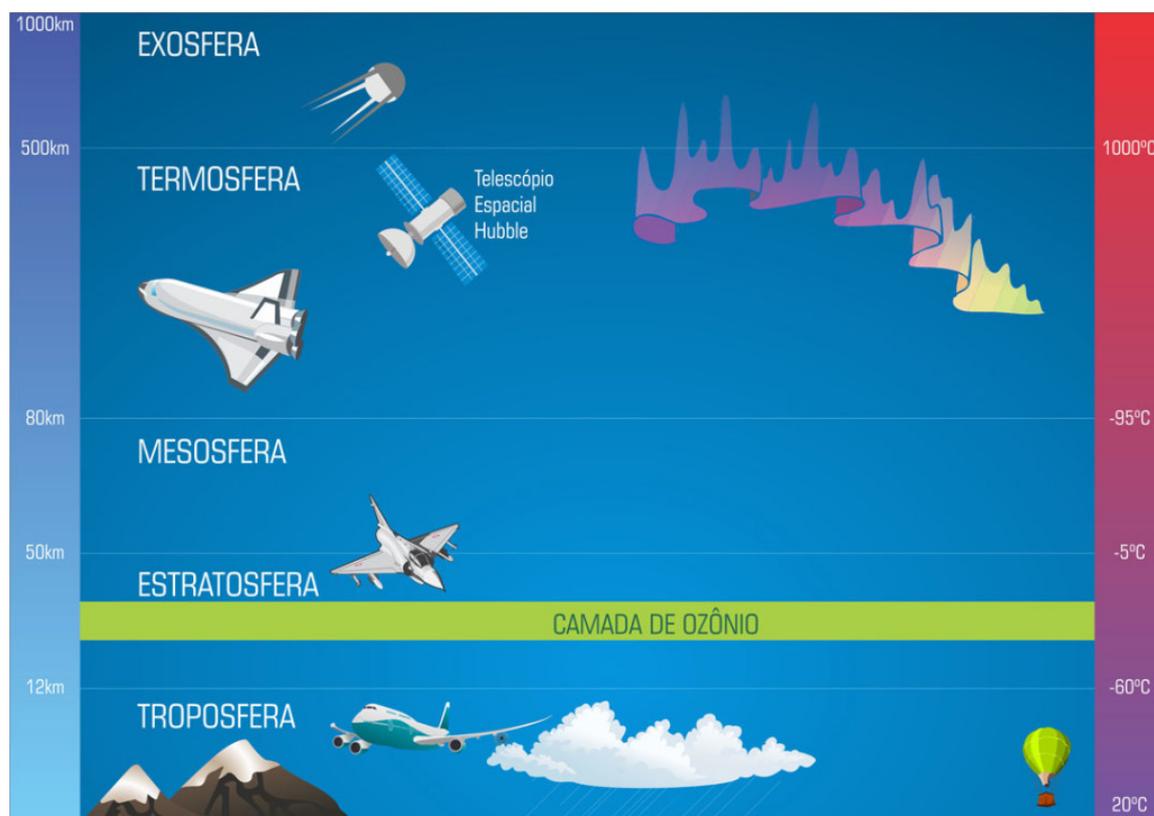
A troposfera concentra a maior quantidade de gases indispensável para os seres vivos e nela encontramos quase todas as nuvens que vemos, é nesta camada que circulam os aviões comerciais. A estratosfera possui fortes correntes de ar, por este motivo os aviões comerciais, quando possível, voam abaixo dela, é neste espaço que se concentra a camada de ozônio. A mesosfera é considerada um dos lugares mais frios da terra, nesta camada são formadas as nuvens noctilucentes compostas por pequenos cristais de gelo, esta camada é onde a maioria dos meteoros queimam ao entrar na atmosfera, é a elevação mais alta que uma nuvem pode se formar. A termosfera é onde a radiação ultravioleta e de raio X de alta energia começam a ser absorvidas devido à forte ionização presente na termosfera, é nela que ocorre o fenômeno das auroras. Por fim, a exosfera ocorre a transição entre a atmosfera e o espaço, nesta camada os gases se comportam de forma diferente das outras camadas, os átomos e as moléculas estão tão distantes que podem viajar centenas de quilômetros

² Satélite meteorológico operado pelo NOAA

³ Satélite meteorológico operado pela *Japan Meteorological Agency*

⁴ Satélite meteorológico operado pela Eumetsat

Figura 3.4 – Composição das camadas externas da terra.



Fonte: Vidraria de Laboratório (2015)

sem colidir, nesta camada encontramos uma grande quantidade de satélites em órbita baixa.

3.2 GOES-16

Os satélites da linha GOES tiveram início em 1975 e vem evoluindo desde o primeiro, o GOES-A, nesta seção vamos iniciar a discussão pelo GOES-R mas os detalhes da evolução dos satélites são apresentados no Apêndice A.1. O GOES-R⁵ é um satélite desenvolvido para operar por 10 anos em órbita precedidos por até 5 anos de armazenamento (CORPORATION, 2011b). Seu desenvolvimento dá início a uma nova série de satélites, a série GOES-R, que inclui o GOES-R, GOES-S, GOES-T e GOES-U (NOAA; NASA, 2021b). É conhecida como os satélites meteorológicos mais avançados dos Estados Unidos da América (EUA) e estão atualmente em órbita o GOES-R nomeado de GOES-16 e o GOES-S nomeado como GOES-17 (NOAASIS,

⁵ Antes de um satélite da série GOES entrar em órbita de operação, ele é nomeado com uma letra. Somente ao entrar em operação sem nenhum problema ele recebe um número. Como é o caso do GOES-G, que em 3 de maio de 1986, seria nomeado de GOES-7, mas foi perdido quando seu veículo foi atingido por um raio após a decolagem.

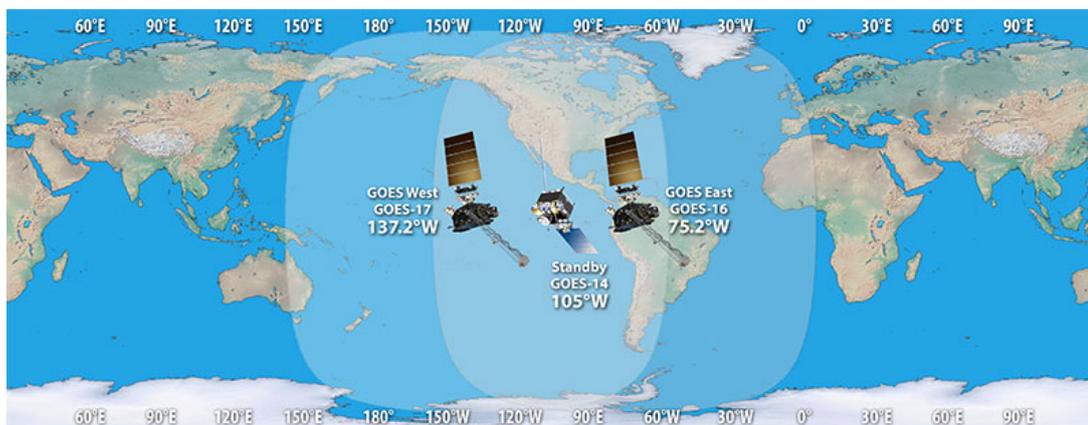
2019; CORPORATION, 2011b). Dentre as melhorias desenvolvidas pela nova geração podem ser citadas, de acordo com NOAA e NASA (2021b):

- Melhoria da trilha de furacões e previsões de intensidade
- Aumento do tempo para o aviso de tempestade e tornado
- Previsão dos riscos de queda de raio
- Melhor detecção de chuvas fortes e riscos de enchentes
- Melhor monitoramento de fumaça e poeira
- Avisos e alertas de qualidade do ar aprimorados
- Melhor detecção de incêndio e estimativa de intensidade
- Detecção aprimorada de nuvens baixas e neblina
- Melhoria da segurança do transporte e planejamento da rota da aviação
- Aviso aprimorado para interrupções de comunicação, navegação e cortes de energia
- Monitoramento mais preciso de partículas energéticas responsáveis pelos riscos de radiação

O sistema GOES é composto atualmente por dois satélites o GOES-16, operando como GOES-East a 75.2° de longitude oeste e o GOES-17, operando como GOES-West a 137.2° de longitude oeste (OSPO, 2022; NOAA; NASA, 2019). Estes satélites ajudam os meteorologistas a observar e prever eventos climáticos locais, incluindo tempestades, tornados, nevoeiros, furacões, inundações repentinas e outras condições climáticas severas. Além disso, as observações do GOES têm se mostrado úteis no monitoramento de tempestades de poeira, erupções vulcânicas e incêndios florestais (HARRIS, 2011). Conforme é demonstrado na Figura 3.5, é possível verificar a órbita dos satélites, a sua área de cobertura e um terceiro satélite (GOES-14). O programa conta com dois satélites ativos, porém permanece um satélite como redundância para os demais, atualmente este satélite é o GOES-14 (NOAASIS, 2019).

O GOES-16 foi lançado no dia 19 de novembro de 2016 na base de Cabo Canaveral na Flórida. A bordo de um foguete Atlas V 541 com o objetivo de supervisionar a América do Norte e do Sul, Oceano Atlântico e a costa oeste da África. Embarcados nos satélites da série GOES-R estão os instrumentos: ABI, EXIS, GLM, MAG, SEISS e SUVI sendo o ABI o instrumento de interesse deste trabalho e descrito na Seção

Figura 3.5 – Posição espacial e cobertura do GOES-East e GOES-West.



Fonte: NOAA e NASA (2021b)

3.2.1. Grande parte dos dados gerados pelo GOES-16 estão disponíveis⁶ em tempo de processamento.

O satélite conta com dois instrumentos para monitorar o sol: O *Extreme Ultraviolet and X-ray Irradiance Sensors* (EXIS) são sensores de radiação ultra violeta extrema e auxiliam no monitoramento do sol para compreender seus efeitos como radiação e explosões solares (ATMOSPHERIC; PHYSICS, 2011) e o *Solar Ultraviolet Imager* (SUVI) um imageador que observa o sol na faixa do ultravioleta extremo, sua aplicação está relacionada a observações de erupções e explosões solares, os seus efeitos e compreender as áreas ativas do sol (CENTER, 2011c).

Para monitorar o ambiente espacial o GOES-16 conta com dois instrumentos dedicados: O *Magnetometer* (MAG) coleta medidas de campo magnético espacial e fornece dados sobre atividade geomagnética, detecção de tempestades magnéticas e aplicações em pesquisa (CENTER, 2011b) e o *Space Environment In-Situ Suite* (SEISS) composto por quatro sensores que monitoram o fluxo de prótons, elétrons e íons pesados na magnetosfera, sua aplicação está relacionada a avaliação de risco de descarga eletrostática e perigo de radiação para astronautas (CORPORATION, 2011a).

Observando a terra o satélite conta com dois instrumentos o *Advanced Baseline Imager* (ABI) descrito na seção seguinte e o *Geostationary Lightning Mapper* (GLM) um detector óptico transiente. Possui uma funcionalidade única, detectar as variações ópticas momentâneas, isto permite que este equipamento seja utilizado para indicar a presença de raios e relâmpagos, sendo esta sua aplicação (CENTER, 2011a).

⁶ <https://noaa-goes16.s3.amazonaws.com/index.html>

3.2.1 Advanced Baseline Imager - ABI

O imageador ABI é o principal instrumento dos satélites da série GOES-R. É responsável por fotografar a terra em 16 diferentes canais espectrais, cada canal com um diferente comprimento de onda, sendo elas, dois canais visíveis, quatro canais de infravermelho próximo e dez canais de infravermelho (HARRIS, 2011). A combinação de diferentes canais pode fornecer informações como rastrear e monitorar a formação de nuvens, movimento atmosférico, convecção, temperatura da superfície terrestre, dinâmica do oceano, fluxo de água, fogo, fumaça, plumas de cinzas vulcânicas, aerossóis, qualidade do ar e saúde vegetal (NOAA, 2021b). O imageador presente no GOES-R possui a capacidade de fornecer três vezes mais informações, quatro vezes a resolução espacial e uma cobertura temporal cinco vezes mais rápida que seu antecessor (HARRIS, 2011). Isso é possível por conta do radiômetro multicanal desenvolvido para operar em diferentes modos de varredura.

A área de cobertura é subdividida em quatro regiões de escaneamento *full disk*, *mesoscale 1*, *mesoscale 2*, *Continental US* (CONUS). Esta condição permite uma frequência de captura diferente para cada região, conforme apresentado na Tabela 3.1. Note que a região mesoscale pode operar em duas frequências distintas, isso porque a mesoscale 1 e 2 podem ser sobrepostas para dobrar a frequência de captura.

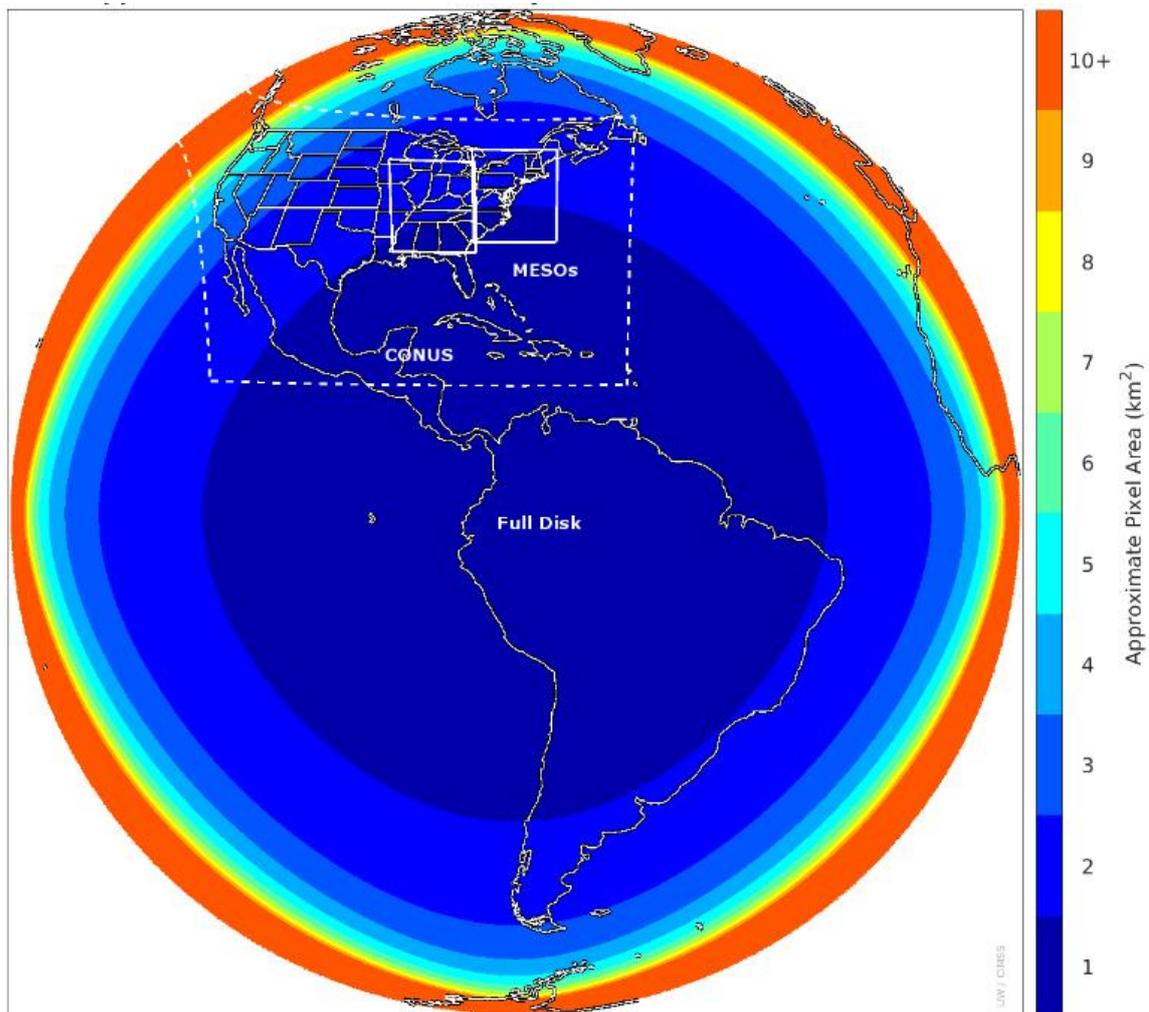
Tabela 3.1 – Tempo de captura em cada uma das regiões e de acordo com o modo de operação.

	Modo 3	Modo 6	Modo 4
<i>Full Disk</i>	15min	10min	5min
CONUS	5min	5min	-
<i>Mesoscale</i>	30s ou 60s	30s ou 60s	-

Fonte: NOAA e NASA (2011).

A cobertura de cada uma das áreas é descrita a seguir, conforme apresentado em NOAA e NASA (2021a) e ilustrada na Figura 3.6. A região *full disk* captura uma imagem que contempla toda a área de cobertura do satélite, é toda a área que contempla a Figura 3.6. O CONUS é uma região retangular de 3.000km (lat) por 5.000km (lon) que olha para todo o EUA, é representado como retângulo tracejado na Figura 3.6. As regiões *mesoscale* são áreas retangulares móveis de tamanho 1.000 km por 1.000 km, e são posicionadas para a execução de análises regionais e em caso de eventos severos. Possuem duas posições definidas, porém podem ser reposicionadas dentro do *full disk* mediante necessidade, são representadas pelos retângulos na Figura 3.6. É importante notar que a curvatura da Terra tem um efeito sobre a imagem gerada, note que a medida com que a imagem se afasta do centro e se aproxima das bordas a resolução espacial diminui.

Figura 3.6 – Posições das regiões de escaneamento do GOES-16 e efeito da curvatura sobre a resolução espacial.

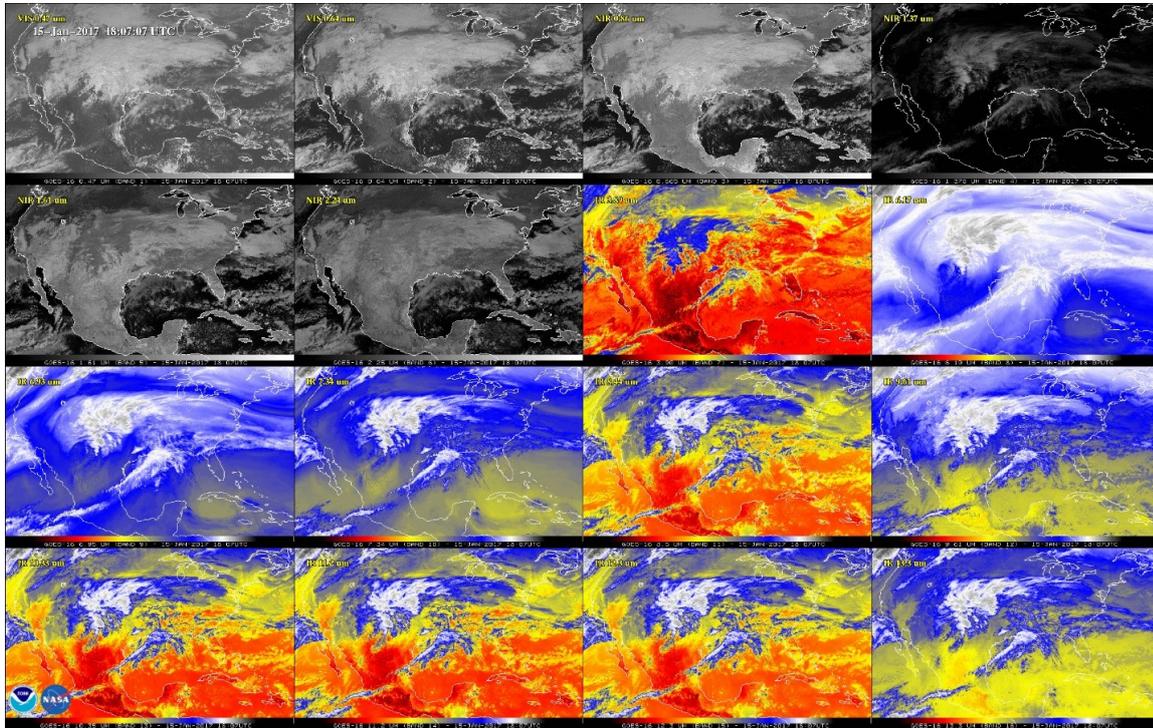


Fonte: STAR (2017).

Todos os canais do ABI têm calibração em órbita, entregando dados com alta acuracidade. Cada um dos canais fornecem informações específicas e podem ser combinados para obter informações derivadas. O imageador entrega dados de irradiância dentro da sua banda espectral em $W.m^{-2}.sr^{-1}.\mu m^{-1}$ nos canais de 1-6 e em $mW.m^{-2}.sr^{-1}.cm$ nos canais de 7-16. Na Figura 3.7 são ilustrados os 16 canais, note que, mesmo com a imagem reproduzindo a mesma região no mesmo momento, cada um dos canais entrega diferentes informações da região.

Todos os canais podem ser utilizados individualmente ou combinados para o desenvolvimento de análises, algumas combinações já são disponibilizadas pelo NOAA. Todos os dados entregues pelo satélite são denominados produtos, conforme descrito na próxima seção. A Tabela 3.2 apresenta a síntese de cada um dos canais, sua aplicação típica dividida em *Visible and Near-Infrared* (VNIR), *Mid-Wavelength Infrared*

Figura 3.7 – Primeira imagem coletada do CONUS pelo sensor ABI em seus 16 canais espectrais. Começando pelos canais 1 até a 16 da esquerda para direita, de cima para baixo.



Fonte: STAR (2020).

(MWIR) e *Long-Wavelength Infrared* (LWIR) e a resolução espacial de cada canal. No Apêndice A.2 são apresentados mais detalhes da aplicação de cada canal.

3.2.2 Produtos

Toda informação disponibilizada pelo GOES-16 é denominada de produto. É a forma de entregar para o usuário informações relevantes da aplicação e são disponibilizados em três níveis. Os níveis relacionam a quantidade de processamento aplicado a imagem, conforme apresentado em NOAA e NASA (2021a) os canais são divididos em:

- Level 0 (L0): São os dados brutos disponibilizados pelos 6 instrumentos a bordo do GOES-16. Estes são recebidos diretamente do satélite por estações terrestres;
- Level 1b (L1b): Os dados entregues pelo L1b passam por um processamento das estações terrestres, ajustando a calibração, correção geográfica e os dados para valores físicos, porém continuam entregando o resultado do instrumento. O produto do ABI entrega dados de irradiância neste nível;

- Level 2+ (L2+): Este produto entrega dados físicos do meio como a temperatura na superfície terrestre, podendo ser uma composição de diferentes sensores e produtos L1b e calculados por algoritmos para entregar seu resultado.

A NOAA disponibiliza os produtos: ABI L1b e L2+, GLM L2+, e SUVI L1b em um repositório da Amazon Web Service (AWS) no formato netCDF no endereço <https://noaa-goes16.s3.amazonaws.com/index.html> (NOAA; NASA, 2021a). Os produtos estão disponíveis em tempo de processamento, ou seja, o tempo de recebimento da imagem do satélite, processamento e carregamento no repositório. No Apêndice A.3 estão listados os produtos entregues na AWS. Para o propósito deste trabalho são considerados apenas os produtos do imageador ABI em *full scale*.

Os produtos ABI-L1b-RadF e ABI-L2-CMIPF são considerados os de maior relevância para este estudo, pois vão fornecer os dados do imageador ABI para o processamento da rede. O primeiro, entrega a irradiância emitida pelos 16 canais espectrais do instrumento, os dados são disponibilizados em irradiância nas unidades $W.m^{-2}.sr^{-1}.\mu m^{-1}$ nos canais de 1-6 e em $mW.m^{-2}.sr^{-1}$ nos canais 7-16. Já o se-

Tabela 3.2 – Canais espectrais disponíveis no imageador ABI e suas aplicações típicas.

	Canal	Comprimento de onda (μm)	Resolução Espacial (km)	Aplicação típica
VNIR	01	0.47	2,0	Aerossol diurno sobre a terra, mapeamento de águas costeiras
	02	0.64	0,5	Nuvens diurnas, nevoeiro, fluxo solar e ventos
	03	0.865	1,0	Vegetação diurna / cicatriz de queimadura e aerossol sobre a água, ventos
	04	1.378	2,0	Nuvens cirros diurnas
	05	1.61	1,0	Fase diurna do topo da nuvem e tamanho de partícula, neve
MWIR	06	2.25	2,0	Propriedades diurnas da terra / nuvem, tamanho das partículas, vegetação, neve
	07	3.90	2,0	Superfície e nuvem, nevoeiro à noite, fogo, ventos
	08	6.185	2,0	Vapor de água atmosférico de alto nível, ventos, chuva
	09	6.95	2,0	Vapor de água atmosférico de nível médio, ventos e chuva
	10	7.34	2,0	Ventos e vapor d'água de nível inferior
	11	8.5	2,0	Água total para estabilidade, fase de nuvem, poeira e chuva
LWIR	12	9.61	2,0	Ozônio total, turbulência e ventos
	13	10.35	2,0	Propriedades de superfície e nuvem
	14	11.2	2,0	Temperatura da superfície do mar, chuva e propriedades da nuvem
	15	12.3	2,0	Água total, cinzas, temperatura da superfície do mar e propriedades da nuvem
	16	13.3	2,0	Temperatura atmosférica e alturas das nuvens

Fonte: Adaptado de NOAA e NASA (2019).

gundo, são as mesmas informações, porém já tratadas para entregar valores de brilho nos canais de 1-6 e em temperatura nos canais de 7-16.

Já os produtos ABI-L2-DSRF, ABI-L2-AODF, ABI-L2-CODF e ABI-L2-ACMF possuem relação com o estudo e podem ser utilizados para otimização do modelo. Onde, o ABI-L2-DSRF entrega uma estimativa da quantidade total de radiação de ondas curtas diretas e difusas que atinge a superfície da terra. Ele é utilizado para a previsão em tempo real e de curto espaço de tempo da geração de energia solar. Já o ABI-L2-AODF utiliza vários comprimentos de ondas para medir a refletância de pontos livres de nuvens. O dado entregue é a taxa de refletância em um determinado ponto.

O ABI-L2-CODF representa a profundidade óptica de uma nuvem. O algoritmo atua na estimativa do topo e da base da coluna atmosférica para estimar o valor. Por fim, o ABI-L2-ACMF entrega uma máscara de nuvens e a classifica em quatro estados: nublado, provavelmente nublado, claro ou provavelmente claro.

3.3 CONSIDERAÇÕES FINAIS

Fica claro neste capítulo a importância dos dados gerados pelos satélites meteorológicos para a previsão da geração de energia solar em painéis fotovoltaicos. A visão destes dispositivos permite o acompanhamento de condições climáticas que influenciam diretamente na dinâmica de geração. Alguns produtos disponibilizados pelo GOES-16 são relacionados ao objetivo desta pesquisa, sendo os principais utilizados neste trabalho o ABI-L1b-RadF e o ABI-L2-CMIPF, os demais produtos mencionados não possuem dados para o período de previsão proposto, tendo iniciado sua operação posteriormente.

4 MACHINE LEARNING

Sempre foi interesse do homem criar máquinas que pensam (GOODFELLOW; BENGIO; COURVILLE, 2016). É um fato que os computadores processam de forma diferente do ser humano, sendo o cérebro um organismo complexo capaz de efetuar processamento extremamente complexo, não linear e de forma paralela (HAYKIN, 2009). O ser humano aprende pela experiência, não precisamos saber a definição científica de uma árvore para conseguirmos identificá-la, nós simplesmente a identificamos. Nosso sistema cognitivo aprende através de dados, nós sabemos o que é uma árvore baseado em milhares de árvores que vimos ao longo da nossa vida (ABU-MOSTAFA; MAGDON-ISMAIL; LIN, 2012).

O aprendizado de máquina busca replicar esta capacidade através de modelos matemáticos e estatísticos tradicionalmente com base em dois tipos de paradigmas de aprendizado: o supervisionado e o não supervisionado. No supervisionado, o algoritmo de aprendizado é alimentado com informações onde as respostas são conhecidas. Nesta metodologia, o sistema busca ajustar os parâmetros para que um determinado conjunto de entradas gere uma saída determinada. Este processo é executado exaustivamente para toda a base de dados, até que o modelo reduza seu erro de saída, se aproximando o máximo possível da saída determinada, isto é o que convencionalmente se chama de aprendizado (HAYKIN, 2009). Na segunda abordagem, não supervisionada, o algoritmo possui uma base de dados de entrada, porém não relaciona com nenhuma saída, sendo ele o responsável por encontrar padrões subjacentes a esses dados.

Neste capítulo serão abordados tópicos relacionados ao aprendizado de máquina que são de interesse deste trabalho. Iniciando pela análise de dados na Seção 4.1. A discussão se aprofunda em redes neurais na Seção 4.2 abordando as *Multi-layer Perceptron* (MLPs) na Seção 4.2.1 e *Convolutional Neural Network* (CNNs) na Seção 4.2.2. Na Seção 4.3, são apresentados trabalhos relacionados, que utilizam de *Artificial Neural Network* (ANNs) dentro do contexto proposto neste trabalho. Ao final, na Seção 4.4, são apresentadas as considerações finais.

4.1 PREPARAÇÃO DOS DADOS

Técnicas de aprendizado de máquina supervisionado utilizam de dados para mapear a relação entre a entrada e a saída. Portanto, para obter bons resultados é importante ter uma boa base de dados. Yu, Wang e Lai (2006) sugerem uma abordagem de três níveis, a pré-análise dos dados, o pré-processamento e a pós-análise. Na primeira etapa é feito um levantamento dos pré-requisitos do problema, determinando quais dados podem ser utilizados para a análise. Com um conjunto de dados determinados, a metodologia segue com a coleta dos dados e seleção de atributos, neste

momento é feito um estudo de quais variáveis possuem maior relação com o problema, a fim de determinar quais serão parte do treinamento. Finalizando a primeira etapa, os dados provenientes de bases diferentes precisam ser combinados em um único conjunto, no caso das séries temporais, os dados são indexados e combinados com base no tempo.

As bases de dados brutas, obtidas diretamente dos sensores e armazenadas por *Data Loggers*, não estão blindadas contra erros, é comum encontrarmos ausência de dados, medidas discrepantes, erros de armazenamento e operação dos equipamentos. Isto porque os sistemas podem sofrer interferências que afetam algumas leituras, gerando problemas de falta de informação, dados em escala errada e erro de sincronização (WITTEN; FRANK, 2005). Para contornar o problema, Yu, Wang e Lai (2006) indicam o pré-processamento dos dados para analisar e entender quais seriam seus valores possíveis, descartando os erros de leitura. No caso da velocidade do vento e irradiância, podemos utilizar de dados históricos para determinar estes valores máximos e mínimos e limitar a amplitude das leituras. Já o caso da refletância, temos a informação na documentação do satélite, onde os valores variam entre 0 e 1, neste caso, os dados fora de escala podem ser tratados como erro. A falta de dados, é um outro problema comum e pode ser tratado de duas formas, caso a base de dados seja ampla e não comprometa o estudo, os dados ausentes podem ser excluídos da análise, caso contrário, podem ser utilizadas de técnicas estatísticas para estimar os dados faltantes, analisando os demais dados da base. Alguns sensores, como é o caso do satélite, além de fornecer os dados da leitura, fornecem a qualidade do dado, esta informação pode ser utilizada como métrica para descartar dados com baixo índice de confiança.

Durante o processo de treinamento das redes neurais a variação das escalas dos dados costuma levar a instabilidade (YU; WANG; LAI, 2006), para isso os dados são tipicamente normalizados em intervalos de -1 a 1 ou 0 a 1. Além da instabilidade a normalização costuma melhorar o desempenho do modelo principalmente quando os dados estão em diferentes escalas. Para Yu, Wang e Lai (2006) a normalização é uma abordagem simples e eficaz, dada por

$$I = I_{min} + \frac{(I_{max} - I_{min}) \times (D - D_{min})}{D_{max} - D_{min}} \quad (4.1)$$

onde I é o dado normalizado, I_{max} e I_{min} é o valor máximo e mínimos da normalização, D é o dado a ser normalizado, D_{max} e D_{min} é o máximo e mínimo do conjunto de dados a ser normalizado.

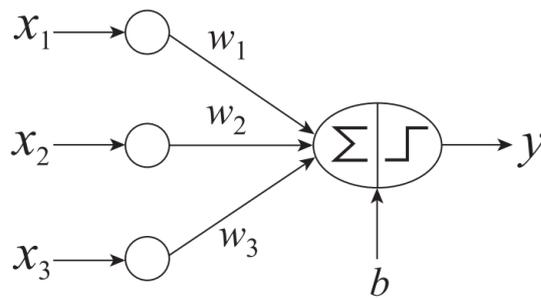
Como último passo, a pós-análise, se concentra na divisão dos dados. O processo de treinamento da rede neural normalmente é dividido em três conjuntos, sendo eles, treinamento, validação e teste (YU; WANG; LAI, 2006). Os dados de treinamento são utilizados para o aprendizado da rede. Os dados de validação são dados utilizados

para ajustar os hiper parâmetros, por fim os dados de teste, são os dados utilizados para a avaliação final.

4.2 ARTIFICIAL NEURAL NETWORKS - ANN

As ANNs são em sua maioria modelos de aprendizado capazes lidar problemas extremamente complexos e não lineares, utilizando conjuntos de dados. São compostas por unidades de processamento chamadas neurônios artificiais, estes podem ser interligados entre si formando as redes neurais (TAN et al., 2019). Uma das ANNs mais simples, composta de um único neurônio artificial, capaz de desempenhar uma função modular, é chamada de Perceptron. A arquitetura básica do Perceptron é apresentada na Figura 4.1,

Figura 4.1 – Modelo do Perceptron.



Fonte: TAN et al. (2019)

onde x_1 , x_2 e x_3 correspondem a entrada da rede e y é uma saída binária da ANN. Note que para cada x_n existe uma conexão correspondente w_n que determina o peso desta conexão na entrada do neurônio. A somatória das entradas com ajuste do peso é somada a um fator b , chamado *bias*, sendo este e os pesos das conexões as variáveis de treinamento do modelo. Por fim o valor passa por uma função de ativação, esta analisa a somatória e determina se o neurônio será ativado ou desativado (TAN et al., 2019). O modelo matemático do Perceptron segue abaixo

$$\hat{y} = \begin{cases} 1, & \text{se } \mathbf{w}^T \mathbf{x} + b > 0 \\ -1, & \text{caso contrário.} \end{cases}$$

onde \hat{y} é a saída. Para simplificar a equação podemos considerar $\tilde{\mathbf{w}} = (\mathbf{w}^T b)^T$, e $\tilde{\mathbf{x}} = (\mathbf{x}^T 1)^T$, desta forma \hat{y} pode ser escrito como

$$\hat{y} = \text{sign}(\tilde{\mathbf{w}}^T \tilde{\mathbf{x}})$$

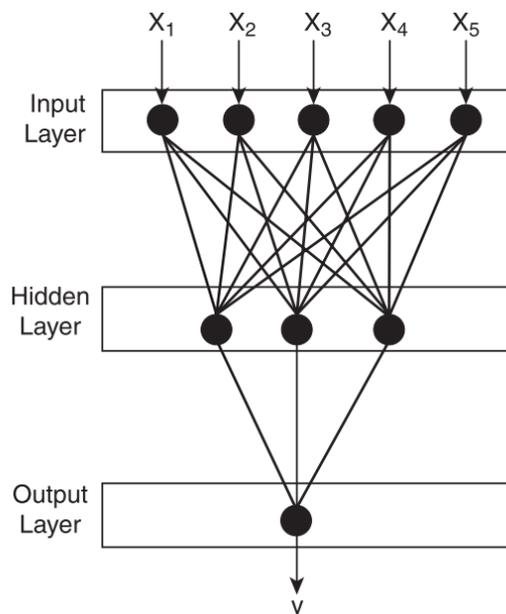
onde *sign* atua como função de ativação, retornando 1 quando o argumento for positivo e -1 quando o argumento for negativo.

A conexão destes neurônios em camadas e/ou em paralelo permite a criação de diversas arquiteturas de rede que são utilizadas para resolver os mais diversos problemas. Nas seções seguintes são apresentadas as MLPs e as CNN, que são de interesse deste estudo.

4.2.1 Multilayer Perceptron - MLP

As MLPs são agrupamentos de neurônios em camadas, conforme Figura 4.2. A primeira camada é chamada de entrada, o seu número de neurônios é determinado pelo número de parâmetros x_n que descrevem o problema. As camadas intermediárias são chamadas de camadas ocultas, responsáveis pelo processamento dos dados, o número de neurônios por camada e o número de camadas vão depender da complexidade do problema. Por fim, temos o resultado y , que possui a função de sintetizar as predições. O número de neurônios de saída vai ser determinado pelo tipo de informação que se espera, em casos de classificação podemos ter n saídas, sendo uma saída por classe, no caso de regressão de uma variável, é comum termos uma saída (TAN et al., 2019).

Figura 4.2 – Exemplo de MLP, onde x_1 a x_5 são as entradas e y a saída.

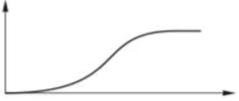
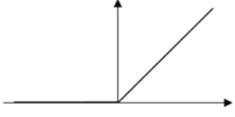


Fonte: TAN et al. (2019)

Cada neurônio da rede passa por um processo de ativação, uma função que determina como os valores serão entregues para a próxima etapa. Este parâmetro é definido durante o processo de modelagem da rede e é determinado de acordo com

o problema que será abordado. As funções de ativação são modelos matemáticos lineares e não lineares (TAN et al., 2019), as mais comuns são apresentadas na Tabela 4.1. A modelagem da rede MLP se dá pela definição do seu número de camadas, o número de neurônios em cada camada e a função de ativação dos neurônios, e deve ser levado em consideração a complexidade e o tipo do problema que se espera resolver. É comum o desenvolvimento de testes de validação de modelo com diversas configurações de rede para determinar a melhor configuração.

Tabela 4.1 – Funções de ativação.

Nome	Equação	Representação Gráfica
Sigmoid	$f(u) = \frac{1}{1 + \exp(-u)}$	
ReLU (Rectified Linear Units)	$f(u) = \begin{cases} u, & \text{se } u > 0 \\ 0, & \text{caso contrário.} \end{cases}$	
Hyperbolic tangent sigmoid (tanh-sig)	$f(u) = \tanh(u)$	

Fonte: Adaptado de Das et al. (2018).

Um dos algoritmos de treinamento mais utilizados é o *backpropagation*, ao aplicar uma entrada conhecida $x(n)$, onde n é o tamanho do vetor de entrada, em par com uma saída conhecida $y(n)$, onde n é o tamanho do vetor de saída, para o ajuste dos parâmetros da rede (pesos e *bias*), isto ocorre de forma iterativa em duas etapas. Primeiramente é aplicado um conjunto de dados na entrada, computando através da sua arquitetura uma saída. No segundo passo o resultado é comparado com a saída esperada e então computando o gradiente do erro de predição δ , o objetivo é otimizar a rede para convergir com o menor erro possível. Utilizando o gradiente do erro é aplicada uma correção ajustada pela taxa de aprendizado η e a saída do neurônio y o que gera uma correção $\Delta\omega$ para cada um dos pesos do neurônio. Este processo é executado para todas as camadas da rede propagando o erro da saída para entrada (HAYKIN, 2009). Os dois passos caracterizam uma iteração, esta se repete exaustivamente até obter o erro aceitável para o problema, quando convergente. Desta forma podemos definir o processo de ajuste do peso sináptico como sendo

$$\begin{pmatrix} \text{Correção} \\ \text{do peso} \\ \Delta\omega(n) \end{pmatrix} = \begin{pmatrix} \text{Taxa de} \\ \text{aprendizado} \\ \eta \end{pmatrix} \times \begin{pmatrix} \text{Gradiente} \\ \text{de erro local} \\ \delta(n) \end{pmatrix} \times \begin{pmatrix} \text{Entrada do} \\ \text{neurônio } j \\ y(n) \end{pmatrix}$$

onde $\Delta\omega_{ji}(n)$ é a correção do peso sináptico, i o número de entradas do neurônio, j o neurônio da camada, n um conjunto de dados com entrada $x(n)$ e saída $y(n)$, η a taxa

de aprendizado, $\delta_j(n)$ o gradiente de erro local do neurônio j no conjunto n e $y_i(n)$ é o a saída do neurônio j pelo estímulo da entrada $x(n)$.

Portanto o algoritmo de *back propagation* entrega uma aproximação para a otimização computando os pesos das sinapses utilizando o método da descida mais acentuada, buscando a forma rápida e menos custosa de minimizar o erro. O processo iterativo leva em conta um parâmetro muito importante, a taxa de aprendizado. Embora uma taxa de aprendizado alta possa acelerar o processo de aprendizado, reduzindo o número de iterações, as grandes mudanças do peso sináptico podem levar a instabilidade (HAYKIN, 2009).

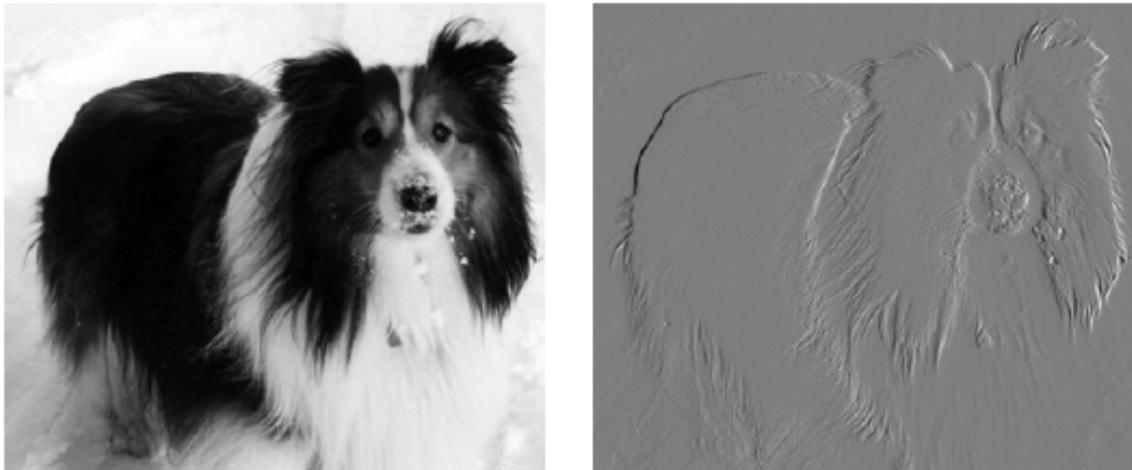
4.2.2 Convolutional Neural Network - CNN

As MLPs tratadas na seção anterior utilizam de todos os atributos individuais para descrever sua interação com a saída, ou seja, toda entrada interage com toda saída. Por exemplo, ao processar uma imagem, em uma MLP todos os pontos serão uma entrada da rede, cada variação mínima da imagem vai gerar uma condição diferente. Porém, normalmente quando processamos imagens, alguns poucos parâmetros, como contornos que formam apenas parte dos pontos da imagem, contraste e coloração são suficientes para que a rede consiga solucionar o problema. Ou seja, podemos reduzir centenas de pontos para uma dezena de características da imagem (HAYKIN, 2009). Para conseguirmos extrair estas informações de um conjunto de dados ou de imagens utilizamos as redes convolucionais ou CNN. Estas redes, diferentes das MLPs possuem iterações esparsas entre a entrada e saída (GOODFELLOW; BENGIO; COURVILLE, 2016). Isso porque, através do operador de convolução, analisamos uma faixa de dados, no caso das imagens uma janela de tamanho $n \times m$, também denotada *kernel*, o operador de convolução analisa a composição de dados dentro deste *kernel* e traduz isso para um conjunto de características e repete este processo de forma iterativa, até traduzir um conjunto de dados ou imagens, em um conjunto de atributos, conforme apresentado na Figura 4.3. Neste exemplo é possível chegar ao mesmo resultado com o operador de convolução ou utilizando multiplicação de matriz, porém neste exemplo a convolução é cerca de 60.000 vezes mais eficiente computacionalmente que o método linear (GOODFELLOW; BENGIO; COURVILLE, 2016).

A convolução discreta em um espaço bidimensional pode ser descrita como

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n)K(i - m, j - n)$$

onde, I é nossa entrada, ou seja, a imagem, K é definido como *kernel*, uma janela de dados definida durante o processo de aprendizado e utilizada como filtro para extração dos parâmetros, S é definido como mapa de características, a saída da convolução e

Figura 4.3 – Detecção de bordas, utilizando um *kernel* de 2x2.

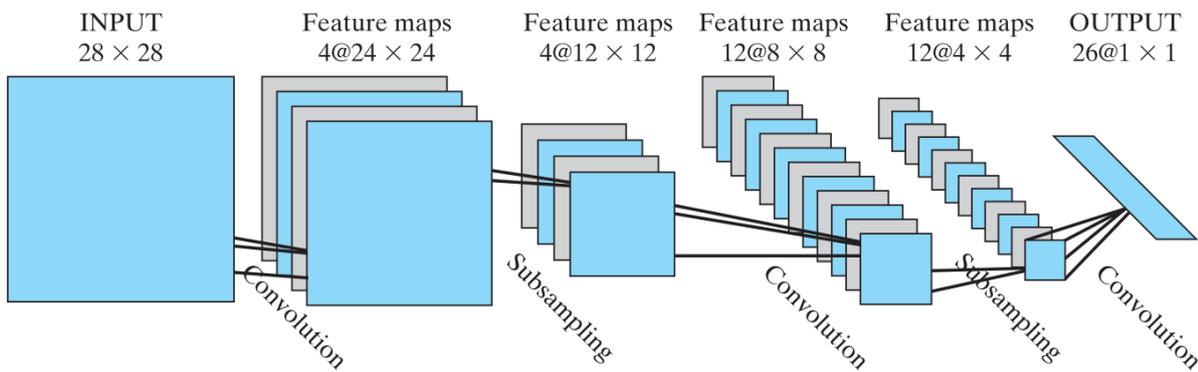
Fonte: Haykin (2009)

m e n as dimensões da imagem (GOODFELLOW; BENGIO; COURVILLE, 2016).

Segundo Goodfellow, Bengio e Courville (2016), uma camada convolucional típica e composta de 3 estágios. O primeiro estágio aplica várias convoluções em paralelo para produzir uma série de características. No segundo estágio é utilizada uma função de ativação não linear para mapear as características, este estágio também é chamado de detector de características. No último estágio são utilizadas funções de sumarização também chamadas de *pooling* para modificar a saída e sumarizar os dados reduzindo o seu volume. A Figura 4.4 apresenta um modelo de rede convolucional aplicado para o processamento de imagem, note que a imagem de tamanho 28x28, ou seja 784 atributos, passa por 3 convoluções, cada uma com os 3 estágios. A primeira camada convolucional consiste em 4 filtros com *kernel* de 5x5, ao fazer o primeiro conjunto de convoluções é aumentada a massa de dados de 28x28 para 4x24x24, ou seja, de 784 para 2304 atributos, porém a cada convolução é aplicada uma camada de *pooling* para reduzir a quantidade de dados sem perder informação, resultando em uma saída de 4x12x12, ou seja 576 atributos uma redução de aproximadamente 26%. As camadas de convolução seguintes seguem o mesmo processo resultando em mapas de 12 e 26 características e respectivamente, reduzindo os dados após o *pooling* de 576 para 192 e 26 respectivamente. Ao final é entregue como saída um vetor com 26 características que podem ser utilizadas para descrever a imagem (HAYKIN, 2009).

O *pooling* é uma função que busca sumarizar um conjunto de dados em sua área de ação, resultando em um conjunto de dados menor, um exemplo é a função MaxPooling, que retorna o maior valor em uma janela de dados (GOODFELLOW; BENGIO; COURVILLE, 2016). Um exemplo do MaxPooling pode ser visto na Figura 4.5,

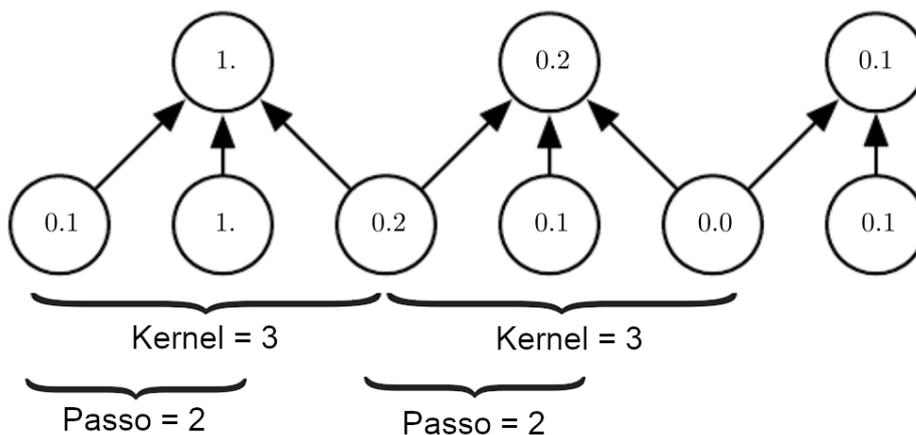
Figura 4.4 – Exemplo de rede convolucional para processamento de imagem.



Fonte: Haykin (2009)

note que um conjunto de 6 neurônios foi reduzido a 3 utilizando a sumarização, neste caso foi utilizado um *kernel* de 3 e um passo de 2.

Figura 4.5 – Aplicação do MaxPooling.



Fonte: Adaptado de Goodfellow, Bengio e Courville (2016)

4.3 TRABALHOS RELACIONADOS

Em um estudo realizado por Pelisson et al. (2020) foram confrontadas as técnicas de MLP, SVR, kNN, *Decision Trees* (DT) e *Linear Regression* (LR) utilizando apenas dados meteorológicos para a previsão de geração de energia fotovoltaica (PV) do dia seguinte. O estudo foi desenvolvido com dados da estação meteorológica e usina Fotovoltaica UFSC, com granularidade de 30 minutos. O autor utilizou o RMSE e oR^2 como métrica de erro e concluiu que os melhores resultados foram obtidos com MLP e SVR. Observou que mesmo com uma correlação mais forte entre a potência gerada

e o GHI, a adição de outras variáveis ao modelo reduziu o RMSE em 30%, levando a sugerir a continuidade do estudo com avaliação de outros parâmetros. Para MLP o autor encontrou que a combinação de 9 neurônios na camada oculta com função de ativação tangente hiperbólica obteve o melhor resultado. E, o melhor desempenho foi obtido na combinação dos parâmetros *Global Horizon Irradiance* (GHI), *Direct Normal Irradiance* (DNI), *Direct Horizontal Irradiance* (DHI), Ultra Violeta (UV), Humidade Relativa (RH), Temperatura Ambiente (AT) e Velocidade do Vento (WS).

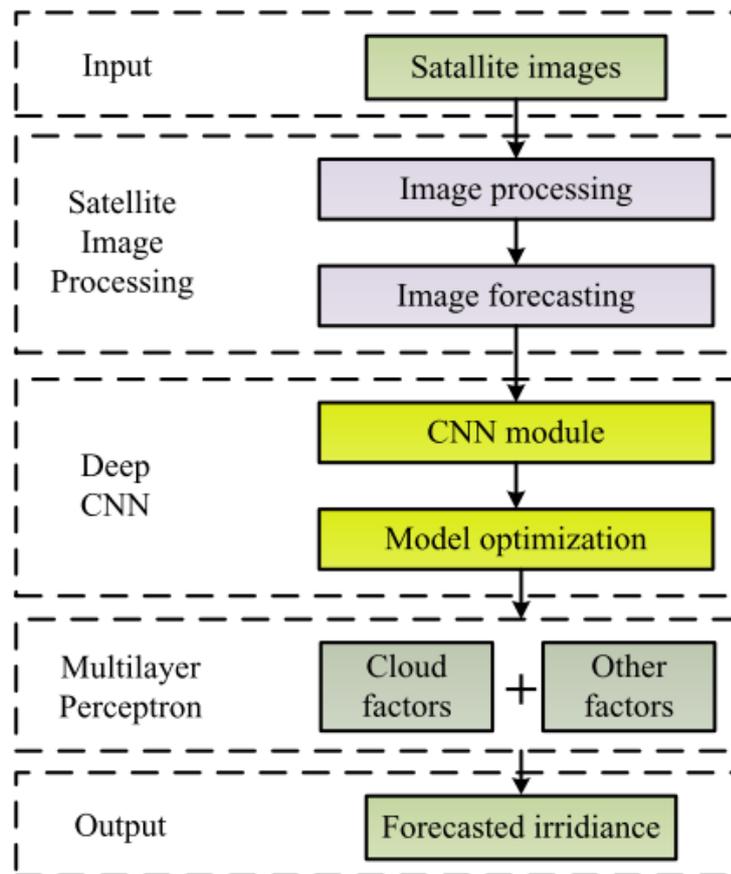
Na pesquisa de Si et al. (2021) foi proposta a determinação do GHI com uma rede convolucional, utilizando imagens de satélite no espectro visível para analisar o comportamento das nuvens, as imagens são obtidas em três intervalos de tempo, no instante t , $t-1$ e $t-2$. Conforme é apresentado na Figura 4.6, as imagens passam por um pré-processamento com o objetivo de normalizar os dados e corrigir a variação do ângulo zenital. Após o pré processamento, as imagens são utilizadas como entrada de uma CNN que vai extrair características para relacionar ao problema. Os dados de saída das duas camadas de convolução são combinados com dados de estações meteorológicas para a entrada de uma MLP, esta responsável pela predição do GHI. Os resultados obtidos pelo autor mostram a importância do pré processamento das imagens, tendo uma variação no RMSE de $28,3 \text{ W/m}^2$ sem o pré processamento para $26,2 \text{ W/m}^2$ com pré processamento em predições de uma hora na primavera. Para predições de duas horas este valor vai de $41,3 \text{ W/m}^2$ para $38,2 \text{ W/m}^2$ nas mesmas condições. Ainda segundo a autora, o agravante para o modelo são as imagens com oclusão de nuvem, ou seja, dias nublados.

Em Marquez, Pedro e Coimbra (2013), é proposta uma abordagem utilizando imagens de satélite do espectro visível e IR. O estudo utiliza das imagens de satélite do espectro visível para a geração de um índice de cobertura de nuvens aplicando uma correção do ângulo zenital e albedo¹. Já as imagens do espectro IR são utilizadas para a geração de um campo vetorial que determina o vetor direção e velocidade das nuvens utilizando o algoritmo *Particle Image Velocimetry* (PIV). Os dados de velocidade e cobertura de nuvens são combinados como entrada de uma ANN para a predição de GHI em horizontes de 30, 60, 90 e 120 minutos com granularidade de 30min. Como resultado foi alcançado um RMSE de $57,84 \text{ W/m}^2$ para horizonte de 30min e $68,11 \text{ W/m}^2$ para horizonte de 30min.

Outra proposta desenvolvida por Eissa et al. (2013), busca uma abordagem utilizando redes em paralelo para diferentes condições climáticas com objetivo de relacionar parâmetros meteorológicos e imagens de satélite para determinar o DNI, DHI e GHI no instante. A autora divide o problema em DNI e DHI, processando de forma separada cada uma das variáveis. As duas redes aplicam a mesma estratégia, utilizar

¹ Refletividade difusa ou poder de reflexão de uma superfície. No caso da superfície terrestres, este valor sofre a influência da composição da atmosfera.

Figura 4.6 – Arquitetura de previsão utilizando imagens de satélite, CNN e MLP proposta por Si et al. (2021).



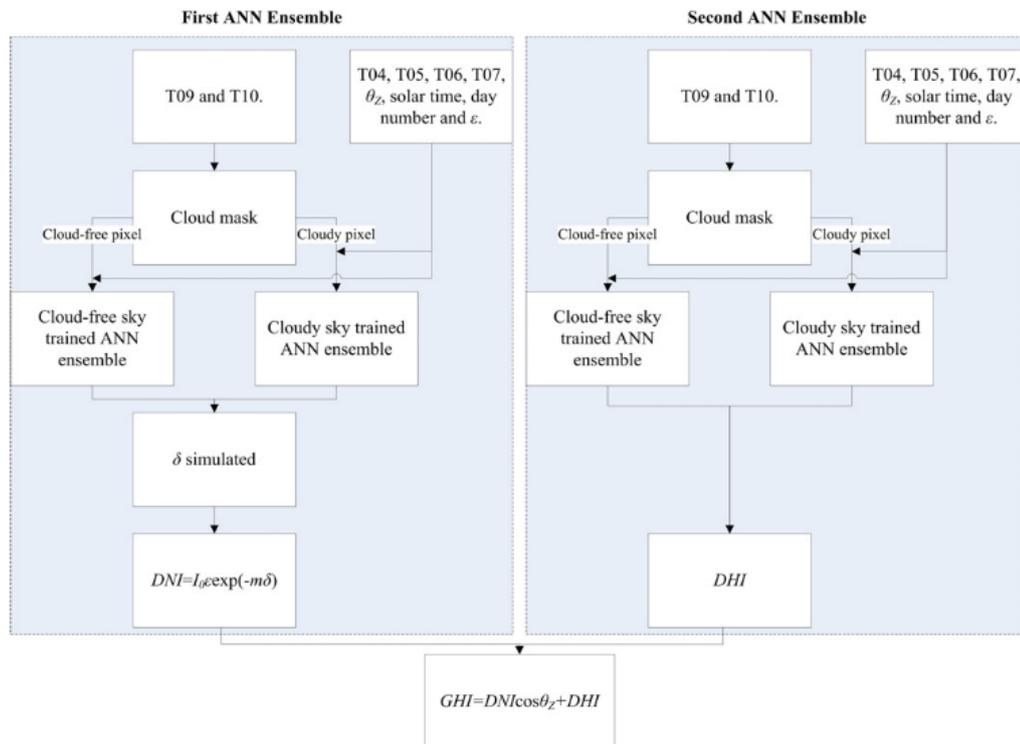
Fonte: Si et al. (2021)

os canais espectrais $10,8\mu m$ e $12\mu m$ para gerar uma máscara de nuvem que divide o problema em sem nuvem e com nuvem. De acordo com o resultado, o algoritmo seleciona qual a rede utilizar e adiciona outros canais espectrais combinados com as variáveis o ângulo zenital, a hora solar, o número do dia e a correção de excentricidade como entrada da ANN. Quatro modelos são treinados, DNI com nuvem e sem nuvem, DHI com nuvem e sem nuvem. Os dados combinados analiticamente determinam o GHI. A topologia pode ser vista na Figura 4.7. Os resultados utilizando dados com 15min de granularidade alcançaram rRMSE para o DNI, DHI e GHI de 26,1%, 25,6% e 12,4% respectivamente combinando os modelos de nuvens e sem nuvens.

4.4 CONSIDERAÇÕES FINAIS

Uma das características das MLPs e das CNNs é encontrar relação entre entrada e saída em problemas complexos e não lineares. Conforme vimos, a metodologia de aprendizado de máquina tem sido utilizada em conjunto de imagens de satélite

Figura 4.7 – Arquitetura de previsão utilizando imagens de satélite, proposta por Eissa et al. (2013).



Fonte: Eissa et al. (2013)

para a previsão de geração de energia PV. É importante salientar que é fundamental para este tipo de desenvolvimento uma boa base de dados, muitas vezes precisando de um pré processamento para a utilização nas ANNs, conforme visto nos trabalhos anteriores.

5 ESTUDO DE CASO

Este capítulo traz um estudo dos dados obtidos da estação no Laboratório Fotovoltaica¹ da UFSC, que possui em sua estrutura painéis fotovoltaicos para produção de energia e estação solarimétrica, e dos dados do satélite GOES-16 nos anos de 2018 e 2019. Na Seção 5.1 serão abordados os dados da estação terrestre e a sua correlação com a geração de energia fotovoltaica (PV). Na Seção 5.2 serão tratados os dados do satélite fazendo uma análise espacial para definir a área de interesse e avaliar o comportamento dos dados ao longo do tempo. A Seção 5.3 consolida as conclusões das seções anteriores e delimita o horizonte e a granularidade que será utilizada na base de dados. A Seção 5.4 correlaciona os dados da estação terrestre com os dados de satélite para identificar os parâmetros que serão utilizados na modelagem da rede neural, fazendo uma análise exploratória nas séries e confrontando as variações temporais. Por fim a Seção 5.5 traz a conclusão sobre a análise.

5.1 ESTAÇÃO TERRESTRE

A estação terrestre possui grande importância para este trabalho, ela é responsável pelo fornecimento dos dados meteorológicos históricos que serão combinados ao modelo neural para a predição de geração de energia PV. Estes dados são utilizados em dois momentos distintos, como saída e entrada da rede. No primeiro e mais importante, utilizaremos os dados da estação para o treinamento do modelo. No segundo caso, combinamos os dados as imagens de satélite para melhorar o desempenho da inferência, os detalhes serão discutidos a frente.

A base de dados da estação terrestre foi obtida do laboratório Fotovoltaica UFSC e compreende dados da estação meteorológica e dados de geração de energia pela usina PV dos anos 2018 e 2019. Os dados foram obtidos com frequência de 1 min. Alguns dos dados disponíveis e de maior relevância para este trabalho são apresentados na Tabela 5.1. Os dados brutos são obtidos com um datalogger Campbell CR6 (PELISSON et al., 2020).

Para a utilização da base de dados foi preciso fazer um trabalho de pré-processamento. Primeiramente a base estava dividida em dois grupos, dados de 2018 e dados de 2019, antes de unificar as bases foi feito um trabalho de remoção de dados duplicados e então foram combinados. Os dados obtidos da estação foram salvos em GMT-0, o mesmo fuso horário foi utilizado para o satélite, não sendo necessário fazer correção de fuso horário. Porém alguns dos dados precisaram ser tratados, como velocidade do vento, DNI, GHI e UV negativos, GHI maior que $1200W/m^2$ e temperatura do ar menor que $8,4^{\circ}C$, com base na distribuição histórica, sendo estes dados removidos das leituras. Da mesma forma os dados de GHI menores que 0 foram substituídos

¹ <https://fotovoltaica.ufsc.br/>

Tabela 5.1 – Dados do laboratório e usina fotovoltaica UFSC.

Parâmetro	Sigla	Sensor	Unidade	Mínimo	Máximo
<i>Direct Normal Irradiance</i>	DNI	SHP1	W/m ²	0	1.041
<i>Global Horizontal Irradiance</i>	GHI	SMP22	W/m ²	0	1.627
<i>Diffuse Horizontal Irradiance</i>	DHI	SMP11	W/m ²	0	1.230
Índice Ultravioleta	UV	CUV 5	W/m ²	0	79,63
Temperatura Ambiente	AT	HMP155	°C	8,40	37,11
Velocidade do Vento	WS	WindSonic1	m/s	0	10,28
Umidade Relativa	RH	HMP155	%	31,03	100
Potência de Saída	PO	ABB Uno 2.0 Inverter	W	0,02	2.739

Fonte: Adaptado de Pelisson et al. (2020).

por 0. Como resultado a base conta com 871.709 pontos de leitura entre os anos de 2018 e 2019, com granularidade de 1min e está disponível para consulta².

Para o início dos estudos foi feito o cálculo da correlação de Pearson entre a potência do gerador e as variáveis GHI, DNI e DHI, segundo a Equação (5.1).

$$e(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \times \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (5.1)$$

Onde x e y são os valores medidos e, \bar{x} e \bar{y} são as médias aritméticas dos conjuntos de dados. O resultado apresentado na Tabela 5.2 como esperado mostra que o GHI tem uma correlação muito forte com a potência gerada, isso se concretiza com os dados levantados na literatura e nos permite construir um modelo de predição que tenha como saída o GHI. Isto dá uma vantagem para o modelo, que pode ser aplicado para diferentes unidades geradoras no mesmo parque através de uma relação entre o GHI e o potencial de geração.

Tabela 5.2 – Correlação da potência do gerador com as variáveis de irradiação.

Componente de Irradiação	Correlação com a Potência
GHI	0,97
DNI	0,77
DHI	0,36

Fazendo uma análise de distribuição média do GHI ao longo do dia, obtemos o gráfico da Figura 5.1. Este resultado permite delimitar nosso horário de interesse para a predição. Se analisarmos a curva podemos observar que a estação começa a ser irradiada as 6h e permanece ativa até as 17h, desta forma, para ter dados do início do dia para as primeiras inferências, foi definido para o estudo que as predições começam a ser feitas as 9h e a última inferência faz uma predição para as 17:30h.

Determinada a saída do modelo para o GHI e o intervalo de predição, é possível avaliar a base de dados para as restrições impostas. Com a limitação horária

² <https://doi.org/10.5281/zenodo.7954973>

Figura 5.1 – Distribuição média do GHI ao longo de 24h nos anos de 2018 e 2019.



reduzimos a base de dados para 218.549 leituras com granularidade de 1min. As correlações das variáveis da estação com o GHI podem ser vistas na Tabela 5.3. Avaliando os parâmetros podemos perceber que a irradiação UV possui uma correlação forte, algo esperado visto que o UV também é um componente de irradiação que sofre atenuações por conta de nuvens e o albedo assim como GHI. A umidade relativa e a temperatura possuem correlação média e serão avaliadas nos modelos, pôr fim a velocidade do vento possui uma correlação fraca.

Tabela 5.3 – Correlação do GHI com as variáveis da estação para os anos de 2018 e 2019 das 8h às 18h.

Variáveis da estação	Correlação com GHI
Irradiação ultravioleta (UV)	0.99
Velocidade do vento (WS)	0.29
Umidade relativa (RH)	-0.49
Temperatura do ar (AT)	0.48

Para finalizar a análise, o gráfico da Figura 5.2 apresenta a distribuição do GHI médio ao longo do ano. Conforme esperado o potencial de irradiação é acentuado durante os meses de novembro a fevereiro e com menor intensidade nos meses de maio a julho, este fator foi abordado nas seções anteriores e é ocasionado pelas estações do ano. Ao distribuir a variável GHI média em um gráfico Mês x Hora, obtemos a Figura 5.3, note que a irradiação é reduzida no meio do ano, como também a janela de maior potencial e corrobora com nossa escolha de intervalo de predição.

Figura 5.2 – Distribuição média do GHI ao longo dos meses nos anos de 2018 e 2019.

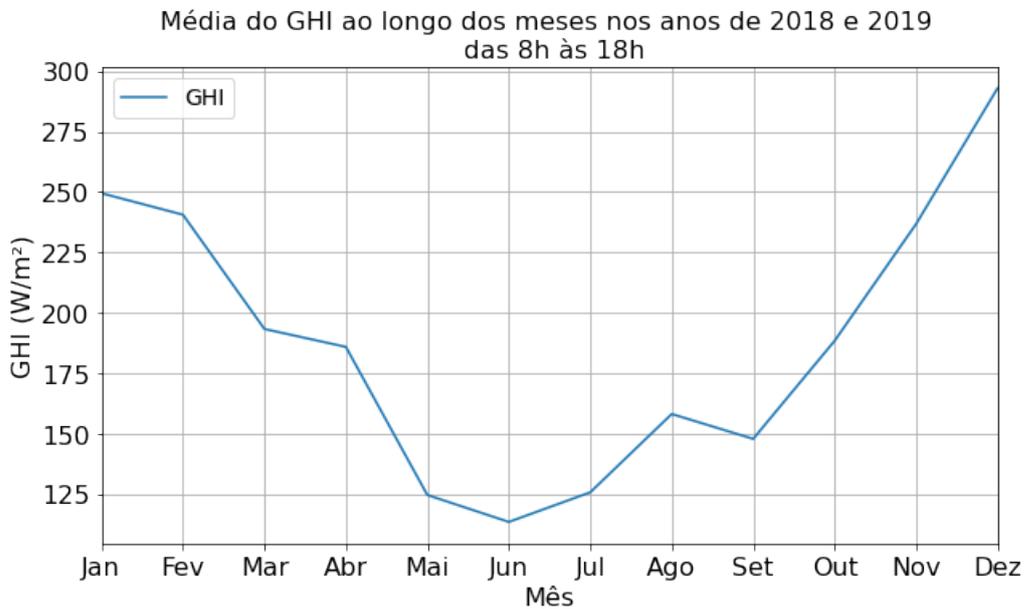
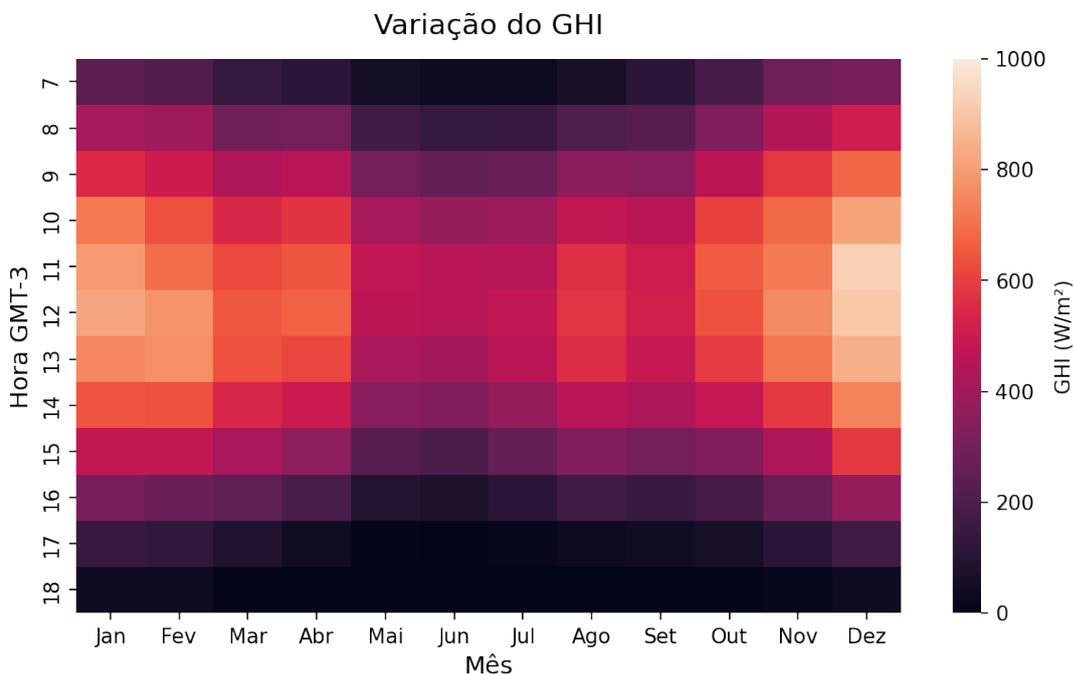


Figura 5.3 – Distribuição média do GHI Hora x Mês para os anos de 2018 e 2019.



5.2 SATÉLITE

Conforme mencionado nos capítulos anteriores, este trabalho se aproveita das melhorias implementadas nos satélites da série GOES para o desenvolvimento de um modelo neural que relaciona as imagens do satélite com a geração de energia PV. Como a estação considerada para o desenvolvimento deste trabalho está localizada

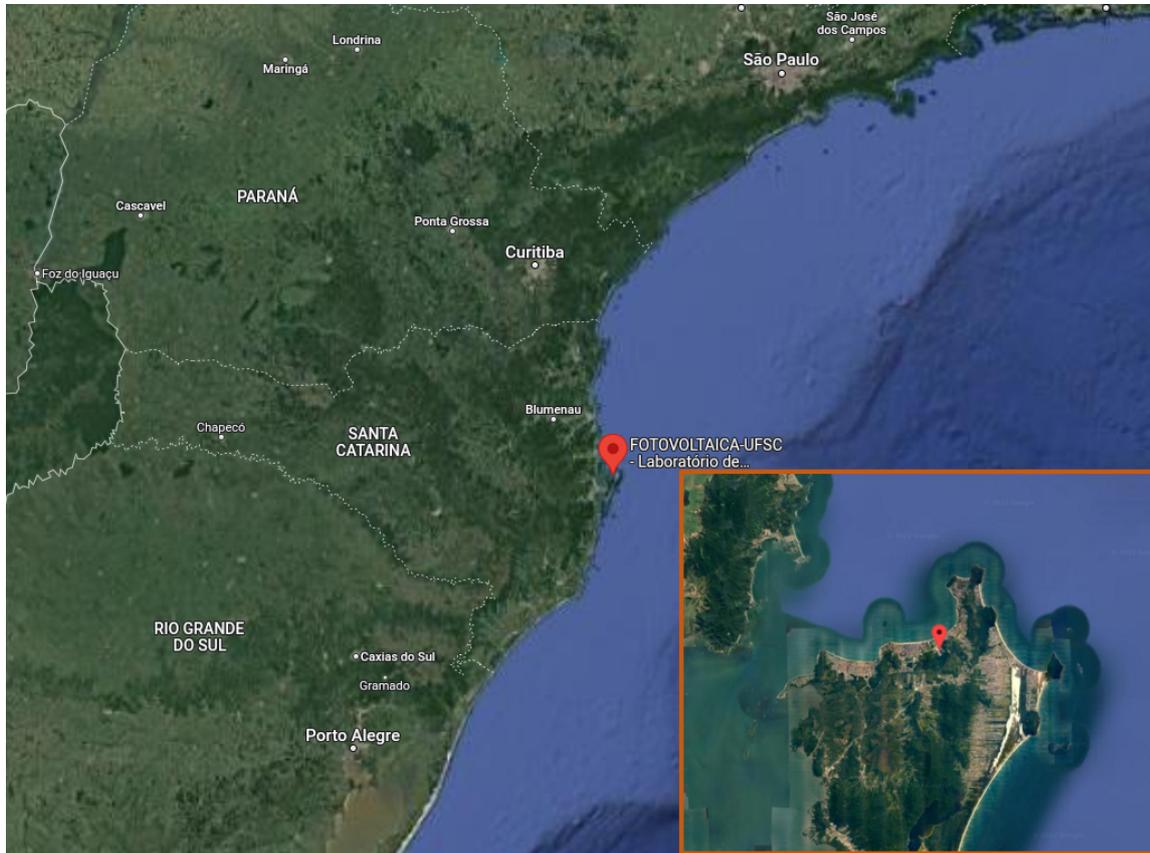
em Florianópolis, o satélite da série GOES que está em órbita de capturar as imagens da estação é o GOES-16. Os dados do satélite foram obtidos no mesmo intervalo que os dados da estação terrestre, 2018 e 2019, sendo estas imagens *full disk*, toda a área de captura do GOES-16.

Recapitulando, o satélite GOES-16 possui 16 faixas espectrais, sendo 6 no intervalo visível e 12 não visível. Como nosso objetivo é relacionar o GHI com a cobertura de nuvens diurnas, o Canal 2 do imageador ABI é uma boa opção, pois além de ser recomendado para esta aplicação é o canal espectral com maior resolução espacial, 0,5 km/pixel, permitindo uma avaliação mais detalhada do nosso ponto de interesse. As imagens deste canal fornecem os dados de refletância, sendo que a cobertura de nuvem apresenta maior refletância que o solo, facilitando a identificação de oclusão (SI et al., 2021). Como as nuvens possuem grande influência sobre a incidência de irradiação, o contraste entre nuvem e solo pode beneficiar a rede convolucional, tornando este canal favorável para a aplicação. Um dos pontos de atenção para os dados obtidos está na mudança de operação do satélite após o lançamento, entre os anos de 2018 e 2019 houve uma mudança no modo operacional do satélite GOES-16 do modo M3 para o modo M6, modificando a frequência de captura dos dados. No modo M3 o satélite executa uma rotina que permite capturar uma imagem a cada 15min, o modo M6 possui uma alteração em sua rotina o que permite capturar uma imagem a cada 10min, desta forma a menor frequência de captura que podemos trabalhar com as duas bases de dados é de 30min sendo esta a granularidade utilizada para este desenvolvimento.

Cada imagem do canal 2 do imageador ABI bruta é de aproximadamente 500MB, para reduzir o volume de dados e delimitar a análise, foi definido o intervalo das 8h até as 18h GMT-3 conforme a seção anterior. Para facilitar o processamento das imagens a base de dados foi obtida em uma janela de 10° de latitude e 10° de longitude, sobre o ponto de interesse nas coordenadas -27,4310° latitude e -48,4414° longitude. O resultado foram imagens que compreendem grande parte do sul do Brasil, conforme Figura 5.4.

Mesmo com o recorte, as imagens resultantes são 2124x875 pixels, ou seja, 1,8 milhões de atributos para a rede, onde cada ponto representa a refletância. Com base no estudo de Chow et al. (2011) foram encontradas velocidades de nuvem de 7,9 m/s, ou seja, 28 km/h. Desta forma a fim de reduzir o tamanho da imagem, foi considerado que as nuvens com maior influência têm uma velocidade não maior que 40 km/h. Considerando que a resolução espacial do satélite é de 0,5 km/pixel e a granularidade dos dados que estamos trabalhando é de 30min, chegamos a uma conta aproximada para o tamanho da imagem mínima, como sendo

Figura 5.4 – Recorte de 10° de latitude e 10° de longitude do ponto de interesse, com zoom na estação Fotovoltaica, sendo esta a área utilizada como entrada da rede.



Fonte: Adaptado de Google (2022).

$$\left(\begin{array}{c} \text{Resolução} \\ \text{da} \\ \text{imagem} \end{array} \right) = \frac{\left(\begin{array}{c} \text{Velocidade} \\ \text{da nuvem} \end{array} \right) \times \left(\begin{array}{c} \text{Granularidade} \\ \text{dos dados} \end{array} \right)}{\left(\begin{array}{c} \text{Resolução} \\ \text{espacial} \end{array} \right)} = \frac{40 \left(\frac{km}{h} \right) \times 0,5 (h)}{0,5 \left(\frac{km}{pixel} \right)} = 40 (pixels)$$

com este resultado é possível reduzir o tamanho da imagem, representada em zoom na Figura 5.4 e resulta em 1600 pontos. Os dados de satélite foram recortados em imagens de 40x40 pixels, ou seja, 20x20 km resultando na Figura 5.5(a). Ao final a base de dados ficou com 7.136 imagens para o ano de 2018 e 7.193 imagens para o ano de 2019.

As imagens da Figura 5.5 representam 6 capturas em 3 intervalos, 8h , 12h e 17h, selecionadas as imagens com maior e menor GHI no intervalo no ano de 2018 e 2019. Note que quanto menor o GHI maior é a refletância, isso é evidente visto que quanto mais claro é o *pixel*, maior é a refletância. A imagens claras apresentam cobertura de nuvem, com exceção dos extremos, como é o caso da Figura 5.5(g), uma

Figura 5.5 – A imagem (a) é uma representação do local obtida do Google (2022). As imagens (b) a (g) foram obtidas do satélite GOES-16 canal 2 em diferentes intervalos de tempo no mesmo local, para representar a refletância em diferentes momentos do dia com os valores mínimos e máximos de GHI.

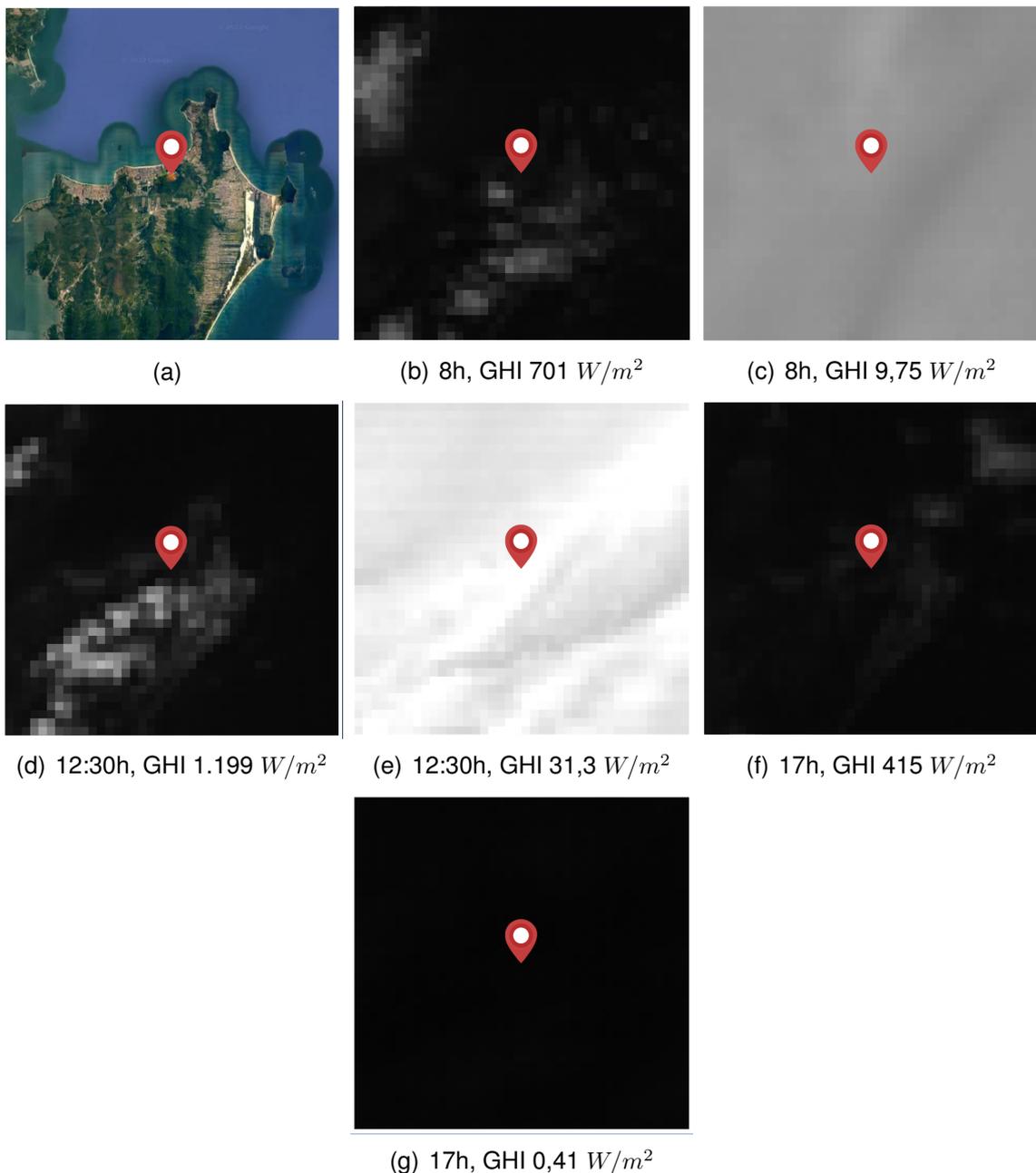
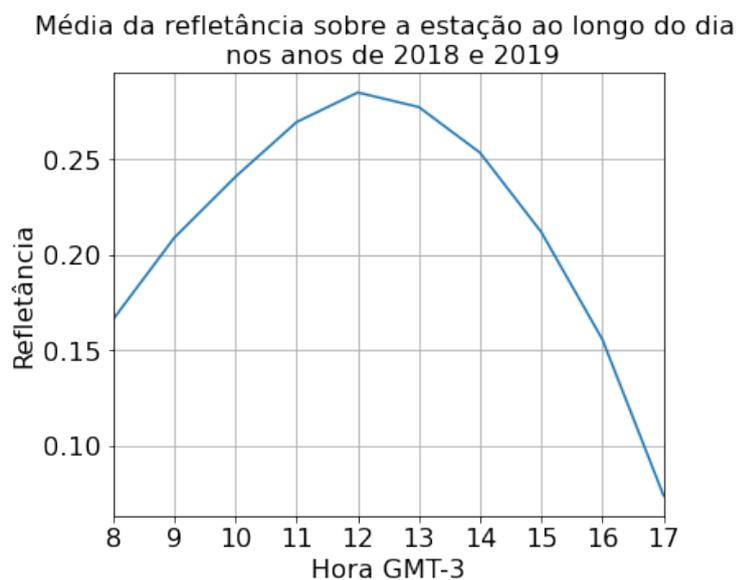


imagem capturada no dia 12/06/2018 e nesta época do ano temos uma redução de exposição ao sol o que ocasionou uma imagem escura. A Figura 5.5(d), está livre de nuvens, podemos comparar os contornos com a Figura 5.5(a) e notar que a refletância do solo é muito menor que a das nuvens neste canal.

Fazendo uma análise temporal da média da refletância ao longo do dia no grá-

fico da Figura 5.6, podemos notar um comportamento similar a irradiação GHI, onde a refletância possui um pico próximo as 12h, este fenômeno pode ser atribuído a posição do sol, gerando maior refletância no solo e no topo da nuvem. Quando comparada a distribuição ao longo do mês Figura 5.7, é possível notar um comportamento diferente, principalmente no mês de dezembro, porém não podemos relacionar a refletância diretamente ao GHI, desta forma vamos avaliar os dados mais a fundo na seção seguinte, comparando a relação da estação terrestre com os dados de satélite.

Figura 5.6 – Distribuição média da refletância ao longo de 24h nos anos de 2018 e 2019.



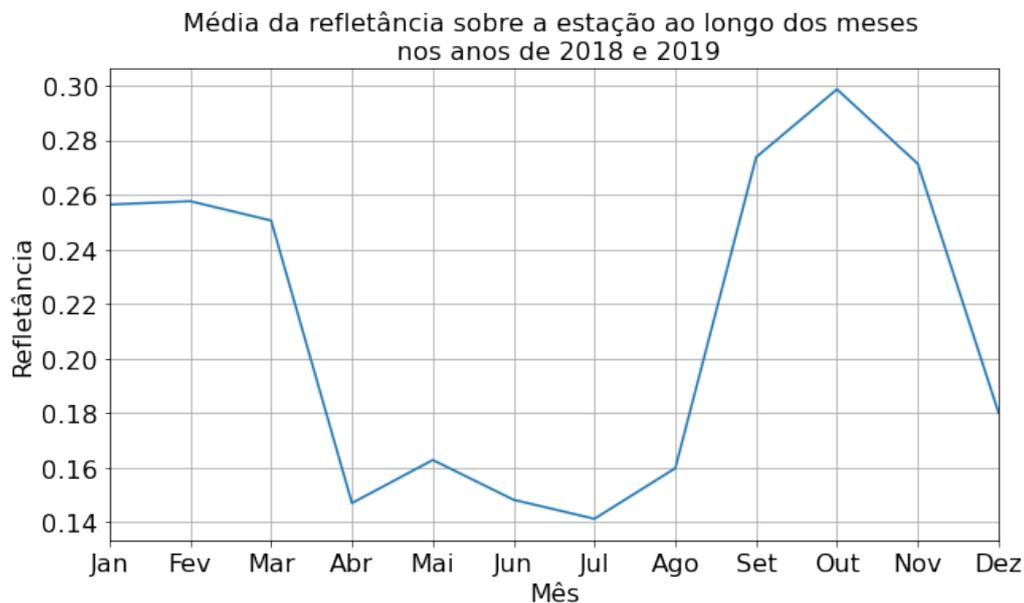
Ao distribuir a refletância média em um gráfico Mês x Hora, obtemos a Figura 5.8, note que a distribuição da refletância é menos uniforme que a do GHI, isto porque a refletância depende da irradiação do sol e do movimento das nuvens, é possível observar uma janela maior entre os meses de abril a agosto com menor refletância e uma descontinuidade no mês de dezembro. Note que os pontos de maior refletância estão em fevereiro, março, setembro e outubro.

5.3 HORIZONTE E GRANULARIDADE

Conforme já abordado na seção anterior, devido a limitação dos modos de operação do GOES-16 vamos trabalhar com a granularidade mínima possível de 30min, ou seja, vamos sumarizar nossa base de dados em intervalos de 30min. Para isso foram descartadas as leituras intermediárias, sendo elas: 15min e 45min para o Modo 3 e 10min, 20min, 40min e 50min para o Modo 6.

O horizonte de predição mínimo é limitado pela granularidade em 30min, nesta condição vamos estimar o GHI 30min a frente. Para ampliar o estudo, entender a

Figura 5.7 – Distribuição média da refletância ao longo dos meses nos anos de 2018 e 2019.



capacidade do modelo proposto e comparar com resultados encontrados na literatura, também serão feitos testes em horizontes de 60, 120 e 180 minutos.

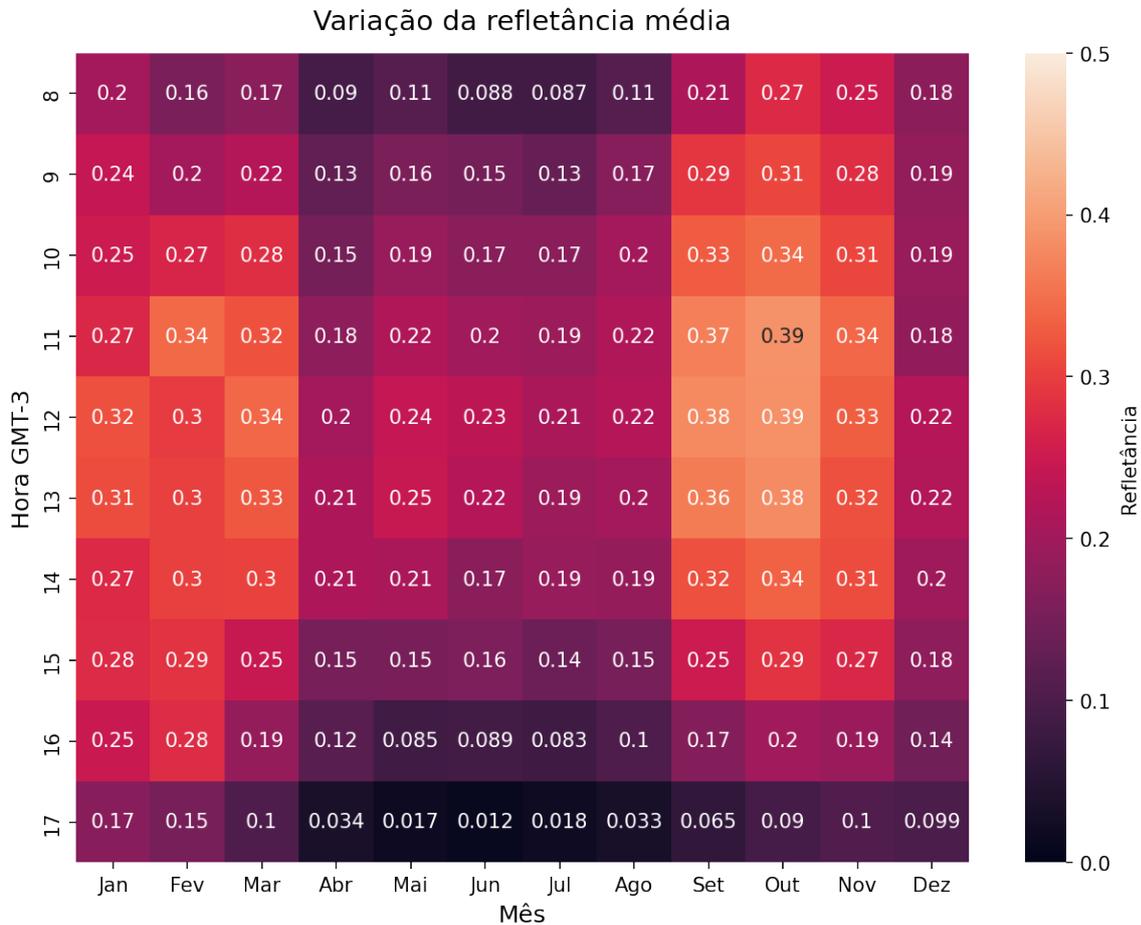
5.4 ANÁLISE EXPLORATÓRIA

Consolidadas as bases de dados foram conduzidos estudos para identificar a correlação dos dados de satélite com os dados de GHI da estação terrestre. Todos os dados apresentados estão em GMT-3 no intervalo das 8h às 17:30h nos anos de 2018 e 2019 para facilitar a interpretação e todas as correlações foram calculadas com uma média da refletância de 4 pontos da imagem, ou seja, uma área de $1km^2$ centrada no ponto de interesse.

Ao correlacionar o GHI com a refletância foi alcançado uma correlação de -0,34, ou seja, uma correlação fraca. Os dados foram dispostos em um gráfico de dispersão apresentado na Figura 5.9(a), podemos observar uma zona de concentração na base, mesmo assim uma grande dispersão dos dados.

Para entender a relação de cada ponto de refletância do satélite com o GHI, foi feita a correlação individual de cada um dos pontos da imagem na área de interesse, Figura 5.9(b), o qual demonstra a distribuição da correlação dos dados por ponto da imagem. É possível notar que a correlação é de maior intensidade no centro da imagem, na estação terrestre, porém os contornos da ilha e a camada de mar mais próxima da costa possui uma correlação de mesmo valor, supomos que devido a análise estar sendo feita em uma ilha, a reflexão da água possa ter influência sobre a refletância.

Figura 5.8 – Distribuição média da refletância Hora x Mês para os anos de 2018 e 2019.

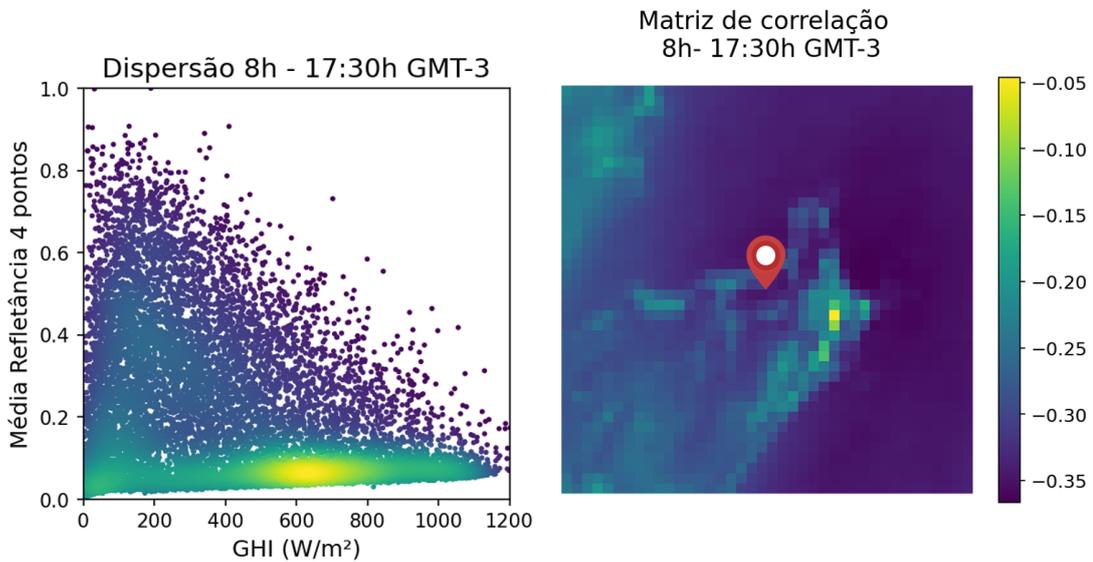


Si et al. (2021) enfatiza a influência do ângulo zenital ao estimar o GHI utilizando imagens de satélite, desta forma, para investigar a influência da posição do sol sobre o resultado obtido, foi calculado a correlação da média de 4 pontos com os dados de GHI em intervalos de 30min, resultando no gráfico apresentado na Figura 5.10. Fica evidente que a posição do sol em relação a usina tem forte influência sobre os dados de geração e que a relação é mais forte próximo das 12h, onde o efeito de sombra não é influenciado pelo ângulo zenital, quando o sol se encontra posicionado sobre a usina.

Para reforçar a correlação da geração com as imagens do satélite, foi calculada a correlação para cada ponto da imagem de satélite com o GHI no intervalo das 8h às 17:30h, conforme mostra a Figura 5.11. É possível notar que os pontos que estão posicionados acima da usina são os pontos de maior influência e podemos perceber a influência do horário na correlação do GHI com os pontos próximos da costa, note que esta correlação diminuiu ao analisar cada imagem individualmente.

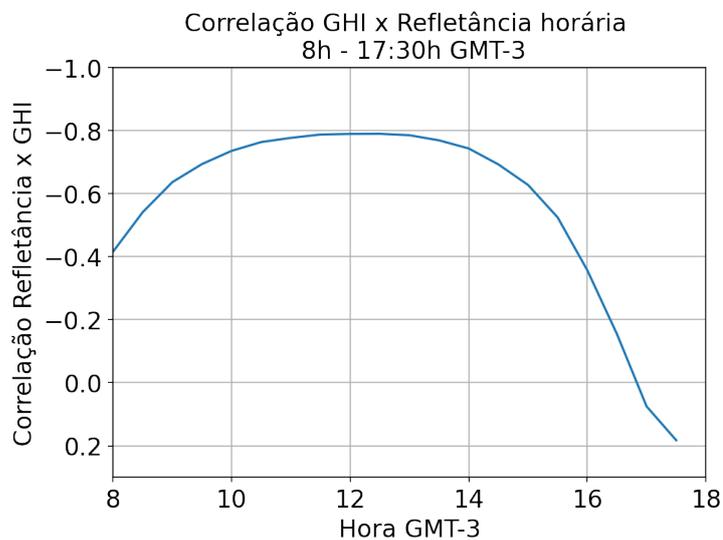
Ao analisarmos a correlação mensal Figura 5.12, fica evidente a influência dos

Figura 5.9 – Relação do GHI da estação terrestre com os dados de satélite no intervalo das 8h às 17:30h GMT-3.



(a) Dispersão da refletância com a irradiação GHI . (b) Mapa de correlação de cada ponto da imagem com a estação terrestre.

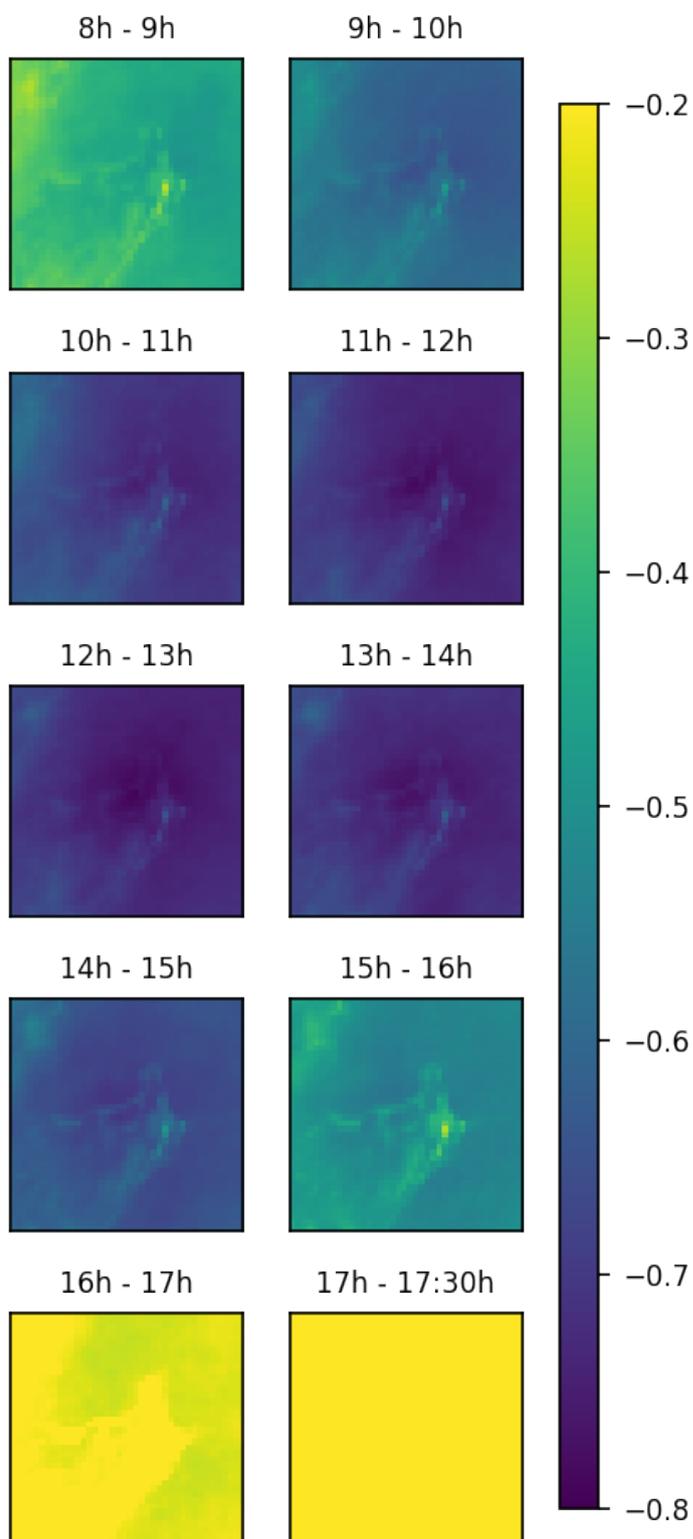
Figura 5.10 – Correlação da média de 4 pontos da imagem de satélite com a estação terrestre a cada 30min, no intervalo das 8h às 17:30h.



meses do ano na correlação dos dados. Se analisarmos a Figura 5.13, correlação entre GHI e refletância mês x hora, podemos observar que no mês de abril a agosto as 17h a correlação possui uma inversão de sinal, sinalizando que a redução da refletância está relacionada a redução da irradiação, se considerarmos que o sol se põe mais cedo na região nesta época, podemos supor que esta situação possui relação com a menor exposição do sol, mas carece de um estudo mais aprofundado.

Podemos notar no gráfico da Figura 5.13 que quando os dados são analisados

Figura 5.11 – Matriz de correlação de cada ponto da imagem de satélite com a estação terrestre a cada 1h, no intervalo das 8h às 17:30h.



mês a mês e hora a hora, a correlação tende a aumentar, podemos entender que neste momento os dados estão sendo analisados sobre uma perspectiva próxima e

isto pode beneficiar a rede na sua inferência. Este resultado sugere que os parâmetros temporais como entrada da rede podem auxiliar no modelo o tornando mais assertivo, permitindo correlacionar hora do dia com as outras entradas.

Figura 5.12 – Correlação da média de 4 pontos da imagem de satélite com a estação terrestre mensalmente, no intervalo das 8h às 17:30h.

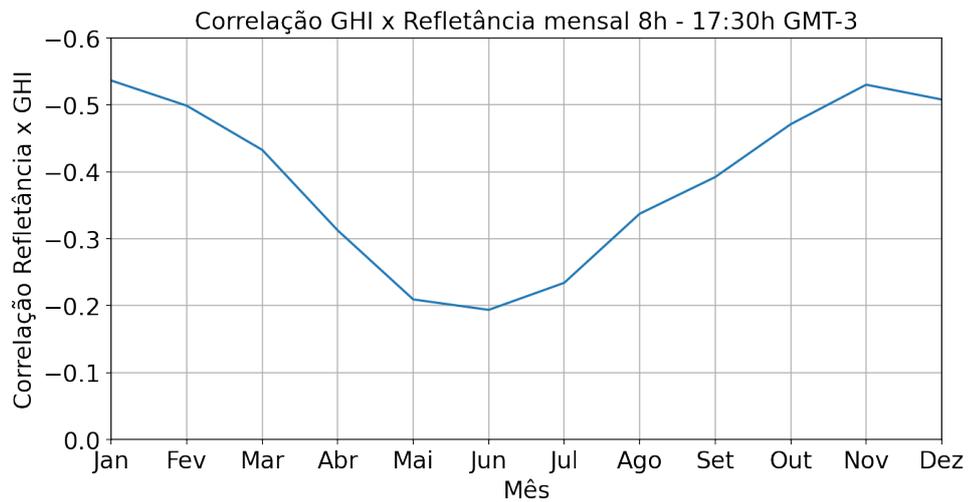
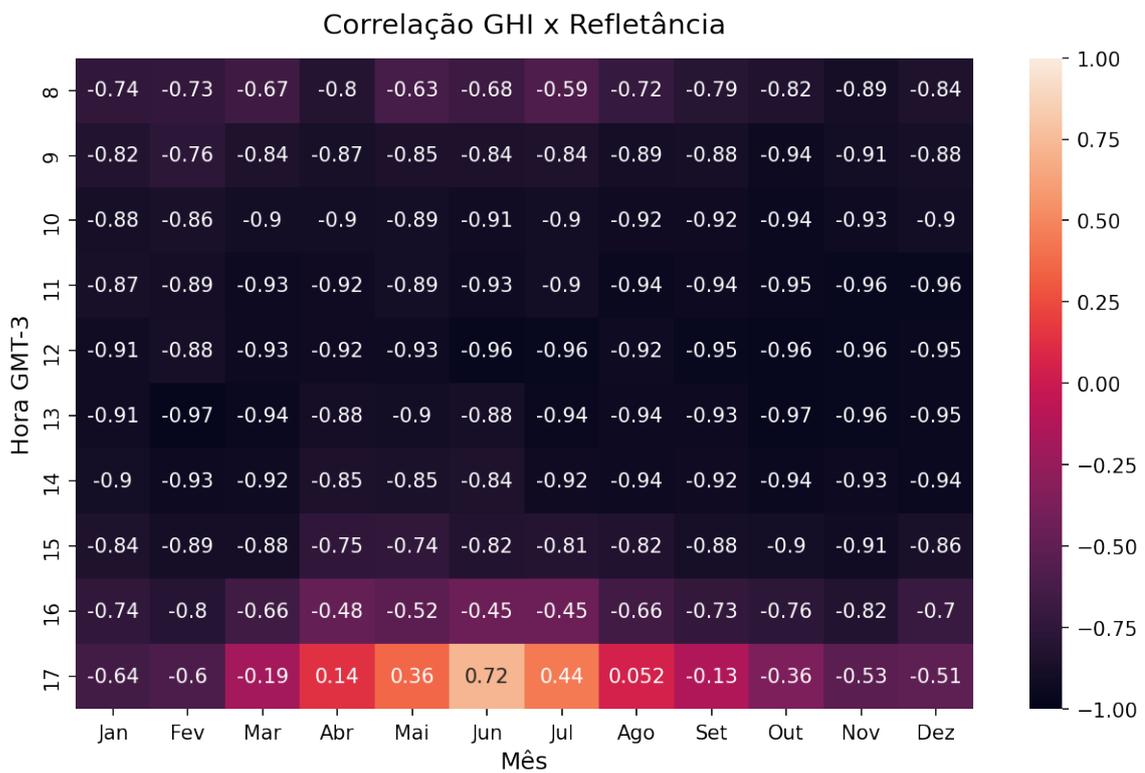


Figura 5.13 – Correlação da média de 4 pontos da imagem de satélite com a estação terrestre mês x hora.



5.5 CONSIDERAÇÕES FINAIS

Com as análises apresentadas fica claro a relação do GHI com a potência gerada na usina, permitindo o estudo seguir com a modelagem de uma rede neural considerando como parâmetro de saída o GHI. Também confirmamos que relação entre a refletância obtida das imagens de satélite com o GHI da estação terrestre são fortes, o que valida a nossa proposição inicial e permite a modelagem de uma rede para a previsão do GHI tendo como base as imagens de satélite. Também fica evidente a influência da hora do dia e dos meses do ano, como esperado estes parâmetros influenciam a capacidade de geração e a correlação dos parâmetros do satélite com a estação terrestre, porém conforme apresentado podemos trabalhar em metodologias para contornar o problema, visto que de uma perspectiva mais próxima os dados possuem forte correlação.

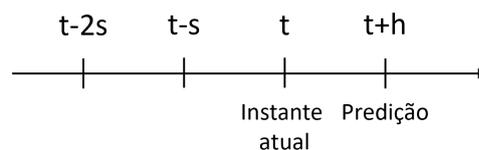
6 MÉTODOS

Este capítulo aborda todo o conjunto experimental utilizado para a construção do modelo neural proposto. A Seção 6.1 aborda todos os elementos da rede proposta, detalhando as camadas convolucionais, a camada de integração dos dados da estação terrestre e a MLP, dando detalhes da configuração experimental que será utilizada. Na Seção 6.2 são apresentados os parâmetros de treinamento da rede proposta, passando pelo otimizador, funções de validação e decaimento, assim como o framework utilizado para a programação da rede. A seção 6.3 detalha cada um dos experimentos executados, dando detalhes de cada configuração de parâmetros, a motivação e as combinações avaliadas. Ao final na Seção 6.4 são apresentadas as considerações finais.

6.1 TOPOLOGIA

Conforme apresentado por Si et al. (2021), é possível determinar o GHI com uma CNN, utilizando 3 imagens como entrada em instantes de tempo diferentes para correlacionar os dados do satélite com o GHI da estação terrestre. Vamos considerar t o instante atual, $t + h$ o horizonte de predição, $t - s$ um passo atrás e $t - 2s$ dois passos de tempo atrás, onde h é o horizonte de predição e s é a granularidade dos dados, conforme apresentado na Figura 6.1.

Figura 6.1 – Sequência temporal para cada predição.

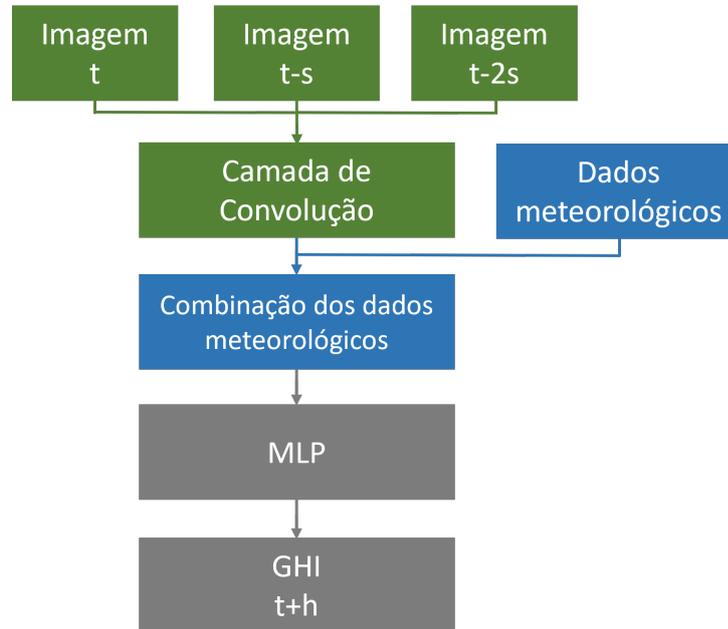


Partiremos de uma topologia base para o estudo, Figura 6.2, sendo ela composta por uma entrada para a rede convolucional, uma camada convolucional para a extração de características das imagens de satélite, uma entrada para os dados meteorológicos, uma camada de combinação dos dados para a adição de dados meteorológicos, uma camada MLP para relacionar os dados e como saída temos uma predição em $t + h$ do GHI. O número de camadas convolucionais é um dos parâmetros do estudo e será tratado a frente.

A composição da rede Figura 6.2 foi dividida em conjuntos. O verde representa o processamento de imagem, descrito na Seção 6.1.1. O azul é atribuído a entrada das variáveis meteorológicas, descrito na Seção 6.1.2. O cinza é a composição da MLP e a saída, descrito na Seção 6.1.3. O conjunto cinza está em todos os testes, já os conjuntos azul e verde podem ou não estar presentes, os grupos de testes são

detalhados a frente. A organização dos dados para o treinamento é tratado na Seção 6.1.4.

Figura 6.2 – Topologia da ANN proposta.



6.1.1 Camada convolucional

A camada convolucional é responsável por processar 3 imagens em intervalos de tempo diferentes, sendo eles t , $t - s$ e $t - 2s$ e entregar as principais características da imagem. Esta camada é necessária para o processamento de imagem pois cada imagem de 40×40 é composta de 1600 pontos, como são 3 imagens estamos falando de 4800 atributos para serem processados. Neste caso a camada convolucional vai reduzir a quantidade de atributos de saída e entregar um dado já processado em características da imagem o que espera-se, irá melhorar o desempenho da rede MLP. Conforme diagrama da Figura 6.3 cada camada é composta por uma convolução 2D, uma normalização e uma camada de *pooling* 2D. Na primeira camada do conjunto convolucional, é aplicada a convolução com ativação ReLU, para terceira camada o *average pooling* 2D, mesmo conceito do MaxPooling mas calculando a média do *kernel*, com passo 2×2 , esta topologia foi adaptada do estudo de Si et al. (2021). Ao final foram feitos os testes com e sem a presença de uma camada densa para a redução dos parâmetros de saída da convolução.

A quantidade de camadas convolucionais, a quantidade de filtros convolucionais e o *kernel*, foram determinadas com base em um estudo preliminar apresentado no Apêndice B, e selecionados os melhores parâmetros conforme apresentado abaixo:

- Camadas de convolução: {1, 2}

- Filtros: {16, 32}
- *Kernel*: {3, 5}
- Camada densa: {sem, 16}

6.1.2 Estratégia de combinação dos dados

As topologias foram restadas em 3 condições, sendo elas somente dados de satélite, Figura 6.4(a), somente dados meteorológicos, Figura 6.4(b) e a combinação dos dois, Figura 6.4(c), neste último caso foi traçada uma estratégia para a combinação dos dados. Seguindo a mesma proposta de combinação dos dados de Marquez, Pedro e Coimbra (2013), o vetor resultante da convolução e os dados da estação meteorológicas são concatenados, formando um único vetor de dados, este vetor será a nova entrada da MLP. Uma das questões existentes é se haveria algum viés na classificação em favor dos dados de satélite, tendo em vista a maior quantidade de atributos extraídos da imagem, em comparação aos dados meteorológicos. Para sanar esta questão, foi considerada uma camada densa na saída da convolução, a qual permitiu extrair, além do resultado da convolução padrão, um conjunto reduzido de 16 atributos, que também foi considerado nos testes.

6.1.3 Camada MLP

A última camada da rede proposta é a MLP, esta camada possui o objetivo de relacionar todos os dados de entrada com a saída. Como função de ativação da MLP

Figura 6.3 – Topologia da camada convolucional.

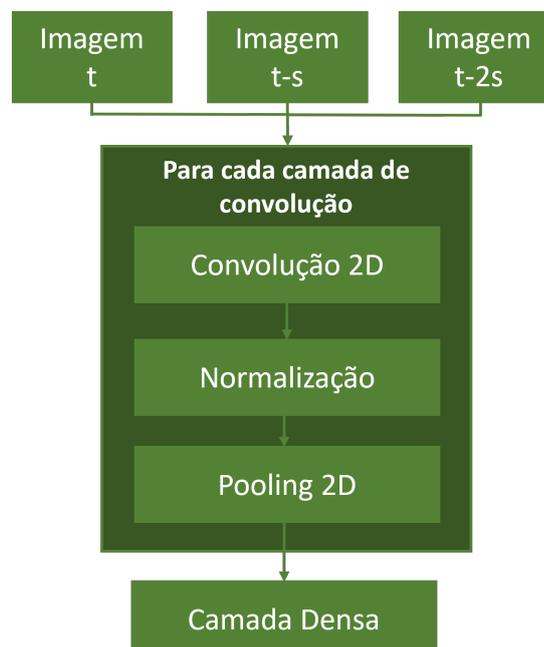
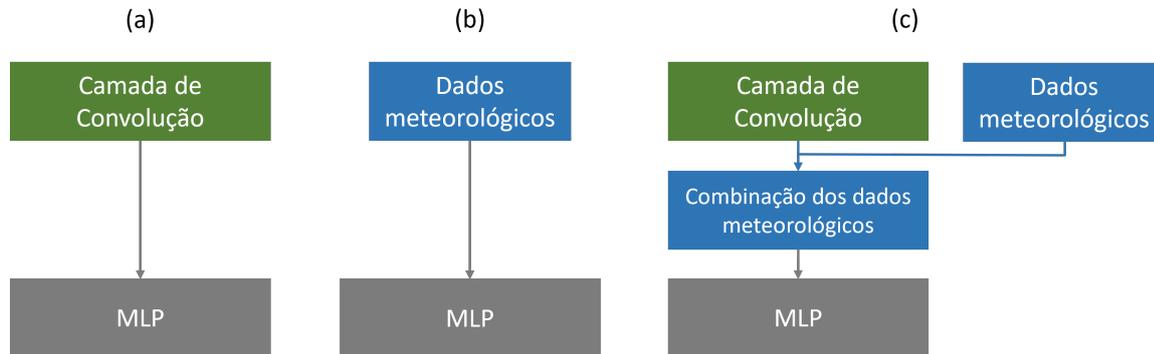


Figura 6.4 – Topologia da camada de combinação dos dados.



foi definido ReLU, já o *dropout*, a quantidade de camadas e neurônios foram limitados baseados em um estudo preliminar apresentado no Apêndice B e definidos para este estudo como sendo:

- Camadas: {1, 2}
- *Dropout*: {0.15}
- Neurônios: {64, 128, 256}

O resultado da MLP é uma previsão no horizonte $t + h$ para o valor de GHI.

6.1.4 Organização dos dados

Definido o intervalo temporal como sendo das 9h às 17:30h, conforme Capítulo 5, com granularidade de 30min e horizonte de predição sendo 30, 60, 120 e 180 minutos, a base de dados foi preparada para ser utilizada na ANN. Como a topologia pode ser variada, os dados foram preparados de forma a atender todos os requisitos.

A composição de cada vetor de dados é dado por: imagens em t , $t - s$ e $t - 2s$ vetorizadas, os dados da estação terrestre: ultravioleta (UV), velocidade do vento (WS), umidade relativa (RH) e temperatura do ar (AT) no intervalo t , e o GHI esperado em $t + h$. Para a granularidade s é utilizado 30min. Para o horizonte h é variado seu valor conforme o horizonte estudado, como é proposto para o estudo 4 horizontes, são preparadas 4 bases de dados conforme a Tabela 6.1.

As bases de dados foram divididas em 3 grupos, 70% dos dados de 2018, com balanceamento homogêneo entre os meses, ou seja, 70% de janeiro, 70% de fevereiro, assim por diante, foi separada para treinar a rede. Os demais 30% dos dados foram separado para a validação durante o treino. O terceiro grupo, dados de 2019 foi separado para testes, ou seja, sem passar pela rede durante o treinamento e utilizado

Tabela 6.1 – Exemplo de representação das bases de dados considerando diferentes horizontes de previsão, $s = 30min$.

Horiz. (h)	Instante (t)	Imagem (t) Entrada	Imagem (t-s) Entrada	Imagem (t-2s) Entrada	Est. Ter. (t) Entrada	GHI (t+h) Saída
30	10/03/18 12:00	Imag. 12:00	Imag. 11:30	Imag. 11:00	Dados 12:00	GHI 12:30
	10/03/18 12:30	Imag. 12:30	Imag. 12:00	Imag. 11:30	Dados 12:30	GHI 13:00
	10/03/18 13:00	Imag. 13:00	Imag. 12:30	Imag. 12:00	Dados 13:00	GHI 13:30
	10/03/18 13:30	Imag. 13:30	Imag. 13:00	Imag. 12:30	Dados 13:30	GHI 14:00
60	10/03/18 12:00	Imag. 12:00	Imag. 11:30	Imag. 11:00	Dados 12:00	GHI 13:00
	10/03/18 12:30	Imag. 12:30	Imag. 12:00	Imag. 11:30	Dados 12:30	GHI 13:30
	10/03/18 13:00	Imag. 13:00	Imag. 12:30	Imag. 12:00	Dados 13:00	GHI 14:00
	10/03/18 13:30	Imag. 13:30	Imag. 13:00	Imag. 12:30	Dados 13:30	GHI 14:30
120	10/03/18 12:00	Imag. 12:00	Imag. 11:30	Imag. 11:00	Dados 12:00	GHI 14:00
	10/03/18 12:30	Imag. 12:30	Imag. 12:00	Imag. 11:30	Dados 12:30	GHI 14:30
	10/03/18 13:00	Imag. 13:00	Imag. 12:30	Imag. 12:00	Dados 13:00	GHI 15:00
	10/03/18 13:30	Imag. 13:30	Imag. 13:00	Imag. 12:30	Dados 13:30	GHI 15:30
180	10/03/18 12:00	Imag. 12:00	Imag. 11:30	Imag. 11:00	Dados 12:00	GHI 15:00
	10/03/18 12:30	Imag. 12:30	Imag. 12:00	Imag. 11:30	Dados 12:30	GHI 15:30
	10/03/18 13:00	Imag. 13:00	Imag. 12:30	Imag. 12:00	Dados 13:00	GHI 16:00
	10/03/18 13:30	Imag. 13:30	Imag. 13:00	Imag. 12:30	Dados 13:30	GHI 16:30

para análise dos resultados. Todas as bases de dados foram disponibilizadas em um repositório online¹.

6.2 PARÂMETROS DE TREINAMENTO

O treinamento é feito utilizando a otimização Adam com decaimento exponencial, sendo seus parâmetros de passo de aprendizado 0.001, passos de decaimento 10000, taxa de decaimento $1e^{-7}$. A função de perda utilizada foi o MSE e como métrica o RMSE. A rede foi treinada até um limite de 10000 épocas, mas utilizando uma função de parada antecipada com 500 épocas de persistência, com tamanho de lote de 512. O framework utilizado foi o Tensorflow (ABADI et al., 2015) com a API do Keras (CHOLLET et al., 2015).

Para identificar qual o melhor conjunto de parâmetros e topologia da ANN, foram testadas 552 combinações diferentes para cada horizonte de predição, sendo elas, 48 combinações somente com imagem, 96 combinações para cada um dos 4 conjuntos de dados climatológicos com imagem e 6 combinações para cada um dos 4 conjuntos de dados climatológicos, totalizando 2208 combinações. E mais as 6400 combinações executadas no estudo preliminar, Apêndice B. Os treinamentos foram conduzidos utilizando a plataforma Colab do Google.

¹ <https://doi.org/10.5281/zenodo.7954973>

6.3 EXPERIMENTOS

Conforme abordado na seção anterior foram conduzidos teste para 4 horizontes, 3 topologias de rede e 7 combinações de parâmetros. Nesta seção são detalhados cada um dos conjuntos de treinamento separados em 3 seções, sendo elas Seção 6.3.1 utilizando apenas imagem de satélite, Seção 6.3.2 utilizando apenas dados da estação terrestre e Seção 6.3.3 a combinação dos dados de satélite com a estação terrestre.

6.3.1 Previsões baseadas em dados de satélite

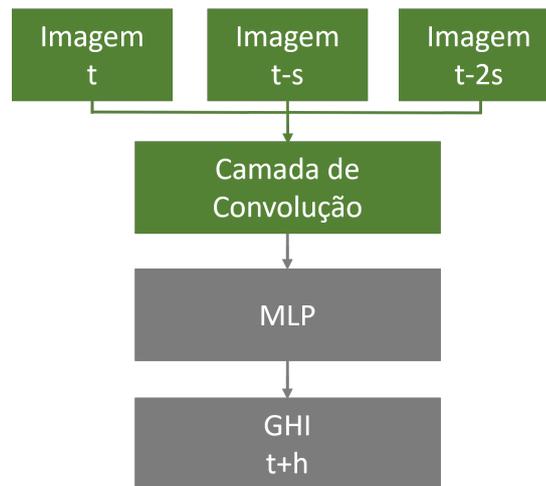
Este teste foi considerado como a base de referência do estudo, utilizar apenas das imagens de satélite para verificar o comportamento da rede e a melhora do desempenho ao adicionar outros parâmetros. Vamos chamar este teste de Conjunto *Img*. O diagrama da Figura 6.5 apresenta a topologia utilizada para este teste. Foram considerados os parâmetros:

- Camadas de convolução: {1, 2}
- Filtros: {16, 32}
- *Kernel*: {3, 5}
- Camada densa: {sem}
- Camadas: {1, 2}
- *Dropout*: {0.15}
- Neurônios: {64, 128, 256}

note que para este conjunto de parâmetros não foi utilizada a camada densa na saída da convolução, isto porque os dados não precisam ser combinados a base da estação terrestre e entram direto na MLP.

Como base de comparação para o primeiro estudo foram adicionados os atributos de dia do ano e minutos do dia na rede, seguindo a topologia do diagrama da Figura 6.2. Os parâmetros utilizados para este estudo foram os mesmos que o anterior com a adição da camada densa de 16 neurônios, isto porque faremos a combinação dos dados temporais com os dados de saída da convolução. Vamos chamar este teste de Conjunto *Img+t*. Se considerarmos a combinação de todos os parâmetros apresentados para cada conjunto, foram realizados 48 combinações para o Conjunto *Img* e 96 combinações para o Conjunto *Img+t* para cada horizonte de predição.

Figura 6.5 – Topologia da rede considerando apenas imagens de satélite.



6.3.2 Previsões baseadas em dados da estação terrestres

O segundo lote de testes tem o objetivo de validar se um determinado conjunto de dados da estação sozinho possui desempenho superior ou inferior ao adicionar as imagens do satélite, por este motivo ele é utilizado como base para comparação do resultado combinado. Este conjunto de testes foi dividido em 4 conjuntos sendo cada um deles a combinação de dados da estação terrestre como se segue:

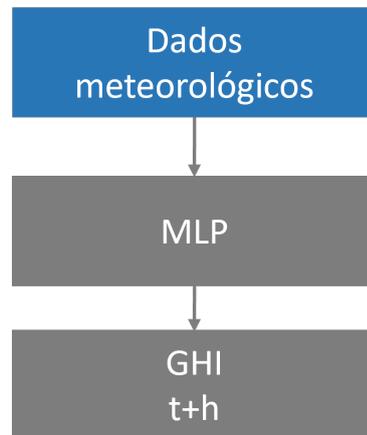
- Conjunto UV+t - Irradiação UV, dia do ano e minutos do dia
- Conjunto UV+AT+RH+WS+t - Irradiação UV, Temperatura do ar, umidade relativa, velocidade do vento, dia do ano e minutos do dia
- Conjunto AT+RH+WS+t - Temperatura do ar, umidade relativa, velocidade do vento, dia do ano e minutos do dia
- Conjunto AT+RH+t - Temperatura do ar, umidade relativa, dia do ano e minutos do dia

e, cada um dos conjuntos foi treinado variando os parâmetros:

- Camadas: {1, 2}
- *Dropout*: {0.15}
- Neurônios: {64, 128, 256}

podemos notar que que neste caso os parâmetros relacionados as convoluções não estão presente, isto porque a topologia não necessita das imagens de satélite, resultando na topologia do diagrama da Figura 6.6.

Figura 6.6 – Topologia da rede considerando apenas dados da estação terrestre.



Para este lote de treinamento são determinadas 6 combinações para cada Conjunto UV+t, UV+AT+RH+WS+t, AT+RH+WS+t e AT+RH+t sendo esta combinação replicada para cada horizonte de predição.

6.3.3 Previsões baseadas na combinação dos dados de satélite com a estação terrestre

Como último lote de testes são replicadas as configurações dos testes da Seção 6.3.2 com a adição das imagens de satélite, resultando em 4 conjuntos como se segue:

- Conjunto $Img+UV+t$ - Imagens de satélite, irradiação UV, dia do ano e minutos do dia
- Conjunto $Img+UV+AT+RH+WS+t$ - Imagens de satélite, irradiação UV, Temperatura do ar, umidade relativa, velocidade do vento, dia do ano e minutos do dia
- Conjunto $Img+AT+RH+WS+t$ - Imagens de satélite, temperatura do ar, umidade relativa, velocidade do vento, dia do ano e minutos do dia
- Conjunto $Img+AT+RH+t$ - Imagens de satélite, temperatura do ar, umidade relativa, dia do ano e minutos do dia

e, cada um dos conjuntos foi treinado variando os parâmetros:

- Camadas de convolução: {1, 2}
- Filtros: {16, 32}
- *Kernel*: {3, 5}
- Camada densa: {sem, 16}

- Camadas: {1, 2}
- *Dropout*: {0.15}
- Neurônios: {64, 128, 256}

Para este lote de testes é utilizada a topologia completa conforme diagrama da Figura 6.2, e são utilizados todos os parâmetros de treinamento. Como resultado neste conjunto de treinamento são determinadas 96 combinações para cada Conjunto Img+UV+t , Img+UV+AT+RH+WS+t , Img+AT+RH+WS+t e Img+AT+RH+t sendo esta combinação replicada para cada horizonte de predição.

6.4 CONSIDERAÇÕES FINAIS

Este capítulo apresentou a metodologia utilizada para a execução dos testes da rede proposta, foram abordados todos os elementos da topologia, os parâmetros que estão sendo variados e avaliados, os conjuntos de dados que são utilizados como entrada e como foram construídos, os parâmetros de treinamento e avaliação da rede e a estrutura de software utilizada. Com as informações apresentadas neste capítulo é esperado que o leitor consiga replicar o estudo proposto.

7 RESULTADOS

Este Capítulo apresenta os resultados dos experimentos propostos no Capítulo 6. A Seção 7.1 expõem as melhores combinações de parâmetros para cada configuração e determina os modelos que serão utilizados para as análises. A Seção 7.2 discute os resultados obtidos com a inclusão dos atributos temporais no modelo. A Seção 7.3 aborda a problemática dos resultados obtidos com a combinação dos dados da estação terrestre e relaciona com os dados de satélite, fazendo uma análise de cada conjunto. A Seção 7.4 discute os principais aspectos da adição da componente de irradiação UV, sendo que esta componente por si só possui uma correlação forte com o GHI. O capítulo termina na Seção 7.5 onde são apresentados os principais resultados em um comparativo com dados encontrados na literatura.

7.1 SELEÇÃO DE MODELOS PARA ANÁLISE DOS RESULTADOS

Para a avaliação dos resultados obtidos foram selecionados as configurações com o melhor RMSE para cada conjunto e horizonte de tempo. Os resultados são apresentados na coluna RMSE nas Tabelas 7.1, 7.2, 7.3 e 7.4, para o horizonte de 30, 60, 120 e 180 minutos, respectivamente. Para cada horizonte foi destacado o melhor resultado de RMSE, importante observar que para todos os horizontes o melhor resultado foi observado no Conjunto *Img+UV+t*.

Tabela 7.1 – Configuração e RMSE dos melhores resultados para cada conjunto em um horizonte de 30min.

#	RMSE <i>W/m²</i>	Conv.	Filtros	Kernel	Fusão Dados	Camadas MLP	Neurônios
<i>Img</i>	182.27	1	32	3	-	1	64
<i>Img+t</i>	128.21	1	16	3	16	2	128
<i>UV+t</i>	101.08	-	-	-	-	2	256
<i>UV+AT+RH+WS+t</i>	101.54	-	-	-	-	2	128
<i>AT+RH+WS+t</i>	180.95	-	-	-	-	2	256
<i>AT+RH+t</i>	188.41	-	-	-	-	2	256
<i>Img+UV+t</i>	94.36	1	16	5	0	2	256
<i>Img+UV+AT+RH+WS+t</i>	94.55	1	16	3	0	2	256
<i>Img+AT+RH+WS+t</i>	124.39	1	16	3	16	2	256
<i>Img+AT+RH+t</i>	123.30	1	16	3	16	2	256

Note que o conjunto *Img* pôr se tratar só de imagem não possui a camada de fusão de dados, assim como os Conjuntos de *UV+t*, *UV+AT+RH+WS+t*, *AT+RH+WS+t* e *AT+RH+t* são apenas dados da estação terrestre, por isto não possuem o conjunto convolucional e a camada de fusão de dados. A camada de fusão de dados está dividida em 2 grupos, sendo eles 0 quando não é adicionada a camada densa e 16 quando é adicionada a camada com 16 neurônios. O Apêndice C traz mais detalhes dos resultados com a quantidade de parâmetros treinados, épocas e tempo de trei-

Tabela 7.2 – Configuração e RMSE dos melhores resultados para cada conjunto em um horizonte de 60min.

#	RMSE W/m^2	Conv.	Filtros	Kernel	Fusão Dados	Camadas MLP	Neurônios
Img	205.85	1	32	5	-	1	128
Img+t	130.87	1	16	3	16	2	256
UV+t	126.38	-	-	-	-	2	256
UV+AT+RH+WS+t	131.67	-	-	-	-	2	256
AT+RH+WS+t	188.56	-	-	-	-	2	256
AT+RH+t	188.04	-	-	-	-	2	256
Img+UV+t	120.33	1	16	5	0	2	256
Img+UV+AT+RH+WS+t	121.24	1	16	5	0	2	128
Img+AT+RH+WS+t	134.40	1	32	5	16	2	128
Img+AT+RH+t	135.24	1	16	3	16	2	128

Tabela 7.3 – Configuração e RMSE dos melhores resultados para cada conjunto em um horizonte de 120min.

#	RMSE W/m^2	Conv.	Filtros	Kernel	Fusão Dados	Camadas MLP	Neurônios
Img	234.94	1	16	5	-	1	64
Img+t	148.80	1	16	3	16	2	256
UV+t	149.14	-	-	-	-	2	128
UV+AT+RH+WS+t	161.38	-	-	-	-	2	256
AT+RH+WS+t	187.18	-	-	-	-	2	256
AT+RH+t	192.48	-	-	-	-	2	256
Img+UV+t	143.86	1	16	3	0	2	128
Img+UV+AT+RH+WS+t	144.56	1	32	5	0	2	256
Img+AT+RH+WS+t	150.15	1	16	3	16	2	128
Img+AT+RH+t	149.54	1	16	3	16	2	256

Tabela 7.4 – Configuração e RMSE dos melhores resultados para cada conjunto em um horizonte de 180min.

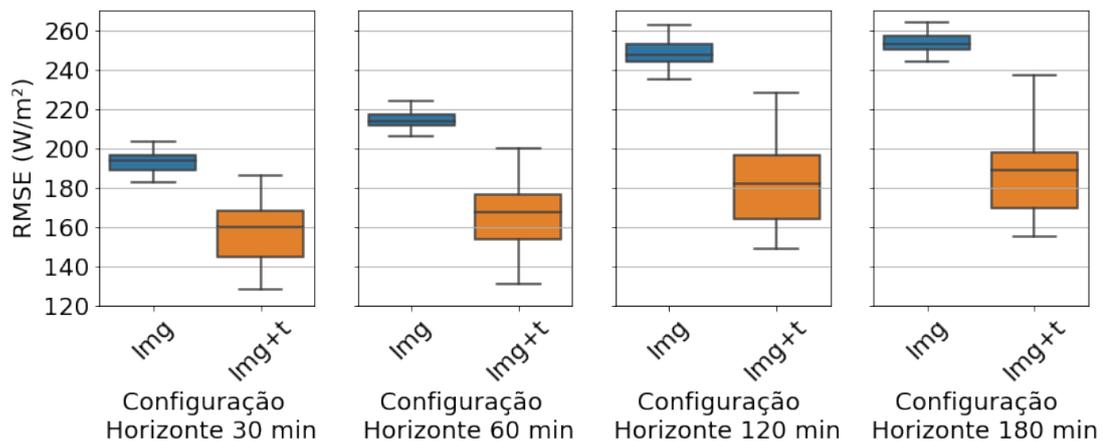
#	RMSE W/m^2	Conv.	Filtros	Kernel	Fusão Dados	Camadas MLP	Neurônios
Img	244.06	1	16	5	-	1	256
Img+t	154.82	1	32	5	0	2	256
UV+t	154.43	-	-	-	-	2	64
UV+AT+RH+WS+t	164.72	-	-	-	-	2	256
AT+RH+WS+t	187.71	-	-	-	-	2	128
AT+RH+t	188.92	-	-	-	-	2	256
Img+UV+t	150.33	1	32	3	0	2	128
Img+UV+AT+RH+WS+t	153.40	1	32	5	0	2	256
Img+AT+RH+WS+t	154.50	1	32	3	16	2	128
Img+AT+RH+t	154.87	1	32	3	0	2	128

namento, e apresenta as 4 melhores configurações para cada conjunto e horizonte, além de avaliar de forma breve a distribuição do RMSE em gráficos de caixa para determinar os parâmetros de maior influência para cada resultado.

7.2 INFLUÊNCIA DE PARÂMETROS TEMPORAIS

Inicialmente foi investigado a influência da adição de informações temporais na rede para identificar a interferência no resultado. Conforme apresentado nos capítulos anteriores, o parâmetro temporal afeta a refletância da imagem ao longo do dia e varia durante o ano devido as estações. Como resultado temos os gráficos da Figura 7.1 que apresentam a distribuição do RMSE de todas as configurações treinadas. Podemos observar que para todas as configurações em todos os horizontes de tempo a inclusão do dia do ano e minutos do dia melhoram significante o resultado, chegando a 29,7%, 36,4%, 36,6% e 36,6%, para o horizonte de 30, 60, 120 e 180 minutos respectivamente, nas melhores combinações de parâmetros.

Figura 7.1 – Distribuição dos resultados de RMSE para todos os modelos treinados nas configurações apresentadas, no ano de 2019.



Se olharmos individualmente para as melhores configurações em uma distribuição do erro, Figura 7.2, fica evidente no resultado que a concentração de erro nas extremidades reduz significativamente, fato visível com a curva mais acentuada, e podemos observar que a concentração no ponto médio 0, teve um aumento significativo.

Ao avaliar a série ao longo do dia, Figura 7.3, notamos uma melhora quase homogênea no RMSE para todos os horizontes, porém com destaque para as previsão ao final do dia nos horizontes de 60, 120 e 180 minutos, que tiveram ganhos mais acentuados nas ultimas predições. Este resultado reforça a importância do parâmetro temporal, ao final do dia a refletância tende a diminuir, não somente pela oclusão das nuvens mas pela redução da exposição do sol, conforme apresentado no Capítulo 5.

Estendendo a análise para os meses do ano, Figura 7.4, fica evidente uma descontinuidade no ganho de desempenho para o mês de fevereiro no horizonte de 60min, para este ponto os dois modelos apresentaram um RMSE muito próximos. Contudo, podemos observar uma melhora mais acentuada nos meses de outubro a janeiro, ficando mais evidente no horizonte de 30 min, podemos atribuir o resultado a inclusão

Figura 7.2 – Densidade do erro para os melhores modelos no ano de 2019.

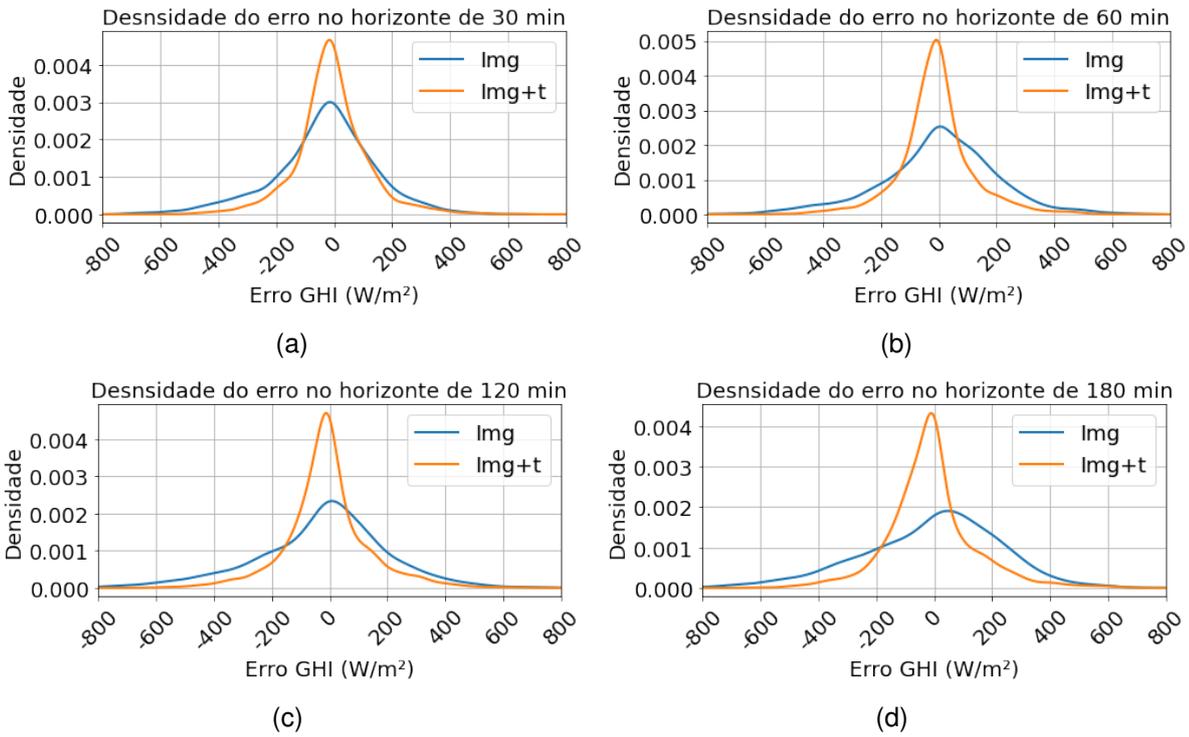
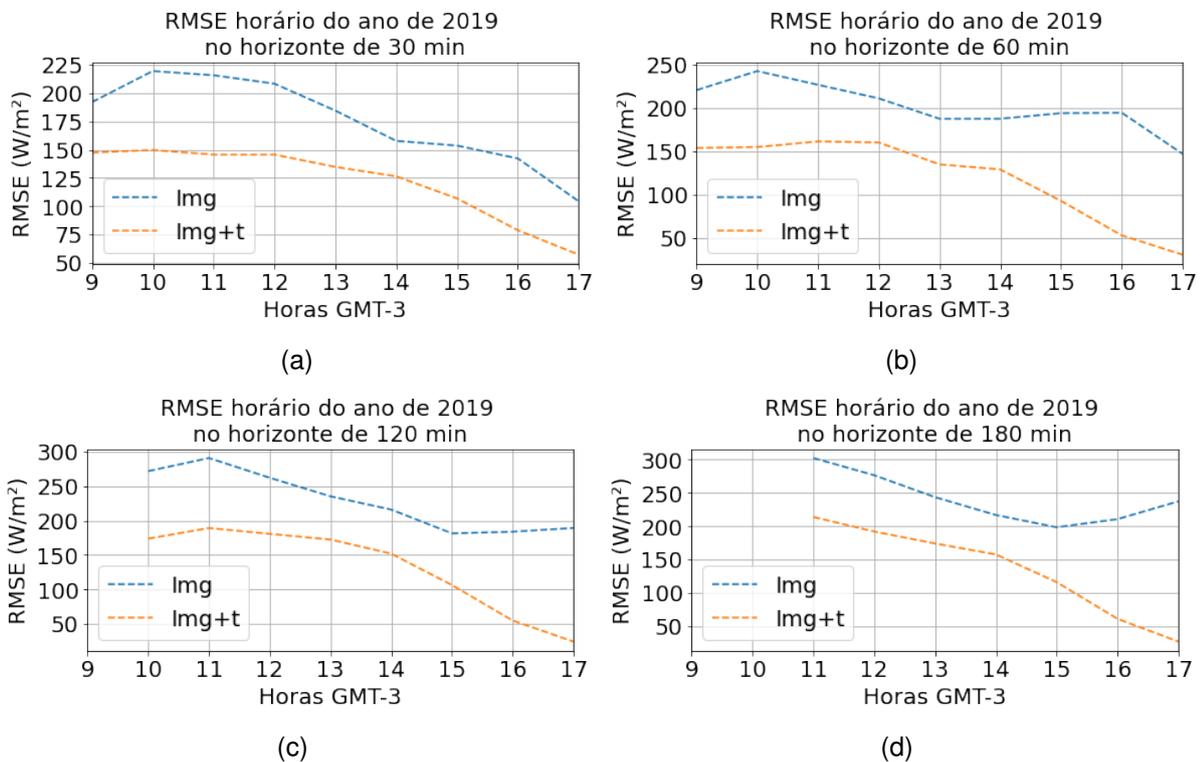
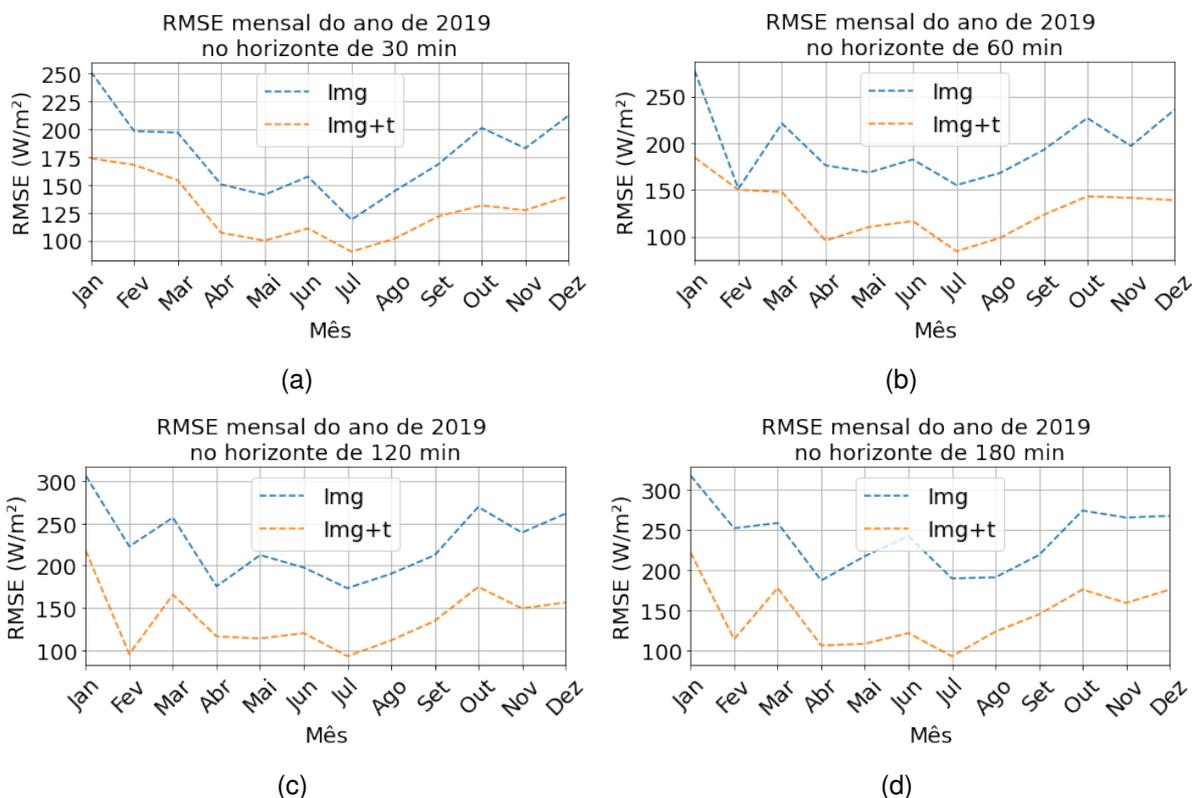


Figura 7.3 – RMSE horário para o ano de 2019, onde o tempo é momento da predição e o resultado se refere a predição mais o horizonte.



dos dias do anos, que auxiliam a rede a correlacionar a variação da intensidade de irradiação ao longo do ano com o GHI.

Figura 7.4 – RMSE mensal para o ano de 2019.



Por conta deste resultado, todas as outras redes foram avaliadas com a inclusão dos dados temporais de dia do ano e minutos do dia, conforme apresentado no capítulo anterior, e todas as análises seguintes foram feitas em relação ao Conjunto *Img+t*, que contempla apenas as imagens de satélite e dados temporais.

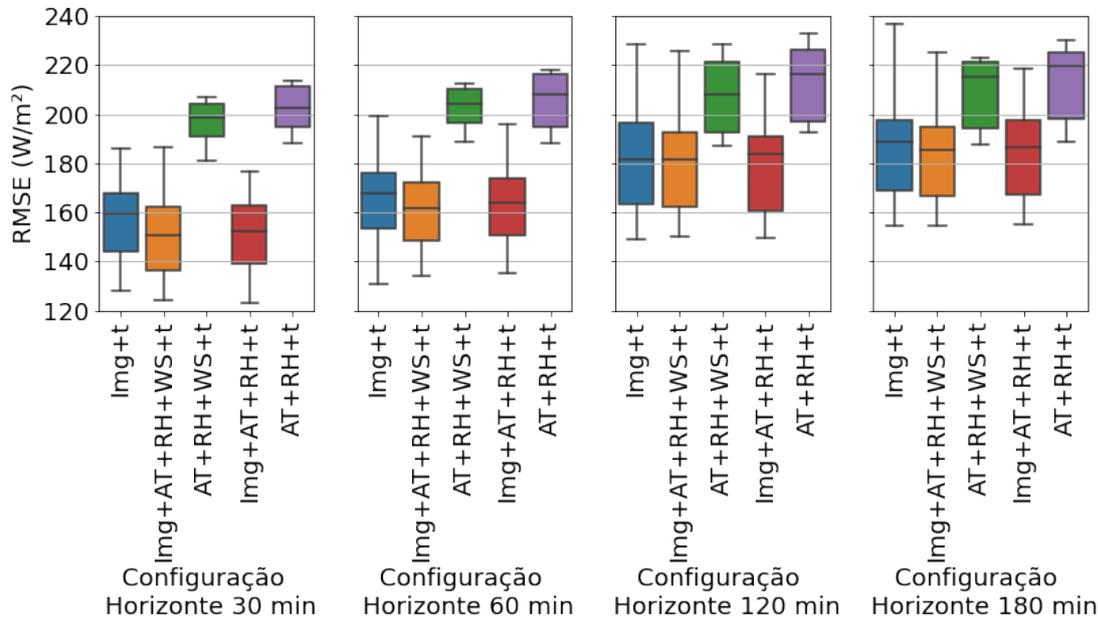
7.3 INFLUÊNCIA DOS DADOS DA ESTAÇÃO TERRESTRE

Para avaliar a influência da combinação dos dados da estação terrestre e entender o ganho de desempenho foram treinados modelos de conjuntos com e sem a presença de imagens de satélite. Nesta seção analisaremos a influência da velocidade do vento, umidade relativa e temperatura do ar, faremos uma análise da influência da irradiação ultravioleta na seção seguinte.

Inicialmente podemos avaliar o gráfico da Figura 7.5, note que para todos os horizontes os Conjuntos *Img+t*, *Img+AT+RH+WS+t* e *Img+AT+RH+t* tiveram resultados muito similares, isso mostra que a combinação dos dados da estação terrestre não teve um ganho significativo nos modelos treinados. Quando comparado com os Conjuntos *AT+RH+WS+t* e *AT+RH+t*, que são dados sem a presença de imagens de satélite, fica evidente o ganho de desempenho que a inclusão das imagens ocasiona no modelo, quando comparado com os modelos sem imagem, além disso no horizonte de 60 e 120 min o Conjunto *Img+t* demonstrou um desempenho levemente superior

aos demais.

Figura 7.5 – Distribuição dos resultados de RMSE para todos os modelos treinados nas configurações apresentadas, no ano de 2019.



É possível observar que com o aumento do horizonte o RMSE dos modelos com as imagens de satélite começa a ter um aumento, enquanto os melhores resultados dos conjuntos sem imagem, permanecem com valores de RMSE similares para os horizontes. É difícil determinar a causa exata, mas dentre as possibilidades podemos citar que o aumento do horizonte torna a janela de captura pequena para conseguir extrair informações das imagens de satélite para o horizonte, ou que em um horizonte mais distante as imagens de satélite começam a ter menor correlação com o GHI do que os parâmetros da estação terrestre. Esta condição fica evidente ao analisar os melhores resultados de cada conjunto, note na Figura 7.6, que a predição considerando apenas dados da estação terrestre possui um RMSE até 46,95% inferior no horizonte de 30min.

Ao avaliar as séries ao longo do dia, gráfico da Figura 7.7, fica evidente uma assertividade maior para as predições ao final do dia para todos os horizontes, e podemos destacar que com o aumento do horizonte a uma aproximação do resultado com ou sem imagens de satélite, ou seja dados puramente da estação meteorológica, puramente do satélite ou a combinação dos dois. É possível observar para o horizonte de 30min que os dados da estação terrestre trouxeram ganhos mais significativos no início do dia, das 9h às 11h.

Olhando as séries ao longo dos meses, gráfico da Figura 7.8, podemos observar no horizonte de 30min uma assertividade maior do Conjunto $\text{Img}+\text{AT}+\text{RH}+\text{t}$ no mês de fevereiro e março. O desempenho dos modelos sem imagens de satélite nos

Figura 7.6 – Variação do RMSE dos modelos em relação ao Conjunto lmg+t.

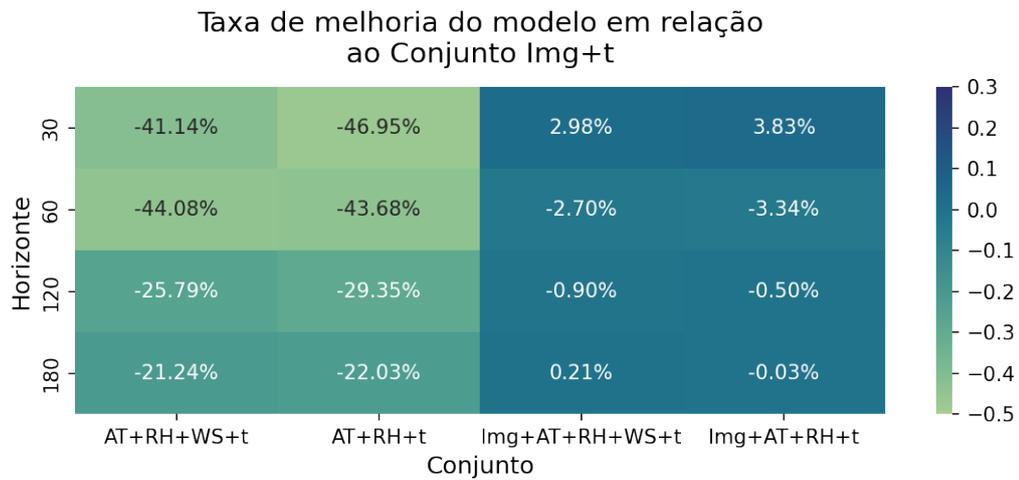
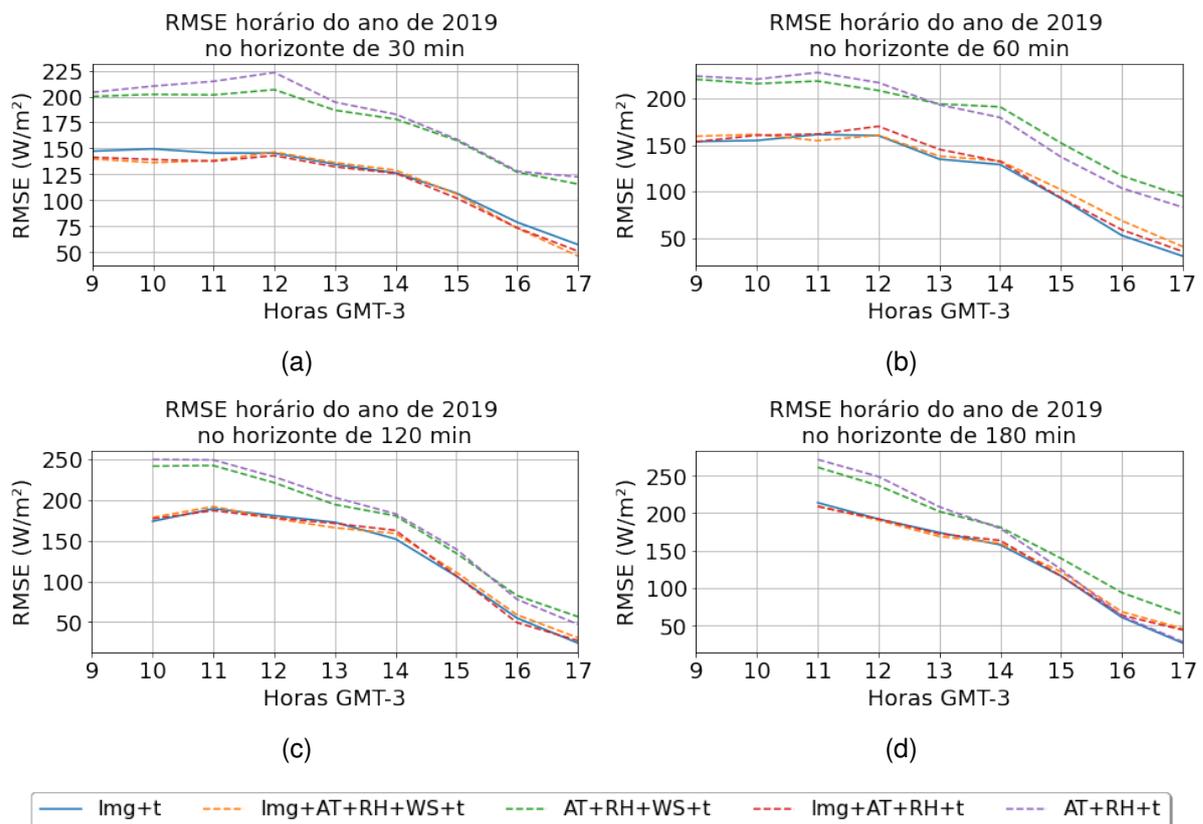
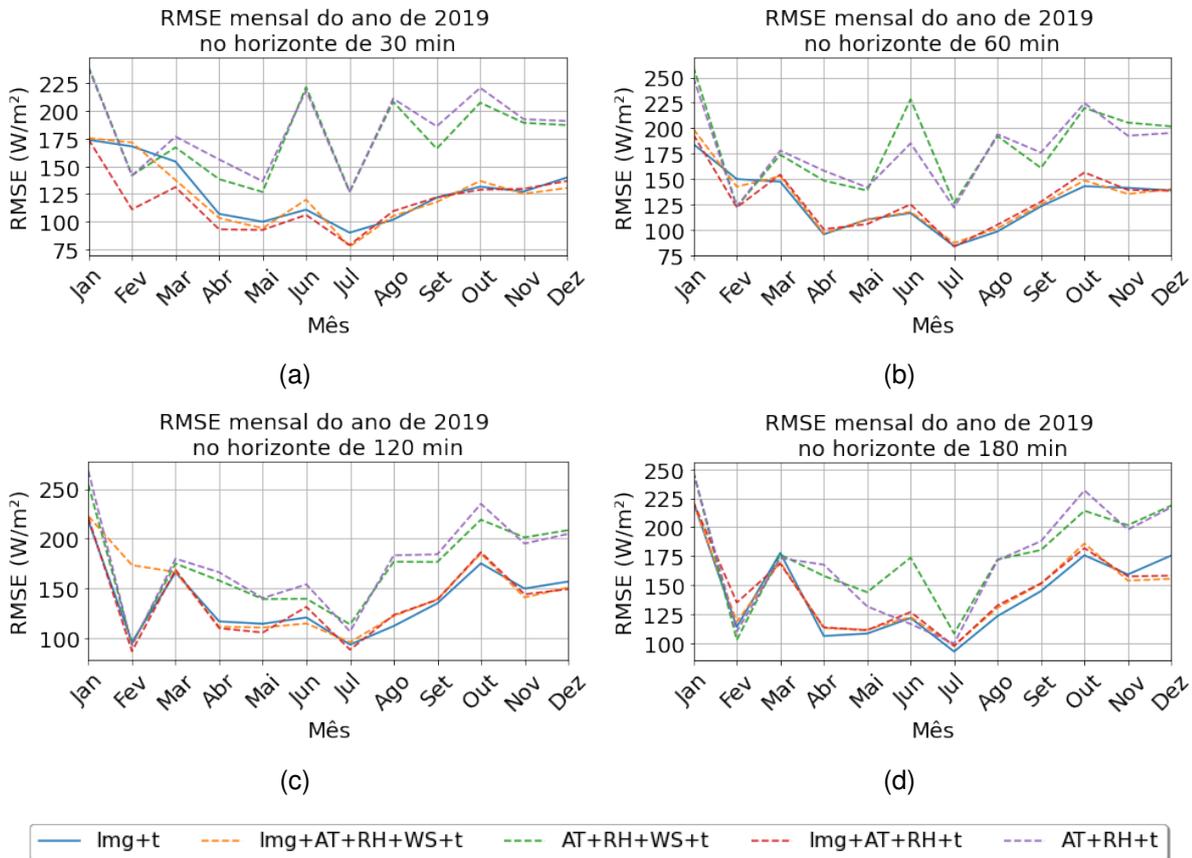


Figura 7.7 – RMSE horário para o ano de 2019, onde o tempo é momento da predição e o resultado se refere a predição mais o horizonte.



horizontes de 30min e 60min apresentaram um pior desempenho nos meses de junho, agosto e outubro, este mesmo resultado não fica tão acentuado nos horizontes de 120min e 180min.

Figura 7.8 – RMSE mensal para o ano de 2019.

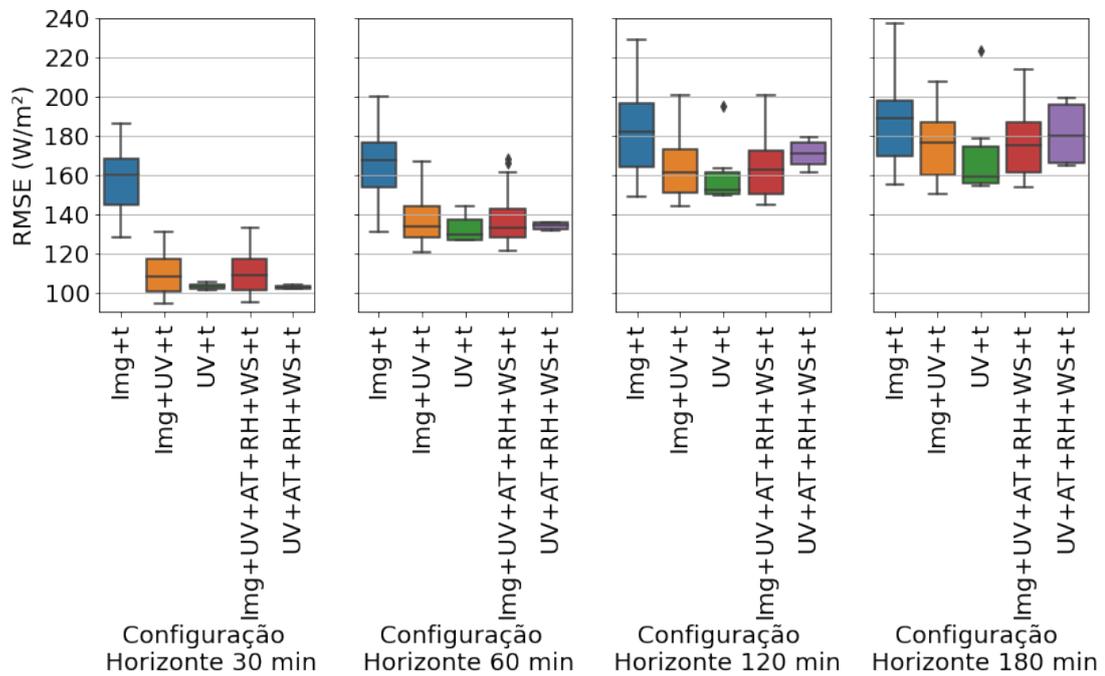


7.4 INFLUÊNCIA DO UV

Como a irradiação ultravioleta apresentou uma correlação forte com o GHI, foi separada esta seção dedicada para analisar o desempenho desta componente no modelo. Separamos a análise em dois conjuntos, inicialmente vamos tratar apenas imagens de satélite e a componente UV, no segundo momento vamos adicionar as demais leituras da estação terrestre ao estudo. Começamos a análise avaliando o desempenho dos modelos nos Conjuntos lmg+t, lmg+UV+t e UV+t, sendo o primeiro apenas imagens de satélite, o segundo a combinação de imagens de satélite e UV e o terceiro apenas UV. No gráfico da Figura 7.9, observamos um aumento significativo no desempenho do modelo em todos os horizontes ao adicionar a componente UV ao estudo, contudo esta influência é muito mais forte no horizonte de 30min, resultado esperado, como vimos nos capítulos anterior a correlação do GHI com o UV é forte, mas pode perder sua eficiência em horizontes maiores, sendo que o valor de UV deste instante pode ter menos influência na irradiação futura.

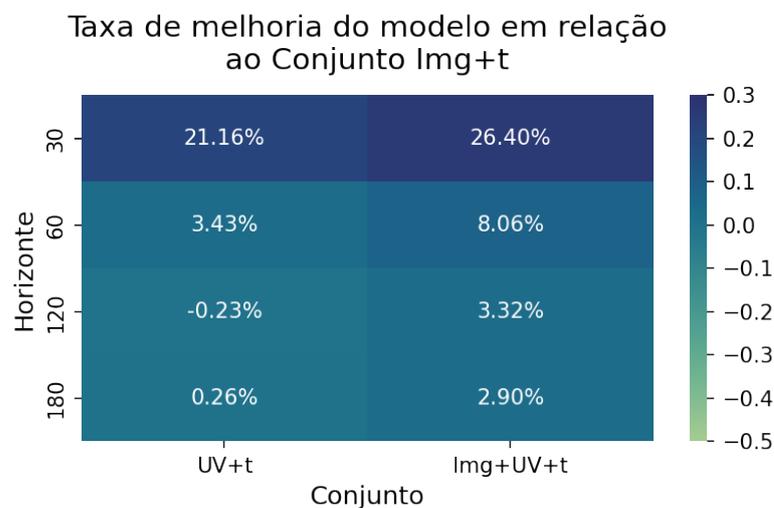
Ao selecionar os melhores modelos e avaliar o desempenho individual para cada horizonte, Figura 7.10, fica evidente o ganho de desempenho no horizonte de 30 min, 21,16% utilizando somente a componente UV e 26,4% ao combinar o UV com

Figura 7.9 – Distribuição dos resultados de RMSE para todos os modelos treinados nas configurações apresentadas, no ano de 2019.



os dados de satélite. Porém quando olhamos para horizontes maiores, a componente UV começa a reduzir sua influência no modelo, sendo que no horizonte de 180min ela apresenta um desempenho levemente superior a imagem de satélite 0,26% e um adicional de 2,9% quando combinada.

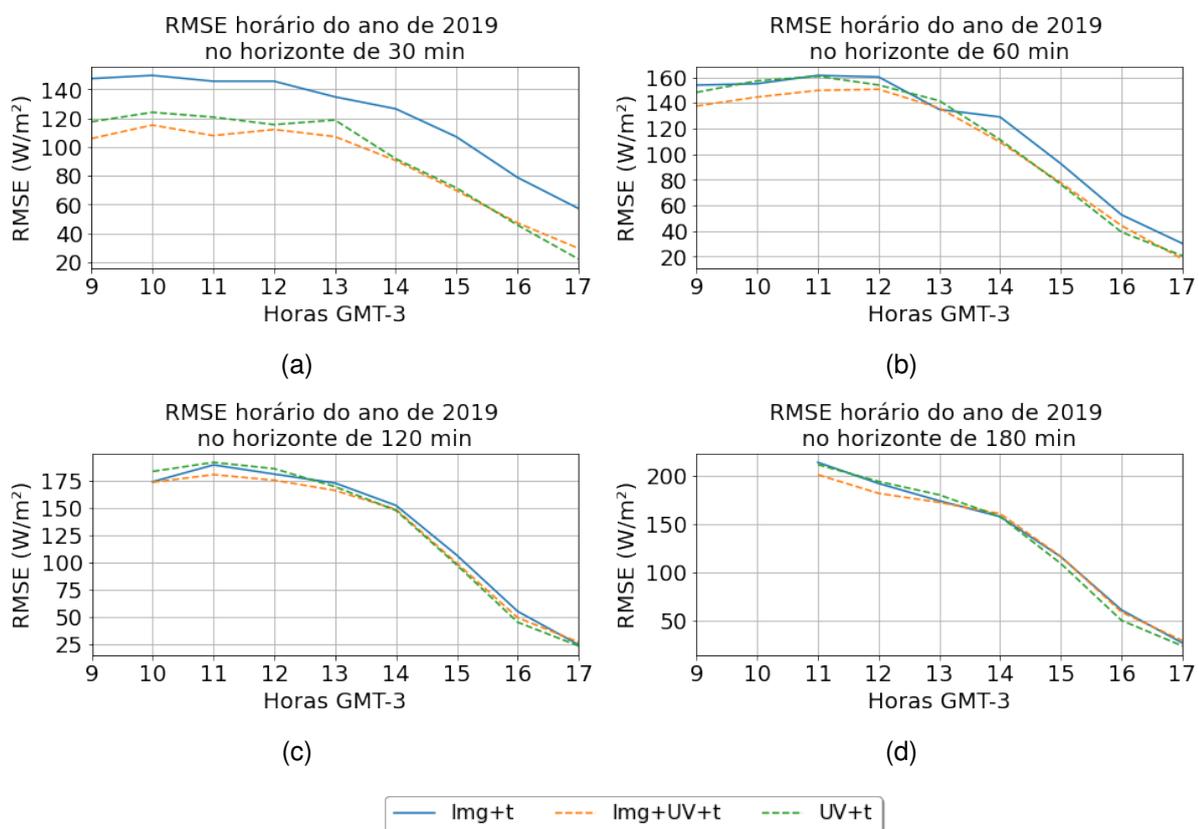
Figura 7.10 – Comparação do ganho de desempenho dos modelos em relação ao Conjunto Img+t, sem dados da estação terrestre.



Quando avaliamos a série temporal do RMSE ao longo do dia, Figura 7.11, o

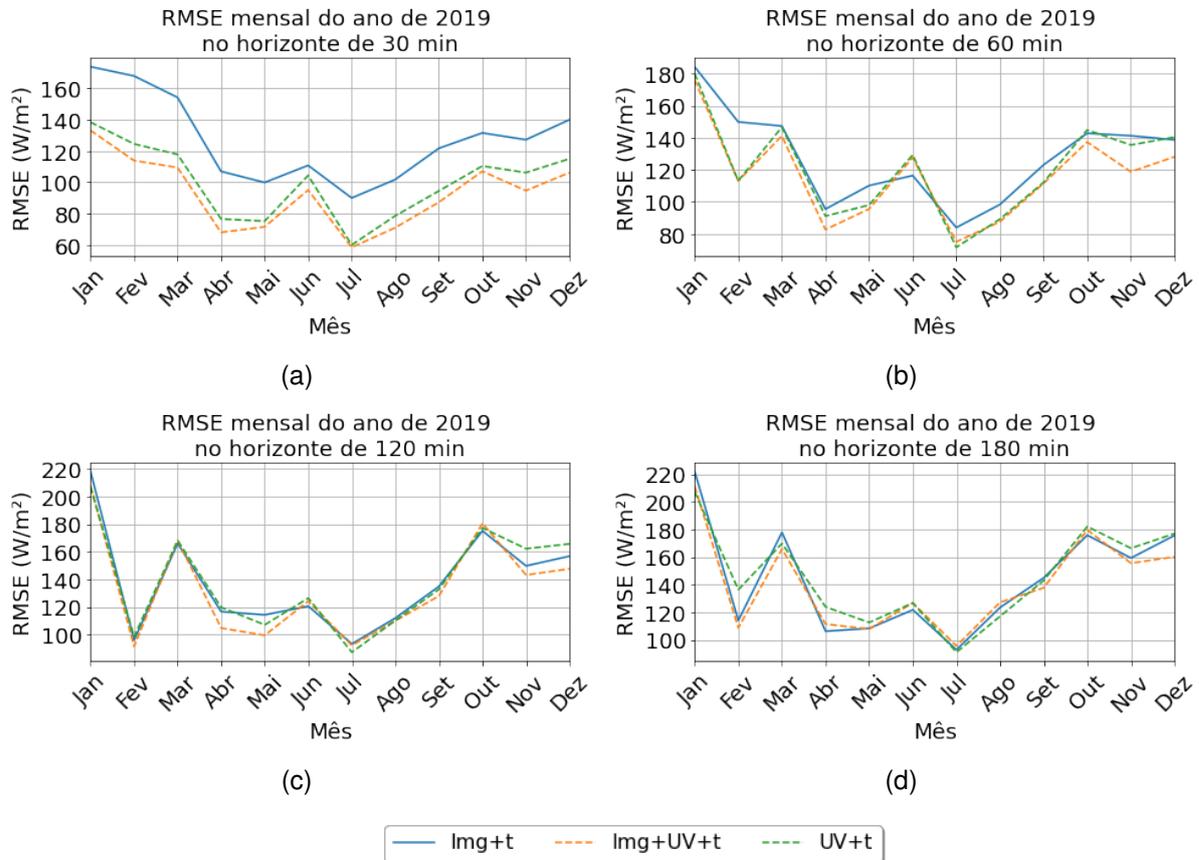
grande destaque vai para o horizonte de 30min que se descola de forma homogênea dos demais modelos quando em combinação com o UV, porém para os demais horizontes o resultado fica muito próximo. A mesma situação se reflete ao analisar a série ao longo dos meses, Figura 7.12, o destaque fica para os meses de janeiro a março no horizonte de 30min, como sendo os meses com maior ganho de desempenho e o mês de fevereiro no horizonte de 60min, podemos apontar este ganho pela maior incidência de irradiação nestes meses, conseqüentemente, maior influência do UV.

Figura 7.11 – RMSE horário para o ano de 2019, onde o tempo é momento da predição e o resultado se refere a predição mais o horizonte.



Para a continuidade da análise, vamos mudar o modelo de referência, até então estávamos avaliando o ganho de desempenho na inclusão dos dados da estação terrestre a imagem de satélite, como nosso objetivo agora é avaliar este mesmo ganho porém com a adição da componente UV, vamos olhar o ganho de desempenho em relação ao Conjunto $lmg+UV+t$, imagem de satélite combinado com a irradiação UV. Podemos observar no gráfico da Figura 7.9, a distribuição de todos os treinamentos para os Conjuntos $lmg+UV+t$, $lmg+UV+AT+RH+WS+t$ e $UV+AT+RH+WS+t$, e já fica claro que a inclusão de outras variáveis ao modelo não melhorou o desempenho, mas se manteve igual, quando olhamos para a distribuição do Conjunto $UV+AT+RH+WS+t$ é perceptível uma pequena queda de desempenho em relação aos demais, mas vale apontar que este conjunto não contempla as imagens de satélite.

Figura 7.12 – RMSE mensal para o ano de 2019.



Olhando os melhores modelos de cada conjunto, Figura 7.13 podemos perceber que nenhuma das combinações teve desempenho superior a combinação de imagens de satélite com UV, quando adicionamos as componentes meteorológicas, tivemos um decréscimo de 7,6% do desempenho no horizonte de 30min e chegando a 12,18% no horizonte de 120min. Mas quando tiramos a imagem de satélite da análise e mantemos apenas as componentes meteorológicas sendo elas, velocidade do vento, temperatura do ar, umidade relativa e irradiação UV, temos um desempenho muito próximo, com uma redução mínima de 0,2% no horizonte de 30min e 2,04% no horizonte de 180min. Este resultado é especialmente interessante, pois observamos que para o estudo sem UV um leve aumento no desempenho, mas quando adicionamos os dados da estação terrestre e UV ao modelo observamos claramente que estes atributos confundem a rede e temos uma queda no desempenho geral.

Quando avaliamos as séries ao longo do dia e dos meses, Figura 7.14 e 7.15 respectivamente, notamos uma piora no desempenho nas predições no início e meio do dia, quando removemos a imagem de satélite, assim como uma piora nos resultados nos meses de agosto a dezembro para o horizonte de 30min e 60min e nos meses de outubro a dezembro para os horizontes de outubro a dezembro.

Figura 7.13 – Comparação do ganho de desempenho dos modelos em relação ao Conjunto lmg+t, sem dados da estação terrestre.

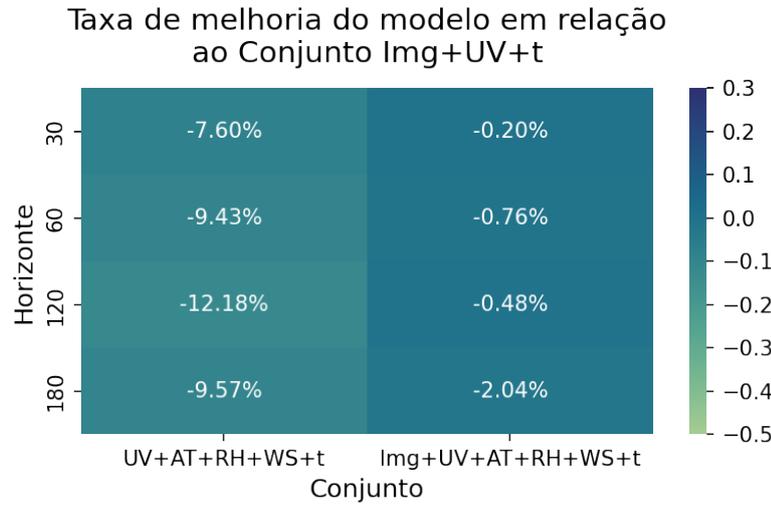
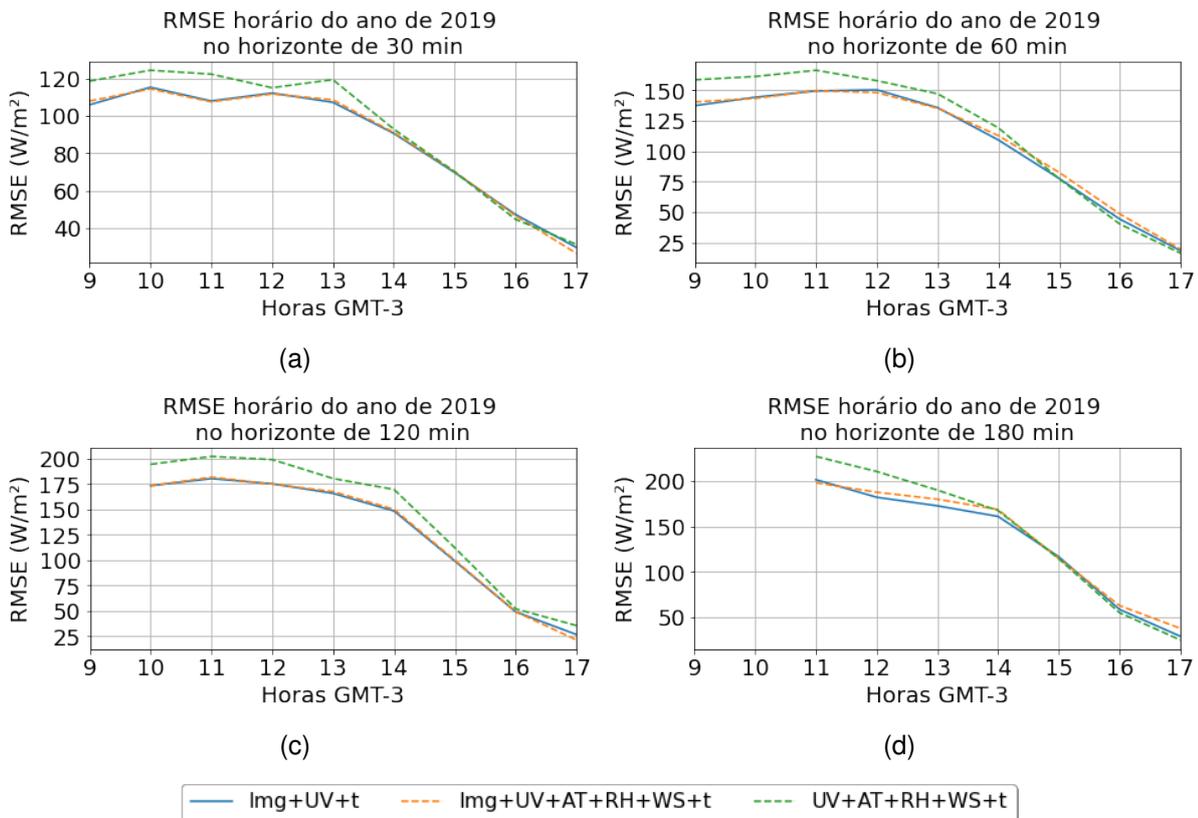


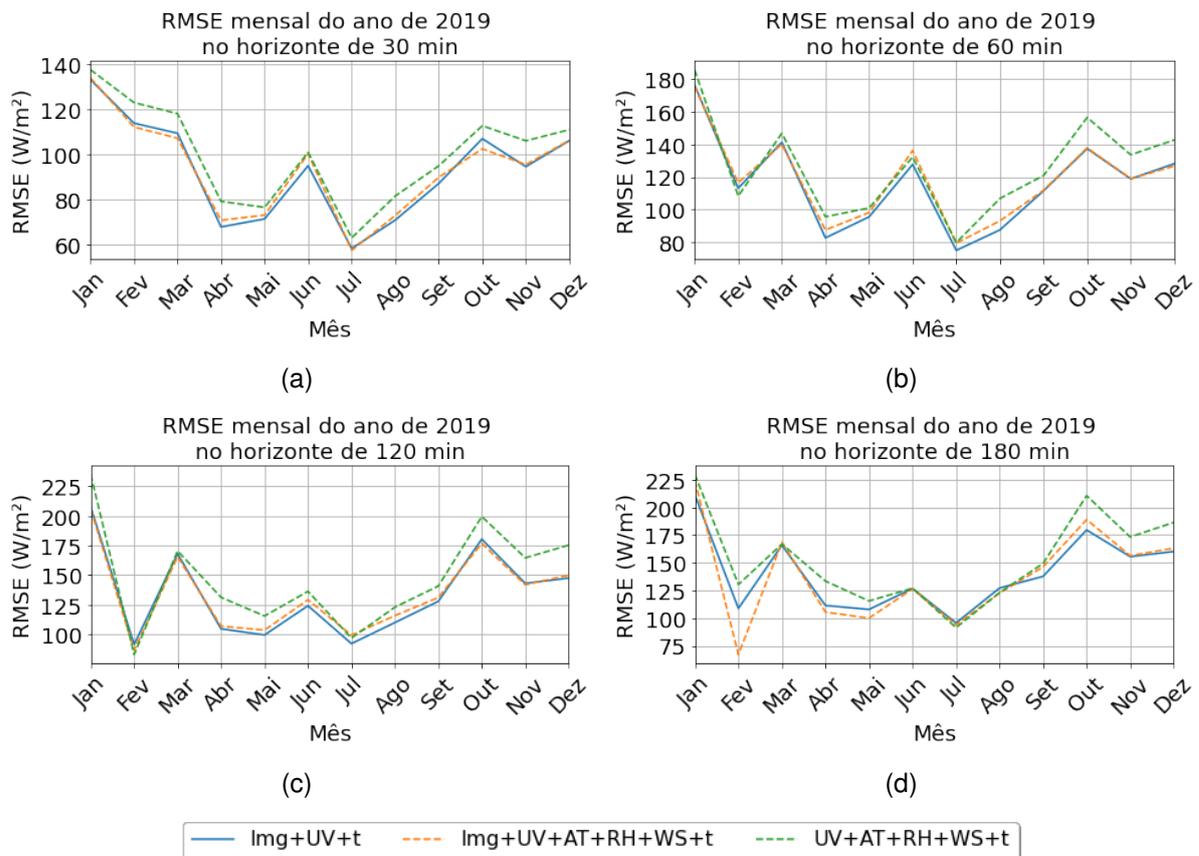
Figura 7.14 – RMSE horário para o ano de 2019, onde o tempo é momento da predição e o resultado se refere a predição mais o horizonte.



7.5 MELHORES RESULTADOS POR CATEGORIA E ANÁLISE GERAL

Para análise dos resultados foram considerados os Conjuntos lmg+t, lmg+AT+RH+t e lmg+UV+t, sendo estes os conjuntos com os melhores resultados

Figura 7.15 – RMSE mensal para o ano de 2019.

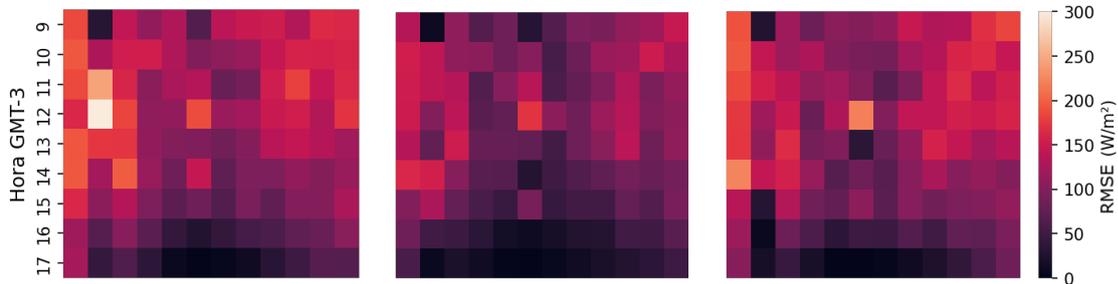


gerais e que conseguem representar de forma geral o estudo. O Conjunto lmg+t apresenta apenas imagens de satélite, o Conjunto lmg+AT+RH+t, combina imagens de satélite com dados da estação terrestre, menos UV, e o Conjunto lmg+UV+t combina os dados de satélite com os dados de irradiação UV. A Figura 7.16 apresenta a distribuição do RMSE do Conjunto lmg+t, onde alguns resultados chamam a atenção para o horizonte de 30min no mês de fevereiro as 12h temos um aumento do RMSE, situação que se repete para o horizonte de 60min porém com menor intensidade, podemos observar que todos os horizontes tiveram melhor desempenho ao final do dia, em especial no período do inverno, e tiveram um desempenho menor próximo do início do dia, situação que se intensifica nos horizontes de 120 e 180min.

Ao adicionarmos os dados da estação meteorológica Conjunto lmg+AT+RH+t, Figura 7.16, tivemos uma leve melhora nos modelos ao final do dia, se intensificando no inverno. Podemos observar um resultado mais acentuado para o horizonte de 30min no mês de junho as 13h e no horizonte de 120min para o mês de junho as 12h, ambos representando a predição para as 14h. Para o horizonte de 120min tivemos uma queda de desempenho do modelo para o mês de janeiro, o mesmo pode ser observado para o horizonte de 180min mês de fevereiro, em especial para as 12h.

Quando adicionada a componente UV no estudo, Conjunto lmg+UV+t, e com-

Figura 7.16 – RMSE dos 3 melhores conjuntos por categoria para os dados de validação em W/m^2 .



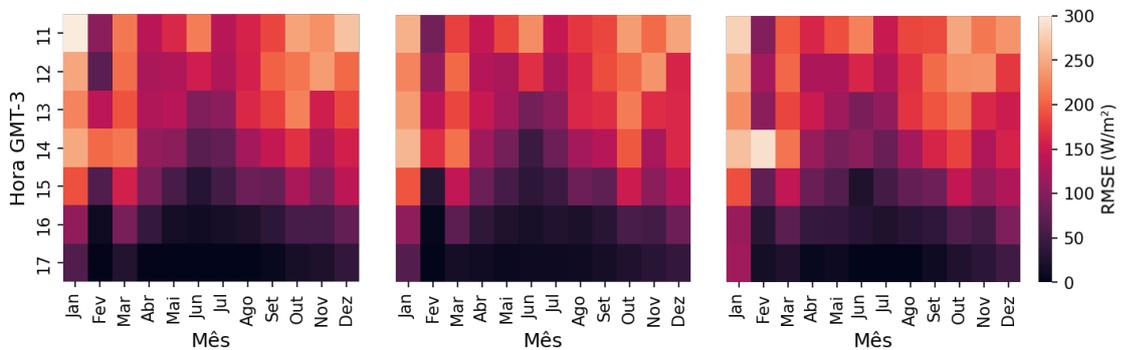
(a) Horizonte de 30min, Conjunto lmg+t. (b) Horizonte de 30min, Conjunto lmg+UV+t. (c) Horizonte de 30min, Conjunto lmg+AT+RH+t.



(d) Horizonte de 60min, Conjunto lmg+t. (e) Horizonte de 60min, Conjunto lmg+UV+t. (f) Horizonte de 60min, Conjunto lmg+AT+RH+t.



(g) Horizonte de 120min, Conjunto lmg+t. (h) Horizonte de 120min, Conjunto lmg+UV+t. (i) Horizonte de 120min, Conjunto lmg+AT+RH+t.



(j) Horizonte de 180min, Conjunto lmg+t. (k) Horizonte de 180min, Conjunto lmg+UV+t. (l) Horizonte de 180min, Conjunto lmg+AT+RH+t.

parada ao Conjunto $\text{Img}+\text{AT}+\text{RH}+\text{t}$, podemos observar uma melhora homogênea e significativa para o horizonte de 30min, similar para horizonte de 60min, porém menor intensidade e quase imperceptível para os horizontes de 120min e 180min, situação esperada conforme vistos nas distribuições do erro nas seções anteriores. Podemos pontuar os resultados do modelo para o horizonte de 120min no mês de fevereiro as 12h onde teve uma correção na descontinuidade observada no modelo anterior, assim como para o horizonte de 120min, mês de janeiro as 13h. Para o final do dia, principalmente no inverno, não tivemos melhoras significativas no modelo.

Ao comparar o resultado com o estudo de Marquez, Pedro e Coimbra (2013), podemos observar que o RMSE alcançado para os Conjuntos $\text{Img}+\text{t}$, $\text{Img}+\text{AT}+\text{RH}+\text{t}$ e $\text{Img}+\text{UV}+\text{t}$ é maior. Importante ressaltar que no estudo mencionado os dados foram coletados de uma estação meteorológica em Merced, nos períodos de 01/01/2011 até 06/06/2011 e 23/11/2011 e 31/01/2012 e Davis no período de 23/11/2011 até 31/01/2012. Foram utilizados para o treinamento 3.756 leituras e 1.032 leitura para validação, sendo que o maior valor de GHI é próximo dos $1.000 \text{ W}/\text{m}^2$ enquanto neste trabalho o maior valor de GHI é de $1.200 \text{ W}/\text{m}^2$ o que influencia diretamente o resultado do RMSE alcançado. Os resultados são apresentados na Tabela 7.5.

Tabela 7.5 – Comparação dos resultados com o estudo de Marquez, Pedro e Coimbra (2013). Os valores representam o RMSE dos modelos e expressos em W/m^2 .

Modelo	Horizonte			
	30min	60min	120min	180min
Marquez, Pedro e Coimbra (2013)	57,84	68,11	80,21	-
$\text{Img}+\text{UV}+\text{t}$	94,36	120,33	143,86	150,33
$\text{Img}+\text{AT}+\text{RH}+\text{t}$	123,30	135,24	149,54	154,87
$\text{Img}+\text{t}$	128,21	130,87	148,80	154,82

Fonte: Adaptado de Marquez, Pedro e Coimbra (2013).

Outro estudo que pode ser utilizado para a comparação é o trabalho de Si et al. (2021), neste os resultados foram segmentados nas estações do ano e assim apresentados na Tabela 7.6. Os dados foram coletados da província de Shandong, China no intervalo de Janeiro de 2017 e novembro de 2019, sendo que neste estudo o GHI máximo é menor que $1.000 \text{ W}/\text{m}^2$, mais uma vez este trabalho o maior valor de GHI é de $1.200 \text{ W}/\text{m}^2$ o que afeta a comparação do resultado. Ao segmentarmos o estudo em estações do ano, fica evidente os dados apresentados nas seções anteriores, os modelos treinados ficaram mais precisos no outono e no inverno e com desempenho inferior na primavera e verão, respectivamente.

Outro fator que influencia diretamente o resultado se dá ao fato de tratarmos de usinas e estações localizadas em locais muito diferentes, o que pode gerar influências sobre o modelo. Além disso é importante ressaltar que os dados coletados da

Tabela 7.6 – Comparação dos resultados com o estudo de Si et al. (2021). Os valores representam o RMSE dos modelos e expressos em W/m^2 .

Modelo	Estação	Horizonte			
		30min	60min	120min	180min
Si et al. (2021)	Primavera	-	27,2	36,9	54,8
Img+UV+t		102,44	129,59	164,15	117,37
Img+AT+RH+t		132,03	141,71	165,92	170,66
Img+t		133,08	147,84	160,7	169,95
Si et al. (2021)	Verão	-	28,6	41	64,6
Img+UV+t		120,09	161,88	187,85	191,92
Img+AT+RH+t		152,78	169,81	189,06	192,33
Img+t		162,01	154,31	180,56	182,65
Si et al. (2021)	Outono	-	29,6	39,7	58,1
Img+UV+t		75,14	96,85	121,38	119,7
Img+AT+RH+t		97,37	109,94	116,66	120,14
Img+t		108,29	110,97	113,64	118,9
Si et al. (2021)	Inverno	-	25	35,2	53,9
Img+UV+t		74,62	94,31	112,54	119,84
Img+AT+RH+t		106,06	106,27	116,22	126,87
Img+t		106,33	103,79	110,44	119,24

Fonte: Adaptado de Si et al. (2021).

usina Fotovoltaica-UFSC e sua estação meteorológica são para propósitos acadêmicos, podendo haver inconsistências nas bases de dados ocasionadas pela operação, manutenção e modificação dos equipamentos de coleta dos dados utilizados neste trabalho.

8 CONSIDERAÇÕES FINAIS

Este trabalho avaliou diferentes combinações de parâmetros, topologias e entradas de uma ANN para a previsão do GHI em horizontes de 30, 60, 120 e 180 minutos, com foco na integração de imagens de satélite com dados da estação meteorológica para investigar sua influência nos resultados. Com este propósito, foram apresentados cenários de investigação onde foram analisados para cada um dos horizontes a influência das variáveis de maior peso, sendo elas, dia do ano, minutos do dia, imagens de satélite e irradiação UV.

Em nossa análise o Canal 2 do imageador ABI do satélite GOES-16 possui maior relevância para este trabalho devido a sua banda dentro do espectro visível possuir um bom contraste entre as camadas de nuvem e o solo. Além disso, este canal possui maior resolução espacial se comparado aos demais disponibilizados. Importante ressaltar que as informações foram avaliadas para os anos de 2018 e 2019, os anos seguintes tiveram novos produtos habilitados que podem beneficiar trabalhos como este.

Neste estudo identificamos que a maior correlação do GHI com os dados disponíveis é com a irradiação UV. Ressaltamos que este resultado era esperado visto que a irradiação UV é uma parcela do espectro do GHI, desta forma ela apresenta um comportamento similar ao GHI, sendo suscetível as variações climáticas e atenuações ocasionadas pelas variações da atmosfera. As imagens do satélite apresentaram uma correlação fraca em comparação direta com o GHI da estação terrestre, ficou claro no estudo que uma das causas da baixa correlação das imagens de satélite é o deslocamento do sol em relação a usina ao longo do dia, afetando diretamente a correlação deste atributo. Ao fazer uma análise mais próxima, considerando apenas imagens de um determinado mês e de uma determinada hora, observamos que a correlação teve um aumento significativo, reposicionando as imagens de satélite como a segunda componente de maior influência.

Observamos que ao adicionarmos dados temporais como hora do dia e dia do ano, tivemos um ganho significativo do resultado para todos os horizontes de predição. Os modelos propostos se beneficiaram destas componentes para ajustar os parâmetros da rede de forma a torna-la mais sensível as variações de GHI ocasionadas pela movimentação do sol em relação a usina. Este ganho se intensificou com o aumento do horizonte de predição, o que demonstra que para horizontes mais longos os dados temporais tem maior influência. Estes resultados mostram que os dados temporais tendem a influenciar positivamente a predição.

Identificamos que os modelos utilizando apenas de dados climáticos provenientes da estação meteorológica, sendo eles temperatura do ar, umidade relativa, velocidade do vento e dados temporais, tiveram desempenho inferior aos modelos utilizando

apenas dados de satélite e dados temporais, chegando a um ganho de desempenho na ordem de 41,14% no horizonte de 30min e 21,24% no horizonte de 180min quando comparado com o modelos sem imagens de satélite. Com isso observamos que as imagens de satélite melhoram de forma significativa estes modelos em todos os horizontes propostos, porém com o aumento do horizonte de predição as imagens de satélite começam a reduzir sua influência sobre resultado, tendo um desempenho inferior com o aumento do horizonte. Quando combinados os dados meteorológico com os dados de satélite notamos um ganho pouco expressivo quando comparado com os modelos utilizando apenas imagens de satélite e dados temporais. Estes resultados validam nossa proposição inicial e demonstram os ganhos que os modelos de predição de geração de energia podem obter com a inclusão de dados de satélite nos modelos.

Observamos que a componente UV possui uma grande influência sobre o desempenho da rede nas predições com horizonte menor. Quando comparado os modelos de predição utilizando apenas imagens de satélite com os que utilizam apenas a irradiação UV, o segundo tem desempenho 21,16% maior comparado com apenas imagens de satélite e dados temporais no horizonte de 30 minutos. Este resultado é atenuado com o aumento do horizonte, onde os ganhos da adição da componente UV acabam sendo menores. Ao adicionamos as outras componentes da estação terrestre com o UV e as imagens de satélite notamos que estas componentes influenciaram negativamente todos os modelos em todos os horizontes.

Por fim, os resultados encontrados permitem confirmar que as imagens de satélite influenciam positivamente os modelos neurais para a predição de geração de energia em painéis fotovoltaicos, mesmos no caso da adição da componente de irradiação UV que possui uma alta correlação com GHI. A adição das imagens de satélite resultaram na melhora dos modelos principalmente em horizontes mais longos. Vale ressaltar que embora a componente UV, ou o próprio GHI seja o ideal, nem sempre é possível instalar uma estação solarimétrica para análise do potencial energético de uma determinada área, nesta condição os satélites estão acessíveis e fornecem uma alternativa barata para o mapeamento e possuem a capacidade de funcionar tão bem quanto a estação para a predição de GHI.

8.1 TRABALHOS FUTUROS

Para a continuidade deste trabalho é de suma importância a avaliação das novas tecnologias e produtos entregues pelo satélite GOES-16, devido a base de dados da estação terrestre ser limitada aos anos de 2018 e 2019, não foi utilizado de produtos que tiveram suas operações iniciadas no ano de 2019. Como exemplo o ABI-L2-DSRF que entrega uma estimativa da quantidade total de radiação de ondas curtas diretas e difusas que atinge a superfície da terra.

Como uma segunda abordagem para a continuidade, podemos citar a investigação de formas para fazer o pré-processamento das imagens de satélite com o objetivo de corrigir a influência da variação do ângulo zenital sobre o resultado. Esta abordagem foi a inclusão dos dados temporais no modelo, mas fica evidente que o modelo pode se beneficiar substancialmente com a otimização deste parâmetro.

Por fim, sugerimos a utilização de diferentes metodologias para a avaliação das imagens de satélite, neste trabalho foram utilizadas de três imagens como entrada de uma CNN, porém metodologias de predição de movimento de nuvem utilizando campos vetoriais ou algoritmos PIV, podem melhorar o resultado ao substituir ou agregar a metodologia atual.

REFERÊNCIAS

- ABADI, M. et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. 2015. Software available from tensorflow.org. Disponível em: <<https://www.tensorflow.org/>>.
- ABU-MOSTAFA, Y. S.; MAGDON-ISMAIL, M.; LIN, H.-T. *Learning From Data*. [S.l.]: AMLBook, 2012. 213 p. ISBN 9781600490064.
- AHMED, R. et al. A review and evaluation of the state-of-the-art in pv solar power forecasting: Techniques and optimization. *Renewable and Sustainable Energy Reviews*, Elsevier Ltd, v. 124, p. 109792, 2020. ISSN 18790690. Disponível em: <<https://doi.org/10.1016/j.rser.2020.109792>>.
- ALMEIDA, M. P.; PERPIÑÁN, O.; NARVARTE, L. Pv power forecast using a nonparametric pv model. *Solar Energy*, v. 115, p. 354–368, 2015. ISSN 0038092X.
- ALMONACID, F. et al. A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a pv generator. *Energy Conversion and Management*, Elsevier Ltd, v. 85, p. 389–398, 2014. ISSN 01968904. Disponível em: <<http://dx.doi.org/10.1016/j.enconman.2014.05.090>>.
- ALBAIDI, M. H. et al. Mapping of the solar irradiance in the uae using advanced artificial neural network ensemble. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, v. 7, p. 3668–3680, 2014. ISSN 21511535.
- ANTONANZAS, J. et al. Review of photovoltaic power forecasting. *Solar Energy*, Elsevier Ltd, v. 136, p. 78–111, 2016. ISSN 0038092X. Disponível em: <<http://dx.doi.org/10.1016/j.solener.2016.06.069>>.
- ATMOSPHERIC, L. for; PHYSICS, S. *Instruments: Extreme Ultraviolet and X-ray Irradiance Sensors (EXIS)*. 2011. Disponível em: <<https://www.goes-r.gov/spacesegment/exis.html>>.
- BACHMEIER, S.; SCHMIT, T.; GERTH, J. *Quick Guide ABI Band 10*. 2017.
- BACHMEIER, S.; SCHMIT, T.; GERTH, J. *Quick Guide ABI Band 13*. 2017.
- BACHMEIER, S.; SCHMIT, T.; GERTH, J. *Quick Guide ABI Band 8*. 2017.
- BACHMEIER, S.; SCHMIT, T.; GERTH, J. *Quick Guide ABI Band 9*. 2017. 2 p. Disponível em: <http://cimss.ssec.wisc.edu/goes/OCLOFactSheetPDFs/ABIQuickGuide_Band04.pdf>.
- CENTER, L. M. A. T. *Instruments: Geostationary Lightning Mapper (Glm)*. 2011. Disponível em: <<https://www.goes-r.gov/spacesegment/glm.html>>.
- CENTER, L. M. A. T. *Instruments: Magnetometer (MAG)*. 2011. Disponível em: <<https://www.goes-r.gov/spacesegment/mag.html>>.
- CENTER, L. M. A. T. *Instruments: Solar Ultraviolet Imager (SUVI)*. 2011. Disponível em: <<https://www.goes-r.gov/spacesegment/suvi.html>>.
- CHOLLET, F. et al. *Keras*. 2015.

CHOW, C. W. et al. Intra-hour forecasting with a total sky imager at the uc san diego solar energy testbed. *Solar Energy*, v. 85, p. 2881–2893, 11 2011. ISSN 0038092X.

CHU, Y.; PEDRO, H. T.; COIMBRA, C. F. Hybrid intra-hour dni forecasts with sky image processing enhanced by stochastic learning. *Solar Energy*, Elsevier Ltd, v. 98, p. 592–603, 2013. ISSN 0038092X. Disponível em: <<http://dx.doi.org/10.1016/j.solener.2013.10.020>>.

CHUVIECO, E.; HUETE, A. *Fundamentals of satellite remote sensing*. [S.l.: s.n.], 2009. 1-419 p. ISBN 9781420021516.

CORPORATION, A. T. *Instruments: Space Environment In-Situ Suite (SEISS)*. 2011. Disponível em: <<https://www.goes-r.gov/spacesegment/seiss.html>>.

CORPORATION, L. M. S. S. *GOES-R Series Spacecraft Overview*. 2011. Disponível em: <<https://www.goes-r.gov/spacesegment/spacecraft.html>>.

COZZI, L. I. E. A.; GOULD, T. I. E. A. World energy outlook 2021. p. 1–386, 2021. Disponível em: <www.iea.org/weo>.

DAS, U. K. et al. Forecasting of photovoltaic power generation and model optimization: A review. *Renewable and Sustainable Energy Reviews*, Elsevier Ltd, v. 81, p. 912–928, 2018. ISSN 18790690. Disponível em: <<http://dx.doi.org/10.1016/j.rser.2017.08.017>>.

DONG, Z. et al. Satellite image analysis and a hybrid esss/ann model to forecast solar irradiance in the tropics. *Energy Conversion and Management*, Elsevier Ltd, v. 79, p. 66–73, 2014. ISSN 01968904. Disponível em: <<http://dx.doi.org/10.1016/j.enconman.2013.11.043>>.

EISSA, Y. et al. Artificial neural network based model for retrieval of the direct normal, diffuse horizontal and global horizontal irradiances using seviri images. *Solar Energy*, v. 89, p. 1–16, 2013. ISSN 0038092X. Disponível em: <<http://dx.doi.org/10.1016/j.solener.2012.12.008>>.

EMERY, W.; CAMPS, A. *The History of Satellite Remote Sensing*. [S.l.: s.n.], 2017. 1-42 p. ISBN 9780128092545.

GOODFELLOW, I. J.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. MIT Press, 2016. 1-10 p. ISBN 0262035618. Disponível em: <www.deeplearningbook.org>.

GOODMAN, S. J. et al. *GOES: Past, present, and future*. Elsevier, 2017. v. 1-9. 119-149 p. ISBN 9780128032206. Disponível em: <<http://dx.doi.org/10.1016/B978-0-12-409548-9.10315-X>>.

GOOGLE. *Google Maps*. 2022. Disponível em: <<https://www.google.com/maps/place/FOTOVOLTAICA-UFSC+-+Laboratório+de+Sistemas+Solares/@-27.3899152,-48.419887,78000m/data=!3m1!1e3!4m5!3m4!1s0x952743b47e0377c9:0x2ab8cee732b92683!8m2!3d-27.430978!4d-48.4415909>>.

HARRIS, L. *Instruments: Advanced Baseline Imager (ABI)*. 2011. Disponível em: <<https://www.goes-r.gov/spacesegment/abi.html>>.

HAYKIN, S. S. *Neural Networks and Learning Machines*. Third. [S.l.]: Pearson Education, 2009. v. 3. 906 p. ISSN 14337851. ISBN 9780131471399.

HEIDINGER, A. *Algorithm Theoretical Basis Document - ABI Cloud Height*. 2012.

HEIDINGER, A.; STRAKA, W. C. *Algorithm Theoretical Basis Document - ABI Cloud Mask*. 2012.

IEA. Renewables 2021. *International Energy Agency (IEA) Publications International.*, p. 167, 2021. Disponível em: <www.iea.org/t&c/%0Ahttps://webstore.iea.org/download/direct/4329>.

IMPRAM, S.; NESE, S. V.; ORAL, B. Challenges of renewable energy penetration on power system flexibility: A survey. *Energy Strategy Reviews*, Elsevier Ltd, v. 31, p. 100539, 2020. ISSN 2211467X. Disponível em: <<https://doi.org/10.1016/j.esr.2020.100539>>.

JANG, H. S. et al. Solar power prediction based on satellite images and support vector machine. *IEEE Transactions on Sustainable Energy*, v. 7, p. 1255–1263, 2016. ISSN 19493029.

KISHTAWAL, C. Meteorological satellites. *Satellite Remote Sensing and GIS Applications in Agricultural Meteorology*, v. 25, p. 67–79, 2004.

KROMER, P. et al. Support vector regression of multiple predictive models of downward short-wave radiation. *Proceedings of the International Joint Conference on Neural Networks*, p. 651–657, 2014.

KUMAR, D. S. et al. Solar irradiance resource and forecasting: a comprehensive review. *IET Renewable Power Generation*, v. 14, p. 1641–1656, 2020. ISSN 17521424.

LAW, E. W. et al. Direct normal irradiance forecasting and its application to concentrated solar thermal output forecasting - a review. *Solar Energy*, Elsevier Ltd, v. 108, p. 287–307, 2014. ISSN 0038092X. Disponível em: <<http://dx.doi.org/10.1016/j.solener.2014.07.008>>.

LI, Y. Z.; LUAN, R.; NIU, J. C. Forecast of power generation for grid-connected photovoltaic system based on grey model and markov chain. *2008 3rd IEEE Conference on Industrial Electronics and Applications, ICIEA 2008*, p. 1729–1733, 2008.

LINDSTROM, S.; SCHMIT, T.; GERTH, J. *Quick Guide ABI Band 1*. 2017.

LINDSTROM, S.; SCHMIT, T.; GERTH, J. *Quick Guide ABI Band 11*. 2017.

LINDSTROM, S.; SCHMIT, T.; GERTH, J. *Quick Guide ABI Band 12*. 2017.

LINDSTROM, S.; SCHMIT, T.; GERTH, J. *Quick Guide ABI Band 14*. 2017. Disponível em: <http://cimss.ssec.wisc.edu/goes/OCLOFactSheetPDFs/ABIQuickGuide_Band14.pdf>.

LINDSTROM, S.; SCHMIT, T.; GERTH, J. *Quick Guide ABI Band 15*. 2017. Disponível em: <http://cimss.ssec.wisc.edu/goes/OCLOFactSheetPDFs/ABIQuickGuide_Band15.pdf>.

LINDSTROM, S.; SCHMIT, T.; GERTH, J. *Quick Guide ABI Band 16*. 2017.

LINDSTROM, S.; SCHMIT, T.; GERTH, J. *Quick Guide ABI Band 2*. 2017. Disponível em: <http://cimss.ssec.wisc.edu/goes/OCLOFactSheetPDFs/ABIQuickGuide_Band04.pdf>.

LINDSTROM, S.; SCHMIT, T.; GERTH, J. *Quick Guide ABI Band 3*. 2017. Disponível em: <http://cimss.ssec.wisc.edu/goes/OCLOFactSheetPDFs/ABIQuickGuide_Band04.pdf>.

LINDSTROM, S.; SCHMIT, T.; GERTH, J. *Quick Guide ABI Band 4*. 2017.

LINDSTROM, S.; SCHMIT, T.; GERTH, J. *Quick Guide ABI Band 5*. 2017. Disponível em: <http://cimss.ssec.wisc.edu/goes/OCLOFactSheetPDFs/ABIQuickGuide_Band13.pdf>.

LINDSTROM, S.; SCHMIT, T.; GERTH, J. *Quick Guide ABI Band 6*. 2017. 2 p. Disponível em: <http://cimss.ssec.wisc.edu/goes/OCLOFactSheetPDFs/ABIQuickGuide_Band04.pdf>.

LINDSTROM, S.; SCHMIT, T.; GERTH, J. *Quick Guide ABI Band 7*. 2017. Disponível em: <http://cimss.ssec.wisc.edu/goes/OCLOFactSheetPDFs/ABIQuickGuide_Band04.pdf>.

LIU, L. et al. Prediction of short-term pv power output and uncertainty analysis. *Applied Energy*, Elsevier, v. 228, p. 700–711, 2018. ISSN 03062619. Disponível em: <<https://doi.org/10.1016/j.apenergy.2018.06.112>>.

LONG, H.; ZHANG, Z.; SU, Y. Analysis of daily solar power prediction with data-driven approaches. *Applied Energy*, Elsevier Ltd, v. 126, p. 29–37, 2014. ISSN 03062619. Disponível em: <<http://dx.doi.org/10.1016/j.apenergy.2014.03.084>>.

MARQUEZ, R.; PEDRO, H. T.; COIMBRA, C. F. Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to anns. *Solar Energy*, Elsevier Ltd, v. 92, p. 176–188, 2013. ISSN 0038092X. Disponível em: <<http://dx.doi.org/10.1016/j.solener.2013.02.023>>.

MILLER, S. D. et al. Short-term solar irradiance forecasting via satellite/model coupling. *Solar Energy*, Elsevier, v. 168, p. 102–117, 2018. ISSN 0038092X. Disponível em: <<http://dx.doi.org/10.1016/j.solener.2017.11.049>>.

NASA. *NASA Worldview Snapshots*. 2022. Disponível em: <<https://go.nasa.gov/3GSskNn>>.

NEDIS. *Peeling Back the Layers of the Atmosphere*. 2016. Disponível em: <<https://www.nesdis.noaa.gov/news/peeling-back-the-layers-of-the-atmosphere>>.

NOAA. *First complete view of the World's Weather*. 2018. Disponível em: <<https://www.photolib.noaa.gov/Collections/NOAA-In-Space/Imagery/Visible-Imagery/module/707/eitem/23010>>.

NOAA. *GOES Image Viwer*. 2021. Disponível em: <<https://www.star.nesdis.noaa.gov/GOES/conus.php?sat=G16>>.

NOAA. *GOES-R Advanced Baseline Imager Fact Sheet*. 2021. Disponível em: <https://www.goes-r.gov/education/docs/Factsheet_ABI.pdf>.

NOAA; NASA. *Advanced Baseline Imager Scan Mode Information*. 2011. Disponível em: <<https://www.goes-r.gov/users/abiScanModeInfo.html>>.

NOAA; NASA. *GOES-R Series Data Book*. 2019. 240 p. Disponível em: <<https://www.goes-r.gov/downloads/resources/documents/GOES-RSeriesDataBook.pdf>>.

NOAA; NASA. *Beginners Guide to GOES-R Series Data*. 2021.

NOAA; NASA. *Mission Overview*. 2021. Disponível em: <<https://www.goes-r.gov/mission/mission.html>>.

NOAASIS. *GOES Overview*. 2019. Disponível em: <https://www.noaasis.noaa.gov/GOES/goes_overview.html>.

OSPO. *GOES Operational Status*. 2022. Disponível em: <<https://www.ospo.noaa.gov/Operations/GOES/status.html>>.

PAOLI, C. et al. Forecasting of preprocessed daily solar radiation time series using neural networks. *Solar Energy*, Elsevier Ltd, v. 84, p. 2146–2160, 2010. ISSN 0038092X. Disponível em: <<http://dx.doi.org/10.1016/j.solener.2010.08.011>>.

PELISSON, A. et al. Comparative study of photovoltaic power forecasting methods. p. 555–566, 2020.

PENG, Z. et al. Solar irradiance forecast system based on geostationary satellite. *2013 IEEE International Conference on Smart Grid Communications, SmartGridComm 2013*, p. 708–713, 2013.

RAZA, M. Q.; NADARAJAH, M.; EKANAYAKE, C. On recent advances in pv output power forecast. *Solar Energy*, Elsevier Ltd, v. 136, p. 125–144, 2016. ISSN 0038092X. Disponível em: <<http://dx.doi.org/10.1016/j.solener.2016.06.073>>.

RODRÍGUEZ-BENÍTEZ, F. J. et al. A short-term solar radiation forecasting system for the iberian peninsula. part 1: Models description and performance assessment. *Solar Energy*, Elsevier, v. 195, p. 396–412, 2020. ISSN 0038092X. Disponível em: <<https://doi.org/10.1016/j.solener.2019.11.028>>.

SI, Z. et al. Hybrid solar forecasting method using satellite visible images and modified convolutional neural networks. *IEEE Transactions on Industry Applications*, v. 57, p. 5–16, 2021. ISSN 19399367.

SOBRI, S.; KOOHI-KAMALI, S.; RAHIM, N. A. Solar photovoltaic generation forecasting methods: A review. *Energy Conversion and Management*, Elsevier, v. 156, p. 459–497, 2018. ISSN 01968904. Disponível em: <<https://doi.org/10.1016/j.enconman.2017.11.019>>.

SRIVASTAVA, S.; LESSMANN, S. A comparative study of lstm neural networks in forecasting day-ahead global horizontal irradiance with satellite data. *Solar Energy*, Elsevier, v. 162, p. 232–247, 2018. ISSN 0038092X. Disponível em: <<https://doi.org/10.1016/j.solener.2018.01.005>>.

STAR. *STAR Scientists Produce First Light Image From GOES-16 Advanced Baseline Imager*. 2020. Disponível em: <www.star.nesdis.noaa.gov/star/news2017_201702_GOESFirstLight.php>.

- STAR, N. N. *GOES ABI (Advanced Baseline Imager) Realtime Imagery*. 2017. Disponível em: <<http://cimss.ssec.wisc.edu/goes/goesdata.html>>.
- TAN, P.-N. et al. *INTRODUCTION TO DATA MINING*. 2. ed. [S.I.]: Pearson Education Limited, 2019. ISBN 0-273-76922-7.
- URQUHART, B. et al. *Sky-Imaging Systems for Short-Term Forecasting*. Elsevier, 2013. 195-232 p. ISBN 9780123971777. Disponível em: <<http://dx.doi.org/10.1016/B978-0-12-397177-7.00009-7>>.
- Vidraria de Laboratório. *Camadas da Atmosfera*. 2015. Disponível em: <<http://www.vidrariadelaboratorio.com.br/como-funciona-o-telescopio-espacial-hubble/camadas-da-atmosfera/>>.
- VOYANT, C. et al. Machine learning methods for solar radiation forecasting: A review. *Renewable Energy*, Elsevier Ltd, v. 105, p. 569–582, 2017. ISSN 18790682. Disponível em: <<http://dx.doi.org/10.1016/j.renene.2016.12.095>>.
- WALTHER, A.; STRAKA, W.; HEIDINGER, A. K. *ABI Algorithm Theoretical Basis Document For Daytime Cloud Optical and Microphysical Properties (DCOMP)*. 2013. Disponível em: <<https://www.goes-r.gov/products/baseline-cloud-opt-depth.html>>.
- WILLIAMS, B. *The Earths Atmosphere as a Filter*. 2023. Disponível em: <<https://www.briangwilliams.us/solar-hydrogen-generation/the-earths-atmosphere-as-a-filter.html>>.
- WITTEN, I. H.; FRANK, E. *Data Mining: Practical Machine Learning Tools and Techniques*. 2. ed. [S.I.]: Elsevier, 2005. ISBN 0120884070.
- YANG, C.; XIE, L. A novel arx-based multi-scale spatio-temporal solar power forecast model. *2012 North American Power Symposium, NAPS 2012*, 2012.
- YANG, D. et al. Very short term irradiance forecasting using the lasso. *Solar Energy*, Elsevier Ltd, v. 114, p. 314–326, 2015. ISSN 0038092X. Disponível em: <<http://dx.doi.org/10.1016/j.solener.2015.01.016>>.
- YU, L.; WANG, S.; LAI, K. An integrated data preparation scheme for neural network data analysis. *IEEE Transactions on Knowledge and Data Engineering*, v. 18, p. 217–230, 2006. ISSN 10414347.
- ZHANG, J. et al. A suite of metrics for assessing the performance of solar power forecasting. *Solar Energy*, Elsevier Ltd, v. 111, p. 157–175, 2015. ISSN 0038092X. Disponível em: <<http://dx.doi.org/10.1016/j.solener.2014.10.016>>.

APÊNDICE A – SATÉLITES GOES

A.1 EVOLUÇÃO DOS SATÉLITES GOES

O projeto GOES teve início em 1975, um esforço conjunto da NOAA com a NASA. Em 16 de outubro de 1975 o primeiro satélite do programa GOES foi lançado, o GOES-A, e foi renomeado ao atingir a órbita para GOES-1, procedimento aplicado a todos os satélites da série. Os satélites GOES-1 a GOES-3 eram idênticos ao *Synchronous Meteorological Satellite* (SMS), contando com instrumentos *Visible and Infrared Spin Scan Radiometer* (VISSR), *Space Environment Monitor* (SEM) e *Data Collection System* (DCS).

O GOES-4 a GOES-7 foi incorporado o *Visible and Infrared Atmospheric Sounder* (VAS). O VISSR possuía capacidade de prover o espectro visível e três bandas IR, já o VAS fornecia uma banda visível e doze bandas IR, o que permitia monitorar fluxo atmosférico de nível médio, umidade, incêndios e neblina. O VAS foi a primeira sonda geoestacionária cujos dados puderam ser usados para determinar a temperatura atmosférica e os perfis de umidade. As bandas espectrais de IR eram sensíveis ao CO_2 atmosférico, N_2O , H_2O e cada uma detectava camadas diferentes da atmosfera (GOODMAN et al., 2017).

Em 3 de maio de 1986, o GOES-G, projetado para substituir o GOES-4, seria nomeado de GOES-7 foi perdido quando seu veículo foi atingido por um raio após a decolagem. O GOES-H, que se tornou o GOES-7 foi lançado em 28 de abril de 1987, sendo o primeiro satélite GOES capaz de detectar sinais de socorro de 40MHz emitidos por aeronaves e navios e enviá-los para estação terrestre.

O GOES-8, lançado em 1994 introduziu uma nova geração de satélites. Várias melhorias foram introduzidas permitindo com que o satélite capturasse imagens de alta resolução, sendo a principal a modificação do imageador. Em 2001 foi lançado o GOES-12 o primeiro satélite a carregar um imageador de Raio-X direcionado para o sol.

O GOES-R, nomeado quando atingiu órbita de GOES-16 foi lançado em 2016 e introduziu o novo imageador ABI, melhoria que permite o mapeamento completo a cada 15 minutos e fornece 16 bandas espectrais. Além da adição de vários outros sensores que permitem a análise de descargas atmosféricas e campos eletromagnéticos. Nas Tabelas A.1 e A.2 é apresentada a evolução dos imageadores com suas principais características.

A.2 DETALHAMENTO DOS CANAIS DO ABI

Detalhamento das aplicações de cada um dos canais do ABI:

- Canal 01 - *Blue* ($0.47 \mu m$) - É um dos canais visíveis. Fornece dados para o mo-

Tabela A.1 – Evolução dos imageadores geoestacionários ao longo dos anos: visível/IR próximo.

Ano de início	Instrumento	Bandas	Resolução Espacial (km)	Tempo de Captura (min)
1966	ATS-1/2 (Spin-Scan Cloud Camera)	1	4	20
1967	ATS-3 (Multicolor Spin-Scan Cloud Camera)	3	4	24
1974	ATS-6 (Geosynchronous Very High Resolution Radiometer)	1	6	25
1974	SMS-1/2 (Visible Infrared Spin-Scan Radiometer (VISSR))	1	1	20
1975	GOES-1/3 (VISSR)	1	1	30
1980	GOES-4/7 (VISSR+VAS)	1	1	30
1994	Imager-8/12+Sounders	1	1	7-30
2010	Imager-13/15+Sounders	1	1	7-30
2016	ABI	6	0.5-1	0.5-5-15

Fonte: Adaptado de Goodman et al. (2017).

Tabela A.2 – Evolução dos imageadores geoestacionários ao longo dos anos: bandas IR.

Ano de início	Instrumento	Bandas	Resolução Espacial (km)	Tempo de Captura (min)
1966	ATS-1/2 (Spin-Scan Cloud Camera)	N/A	N/A	N/A
1967	ATS-3 (Multicolor Spin-Scan Cloud Camera)	N/A	N/A	N/A
1974	ATS-6 (Geosynchronous Very High Resolution Radiometer)	1	11	25
1974	SMS-1/2 (Visible Infrared Spin-Scan Radiometer (VISSR))	1	8	20
1975	GOES-1/3 (VISSR)	1	8	30
1980	GOES-4/7 (VISSR+VAS)	3	7-14	30
1994	Imager-8/12+Sounders	4	4-8	7-30
2010	Imager-13/15+Sounders	4	4	7-30
2016	ABI	10	2	0.5-5-15

Fonte: Adaptado de Goodman et al. (2017).

nitoramento de aerossóis. Utilizada para observações diurnas de poeira, neblina, fumaça e nuvens (LINDSTROM; SCHMIT; GERTH, 2017a);

- Canal 02 - *Red* ($0.64 \mu m$) - É o canal com maior resolução espacial 0,5km. Ideal para identificar nevoeiros, cobertura de neve e gelo, identificação de cinzas vulcânicas e análise de furacões (LINDSTROM; SCHMIT; GERTH, 2017g);
- Canal 03 - *Veggie* ($0.86 \mu m$) - Detecta nuvens diurnas. Neblina e aerossóis. Utilizada para calcular o índice de vegetação de diferença normalizada (NDVI). Utilizada para identificação de pontos que ocorreram queimadas e determinar potencial de incêndio. Gera grande contraste entre terra e água (LINDSTROM; SCHMIT; GERTH, 2017h);

- Canal 04 - *Cirrus* (1.37 μm) - Utilizada para detectar nuvens altas durante o dia. Cirros muito finos podem ser detectados. Atua na região de absorção do vapor d'água (LINDSTROM; SCHMIT; GERTH, 2017i);
- Canal 05 - *Snow/Ice* (1.61 μm) - Detecta nuvens de água líquida. As nuvens de gelo são escuras, pois o gelo absorve a radiação. Contraste entre terra e água é muito bom o que torna fácil a detecção de lagos nesta imagem. Incêndios podem ser detectados a noite neste canal (LINDSTROM; SCHMIT; GERTH, 2017j);
- Canal 06 - *Cloud particles size* (2.25 μm) - Em conjunto com outros canais permite estimar o tamanho das partículas de nuvem. Utilizada com outros canais para estimar o tamanho das partículas de aerossol, criar máscara de nuvens e detecção de pontos quentes (LINDSTROM; SCHMIT; GERTH, 2017k);
- Canal 07 - *Shortwave window* (3.9 μm) - Utilizada para identificar neblina e nuvens baixas a noite. Identificar pontos quentes de incêndios, cinzas vulcânicas e estimar a temperatura da superfície do mar. Os vetores de vento de nível baixo podem ser estimados. Detecta radiação terrestre emitida e radiação solar refletida durante o dia (LINDSTROM; SCHMIT; GERTH, 2017l);
- Canal 08 - *Upper-level water vapor* (6.19 μm) - É um dos canais de vapor de água. É utilizada para rastrear ventos troposféricos superiores, correntes de jatos, prever a trilha de furacões e movimento de tempestades médias. Identifica regiões onde existe potencial turbulência (BACHMEIER; SCHMIT; GERTH, 2017c);
- Canal 09 - *Midlevel water vapor* (6.9 μm) - É um dos canais de vapor de água. Usada para rastreamento de ventos troposféricos, identificar correntes jato, prever trilha de furacões e tempestades de latitude média. As camadas mostram resfriamento devido a absorção de energia a 6.9 μm pelo vapor de água (BACHMEIER; SCHMIT; GERTH, 2017d);
- Canal 10 - *Lower/midlevel water vapor* (7.3 μm) - Um dos canais de detecção de vapor de água. É usado para rastrear ventos troposféricos inferiores, identificar rajadas de jatos, monitorar o potencial de clima severo, estimar umidade de nível inferior, destacar plumas vulcânicas ricas em dióxido de enxofre (SO₂) (BACHMEIER; SCHMIT; GERTH, 2017a);
- Canal 11 - *Cloud-top phase* (8.5 μm) - Monitorar atividade vulcânica. É chamada janela suja, há mais absorção de vapor de água neste canal do que na janela limpa 10.3 μm (LINDSTROM; SCHMIT; GERTH, 2017b);
- Canal 12 - *Ozone* (9.6 μm) - Fornece informações diurnas e noturnas sobre a dinâmica da atmosfera próxima à tropopausa. Este canal mostra temperaturas

mais frias do que o canal da janela limpa, porque o ozônio e o vapor de água absorvem $9,6\mu m$ de energia atmosférica (LINDSTROM; SCHMIT; GERTH, 2017c);

- Canal 13 - *Clean longwave window* ($10.3\ \mu m$) - É o canal de janela IR menos sensível ao vapor de água. Melhora a identificação / classificação de outras características atmosféricas e na nuvem. Estimativa de temperatura no topo da nuvem (BACHMEIER; SCHMIT; GERTH, 2017b);
- Canal 14 - *Longwave window* ($11.2\ \mu m$) - Há absorção de energia pelo vapor d'água neste canal. Usado em produtos de base como detecção de cinzas vulcânicas, vetores de vento de movimento derivado, perfis atmosféricos legados, incluindo água precipitável, propriedades do topo da nuvem, detecção de aerossol e temperatura da superfície do solo (LINDSTROM; SCHMIT; GERTH, 2017d);
- Canal 15 - *Dirty longwave window* ($12.3\ \mu m$) - Usado em produtos de base como detecção de cinzas vulcânicas, vetores de vento de movimento derivado, perfis atmosféricos legados, incluindo água precipitável, propriedades do topo da nuvem, detecção de aerossol e temperatura da superfície do solo (LINDSTROM; SCHMIT; GERTH, 2017e);
- Canal 16 - CO_2 ($13.3\ \mu m$) - É usada em muitos produtos de linha de base GOES-16. Incluem máscara de nuvem, altura do topo da nuvem, pressão e temperatura. Este canal é uma entrada para os perfis legados de umidade e temperatura. Apesar de sua importância nos produtos, o canal de CO_2 normalmente não é usado para interpretação visual de eventos climáticos (LINDSTROM; SCHMIT; GERTH, 2017f);

A.3 PRINCIPAIS PRODUTOS ENTREGUES PELO GOES-16

Tabela A.3 – Produtos entregues pelo satélite GOES-16 no repositório AWS.

Produto	Instr.	Nível	Descrição
ABI-L1b-RadF	ABI	L1b	<i>Advanced Baseline Imager Level 1b Full Disk</i>
ABI-L2-ACHAF	ABI	L2	<i>Advanced Baseline Imager Level 2 Cloud Top Height Full Disk</i>
ABI-L2-ACHTF	ABI	L2	<i>Advanced Baseline Imager Level 2 Cloud Top Temperature Full Disk</i>
ABI-L2-ACMF	ABI	L2	<i>Advanced Baseline Imager Level 2 Clear Sky Mask Full Disk</i>
ABI-L2-ACTPF	ABI	L2	<i>Advanced Baseline Imager Level 2 Cloud Top Phase Full Disk</i>
ABI-L2-ADPF	ABI	L2	<i>Advanced Baseline Imager Level 2 Aerosol Detection Full Disk</i>
ABI-L2-AODF	ABI	L2	<i>Advanced Baseline Imager Level 2 Aerosol Optical Depth Full Disk</i>
ABI-L2-CMIPF	ABI	L2	<i>Advanced Baseline Imager Level 2 Cloud and Moisture Imagery Full Disk</i>
ABI-L2-CODF	ABI	L2	<i>Advanced Baseline Imager Level 2 Cloud Optical Depth Full Disk</i>
ABI-L2-CPSF	ABI	L2	<i>Advanced Baseline Imager Level 2 Cloud Particle Size Full Disk</i>
ABI-L2-CTPF	ABI	L2	<i>Advanced Baseline Imager Level 2 Cloud Top Pressure Full Disk</i>
ABI-L2-DMWF	ABI	L2	<i>Advanced Baseline Imager Level 2 Derived Motion Winds Full Disk</i>
ABI-L2-DSIF	ABI	L2	<i>Advanced Baseline Imager Level 2 Derived Stability Indices Full Disk</i>
ABI-L2-DSRF	ABI	L2	<i>Advanced Baseline Imager Level 2 Downward Shortwave Radiation Full Disk</i>
ABI-L2-FDCF	ABI	L2	<i>Advanced Baseline Imager Level 2 Fire (Hot Spot Characterization) Full Disk</i>
ABI-L2-LSTF	ABI	L2	<i>Advanced Baseline Imager Level 2 Land Surface Temperature Full Disk</i>
ABI-L2-LVMPF	ABI	L2	<i>Advanced Baseline Imager Level 2 Legacy Vertical Moisture Profile Full Disk</i>
ABI-L2-LVTPF	ABI	L2	<i>Advanced Baseline Imager Level 2 Legacy Vertical Temperature Profile Full Disk</i>
ABI-L2-MCMIPF	ABI	L2	<i>Advanced Baseline Imager Level 2 Cloud and Moisture Imagery Full Disk</i>
ABI-L2-RRQPEF	ABI	L2	<i>Advanced Baseline Imager Level 2 Rainfall Rate (Quantitative Precipitation Estimate) Full Disk</i>
ABI-L2-RSRF	ABI	L2	<i>Advanced Baseline Imager Level 2 Reflected Shortwave Radiation Top-Of-Atmosphere Full Disk</i>
ABI-L2-SSTF	ABI	L2	<i>Advanced Baseline Imager Level 2 Sea Surface (Skin) Temperature Full Disk</i>
ABI-L2-TPWF	ABI	L2	<i>Advanced Baseline Imager Level 2 Total Precipitable Water Full Disk</i>
ABI-L2-VAAF	ABI	L2	<i>Advanced Baseline Imager Level 2 Volcanic Ash: Detection and Height Full Disk</i>
GLM-L2-LCFA	GLM	L2	<i>Geostationary Lightning Mapper Level 2 Lightning Detection</i>
SUVI-L1b-Fe093	SUVI	L1b	<i>Solar Ultraviolet Imager Level 1b Extreme Ultraviolet</i>
SUVI-L1b-Fe131	SUVI	L1b	<i>Solar Ultraviolet Imager Level 1b Extreme Ultraviolet</i>
SUVI-L1b-Fe171	SUVI	L1b	<i>Solar Ultraviolet Imager Level 1b Extreme Ultraviolet</i>
SUVI-L1b-Fe195	SUVI	L1b	<i>Solar Ultraviolet Imager Level 1b Extreme Ultraviolet</i>
SUVI-L1b-Fe284	SUVI	L1b	<i>Solar Ultraviolet Imager Level 1b Extreme Ultraviolet</i>
SUVI-L1b-He303	SUVI	L1b	<i>Solar Ultraviolet Imager Level 1b Extreme Ultraviolet</i>

Fonte: Adaptado de NOAA e NASA (2019).

APÊNDICE B – ESTUDO PRELIMINAR

Para a construção da proposta de desenvolvimento deste trabalho foram realizados estudos preliminares para validar as hipóteses iniciais. Sendo elas: temos acesso e conseguimos utilizar as bases de dados propostas? Os dados de satélite e os dados das estações terrestres possuem relação? Com estes dados preparados a rede neural será capaz de encontrar relação e entregar um resultado minimamente satisfatório?

Estas questões foram respondidas no estudo preliminar e são apresentadas nas seções seguintes. Na Seção B.1 são apresentadas as metodologias utilizadas para a obtenção dos dados do satélite GOES-16, assim como, a construção da base de dados em conjunto com os dados da estação terrestre, que são utilizados para a validação da base e o treinamento do primeiro modelo. Na Seção B.2 é apresentada a construção da ANN utilizada para a validação da hipótese. Na Seção B.3 são apresentados os resultados obtidos no estudo preliminar.

B.1 DADOS

B.1.1 Preparação dos dados

Definido o intervalo temporal do estudo sendo das 9h às 15h GMT-3, com granularidade de 30min e horizonte de predição sendo 30min e 60min, a base de dados foi preparada para ser utilizada na ANN. A proposta da rede neural será apresentada na Seção B.2, mas precisa de 4 informações para ser treinada, o GHI esperado em t +horizonte de predição, a imagem do satélite no momento t , a imagem do satélite em $t-30$ min e a imagem do satélite em $t-60$ min, onde t é o instante atual. Para compor esta estrutura, os dados de GHI foram normalizados conforme a Equação (4.1) e os dados das imagens foram vetorizados, formando vetores de 1600 posições cada. A estrutura pode ser vista na Tabela B.1, onde representa a organização da base de dados para horizonte de 30min e de 60min, é importante pontuar que cada horizonte possui a sua ANN e é treinado separadamente.

Tabela B.1 – Representação dos dados de entrada da ANN.

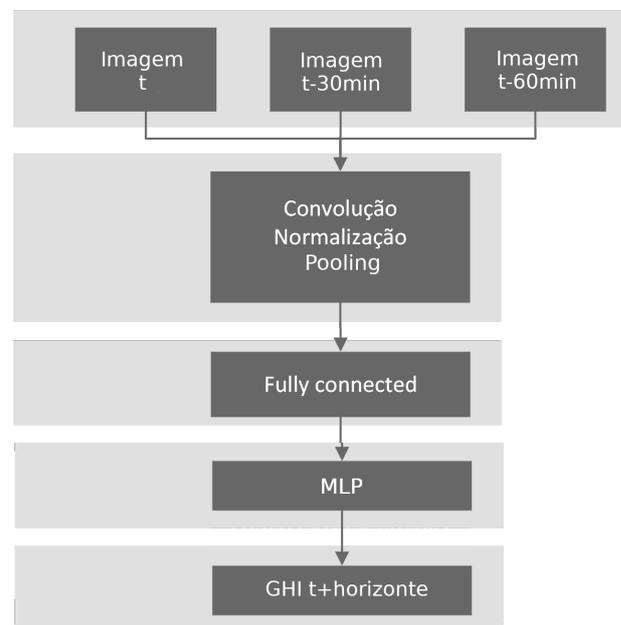
Horizonte	Timestamp Atual Hora do dia	Imagem 1 Entrada	Imagem 2 Entrada	Imagem 3 Entrada	GHI Previsto Saída
30	10/03/2018 12:00	Imagem 12:00	Imagem 11:30	Imagem 11:00	GHI 12:30
	10/03/2018 12:30	Imagem 12:30	Imagem 12:00	Imagem 11:30	GHI 13:00
	10/03/2018 13:00	Imagem 13:00	Imagem 12:30	Imagem 12:00	GHI 13:30
	10/03/2018 13:30	Imagem 13:30	Imagem 13:00	Imagem 12:30	GHI 14:00
60	10/03/2018 12:00	Imagem 12:00	Imagem 11:30	Imagem 11:00	GHI 13:00
	10/03/2018 12:30	Imagem 12:30	Imagem 12:00	Imagem 11:30	GHI 13:30
	10/03/2018 13:00	Imagem 13:00	Imagem 12:30	Imagem 12:00	GHI 14:00
	10/03/2018 13:30	Imagem 13:30	Imagem 13:00	Imagem 12:30	GHI 14:30

Organizada as bases de dados, estas foram divididas em 3 grupos, 70% dos dados de 2018, com balanceamento homogêneo entre os meses, ou seja, 70% de janeiro, 70% de fevereiro, assim por diante, foi separada para treinar a rede. Os demais 30% dos dados foi separado para a validação durante o treino. O terceiro grupo, dados de 2019 foi separado para testes, ou seja, sem passar pela rede durante o treinamento.

B.2 PROPOSTA DE REDE CONVOLUCIONAL

A entrada da rede será as imagens do instante t , $t-30\text{min}$ e $t-60\text{min}$, conforme apresentado na Figura B.1. As imagens serão processadas por camadas convolucionais, cada camada é composta por uma convolução 2D, uma normalização e uma camada de *pooling*. Na primeira camada do conjunto convolucional, é aplicada a convolução com ativação ReLU, seguindo a abordagem de Si et al. (2021) e para terceira camada o *average pooling* 2D, mesmo conceito do MaxPooling mas calculando a média do *kernel*, com passo 2×2 , esta topologia foi adaptada do estudo de Si et al. (2021). A quantidade de camadas convolucionais, a quantidade de filtros convolucionais e o *kernel*, serão variáveis deste estudo e são apresentados à frente.

Figura B.1 – Topologia da ANN para o estudo preliminar.



Com a topologia proposta a rede possui uma entrada de 4800 atributos, 3 imagens de 40×40 , após as camadas de convolução são entregues características das imagens apresentadas. Não adicionamos os dados meteorológicos para o estudo preliminar, são considerados apenas as imagens de satélite, isto para validar o método proposto. Portanto os atributos são aplicados a uma MLP com função de ativação

ReLU, seguindo a abordagem de Si et al. (2021) e a quantidade de camadas e neurônios definidos no estudo, apresentado na sequência. O resultado é uma saída que varia entre 0 e 1, correspondente ao GHI normalizado. O treinamento é feito utilizando a otimização Adam com passo de 0,001, função de perda MSE e como métrica o RMSE. A rede foi treinada até um limite de 1000 épocas, mas utilizando uma função de parada antecipada com 50 épocas de persistência, com tamanho de lote de 512 e os resultados são apresentados na seção seguinte.

Para identificar qual o melhor conjunto de parâmetros e topologia da ANN, foram testadas 3200 combinações diferentes para o horizonte de 30 minutos e 3200 combinações para o horizonte de 60 minutos. Os parâmetros que foram variados são quantidade de camadas de convolução, quantidade de filtros, tamanho do *kernel*, quantidade de camadas ocultas, taxa de *dropout* e quantidade de neurônios, sendo as variações:

- Convoluções: {1, 2, 3, 4}
- Filtros: {16, 32}
- *Kernel*: {2, 3, 4, 5}
- Camadas: {1, 2, 3, 4, 5}
- *Dropout*: {0.0, 0.15, 0.3, 0.45}
- Neurônios: {64, 128, 256, 512, 1024}

Os treinamentos foram conduzidos utilizando a plataforma Colab do Google, utilizando como hardware de vídeo uma Tesla P100 com 16Gb de memória, 27,3 Gb de memória RAM.

B.3 RESULTADOS

Após o treinamento das redes utilizando os dados de 2018, elas foram testadas utilizando os dados de 2019 com a métrica RMSE normalizado ou rRMSE. O melhor resultado para o horizonte de 30 minutos foi um rRMSE de 15,6% e para o horizonte de 60 minutos 17,2% respectivamente. A Tabela B.2 apresenta as configurações dos resultados, podemos notar uma predominância de poucas camadas ocultas. Para as redes de horizonte de 30 minutos, teve uma quantidade de neurônios mais baixa de comparada com o horizonte de 60 minutos, porém de forma geral, os dois resultados tiveram melhor desempenho com menos neurônios. Os demais parâmetros tiveram variações maiores no resultado.

Tabela B.2 – Configuração dos cinco melhores resultados da métrica rRMSE para horizontes de 30 e 60 minutos.

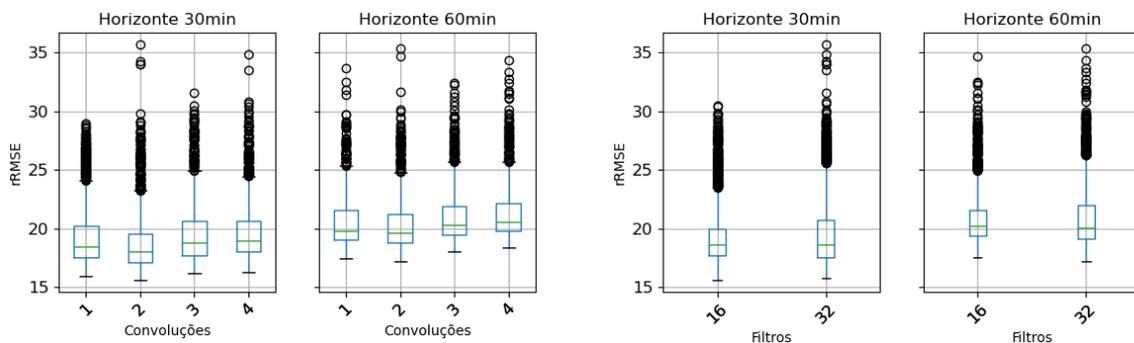
Horizonte	Conv.	Filtros	Kernel	Camadas	Dropout	Neurônios	rRMSE	Modelo
30	2	16	5	1	0,45	64	15,6	M07
	2	32	4	1	0,45	1024	15,8	M08
	2	32	5	1	0,3	64	15,8	M09
	2	16	3	2	0,45	64	15,8	M10
	2	16	3	2	0,3	64	15,8	M11
60	2	32	5	1	0,45	256	17,2	M12
	2	32	3	2	0,15	256	17,2	M13
	1	32	3	5	0,15	256	17,5	M14
	1	16	5	1	0,3	64	17,5	M15
	1	32	5	1	0,45	512	17,5	M16

Para melhor avaliar as diferentes combinações de parâmetros e a distribuição do erro, os dados foram apresentados em gráfico de caixa individual por parâmetro, Figura B.2. Avaliando individualmente cada um dos gráficos, podemos elencar os melhores resultados para determinar uma combinação para a ANN. Avaliando a quantidade de camadas de convolução Figura B.2(a), podemos notar que para os dois horizontes o resultado que possui uma concentração de erro menor e menos disperso, é 2 camadas de convolução. Seguindo a mesma analogia para a quantidade de filtros, Figura B.2(b), podemos selecionar 16 filtros para os dois horizontes. No caso do *kernel*, Figura B.2(c), para os dois casos 3 e 4 possuem uma distribuição muito similar.

Quanto a quantidade de camadas ocultas, Figura B.2(d), a distribuição de 1 camada evidentemente possui um melhor resultado, porém fica claro um aumento da dispersão do erro com o aumento do número de camadas, o que pode sinalizar o aumento da complexidade da rede e uma dificuldade maior de convergência. Já a distribuição da quantidade de neurônios, Figura B.2(e), para o horizonte de 30 minutos os resultados de 256 e 512 neurônios ficam muito próximos, enquanto para o horizonte de 60 minutos o resultado de 512 tem uma concentração dos valores maior e uma taxa de erro dos resultados menor. Por fim, a taxa de *dropout*, responsável por generalizar o modelo aplica uma taxa de perda dos dados para que a rede tenha capacidade de lidar com as variações de entrada, representado o estudo a Figura B.2(f), possui um resultado melhor em 0.15, porém é interessante observar que o aumento da taxa de *dropout* torna a rede mais instável, tornando a dispersão do erro maior, assim como o *dropout* de 0 pode ter um resultado inferior, devido a rede estar se especializando nos dados de treinamento, com pouca margem para se adaptar as variações das entradas que não foram treinadas.

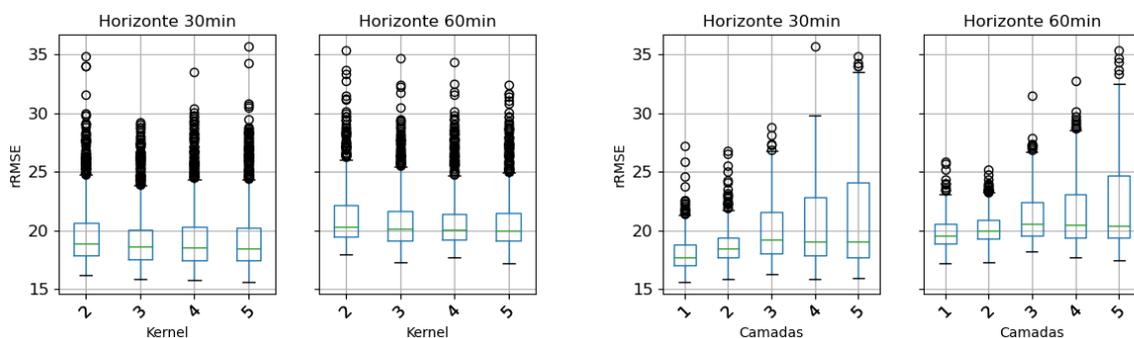
Com base nas observações foi montada a Tabela B.3, que compila as configurações e resultados dos modelos elencados por análise gráfica. É possível notar que para este resultado tivemos um aumento aproximado de 1,2 pontos percentuais no rRMSE para o horizonte de 30 minutos e 2,2 pontos para o horizonte de 60 minutos,

Figura B.2 – rRMSE de um conjunto de 3200 treinos para horizonte de 30min e 3200 treinos para horizonte de 60min, variando os parâmetros do treino.



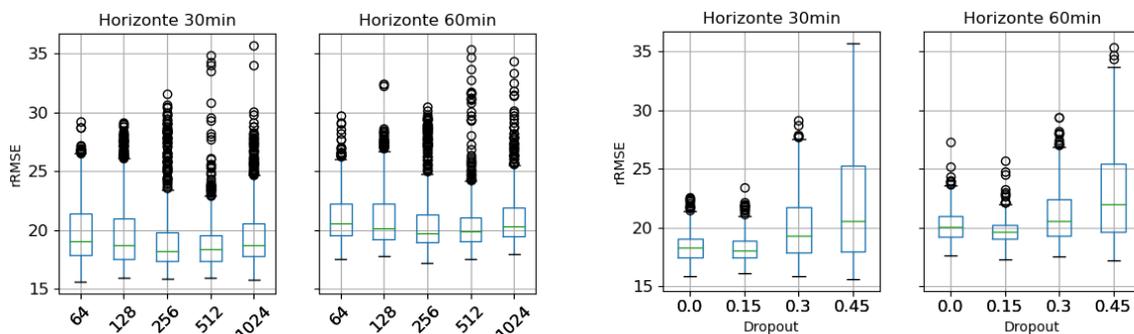
(a) Variação da quantidade de convoluções.

(b) Variação da quantidade de Filtros.



(c) Variação do tamanho do *Kernel*.

(d) Variação do número de camadas.



(e) Variação da quantidade de neurônios nas camadas ocultas.

(f) Variação da taxa de *dropout*.

quando comparado com a Tabela B.2.

Tabela B.3 – Configuração das recomendações baseados no gráfico de caixas para horizontes de 30 e 60 minutos.

Horizonte	Conv.	Filtros	Kernel	Camadas	Dropout	Neurônios	rRMSE	Modelo
30	2	16	3	1	0,15	256	17,1	M01
	2	16	4	1	0,15	256	16,8	M02
	2	16	3	1	0,15	512	17,1	M03
	2	16	4	1	0,15	512	16,8	M04
60	2	16	3	1	0,15	512	19,4	M05
	2	16	4	1	0,15	512	19,7	M06

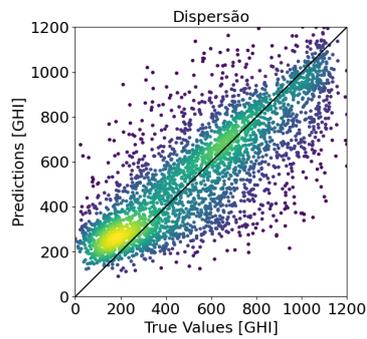
Porém ao compararmos os gráfico de distribuição para o horizonte de 30 minutos dos modelos de melhor desempenho, Figura B.3(a) até Figura B.3(e), com os modelos escolhidos graficamente, Figura B.3(f) até Figura B.3(i), é possível observar que a distribuição dos dados mesmo que com um rRMSE menor, ficam mais agrupados em torno do eixo para as escolhas gráficas, podendo analisar a Figura B.3(d) e Figura B.3(e), que com o GHI próximo de 200 acentuam o erro. Porém a dispersão dos dados é maior no segundo caso. O mesmo efeito pode ser observado para os casos com horizonte de 60 minutos, Figura B.4(a) até Figura B.4(e) e Figura B.4(f) e Figura B.4(g). Porém neste, a distribuição M13, Figura B.4(b), apresenta os dados próximos do eixo e com uma distribuição concentrada.

Ao analisarmos a distribuição do erro para os modelos, conforme Figura B.5, para os dois horizontes, podemos notar que as distribuições são parecidas, e não representam forte influência para o resultado, porém podemos destacar o modelo M08 para o horizonte de 30 minutos e o modelo M12 para o horizonte de 60min tem uma concentração maior de previsões com erro menor. A predição Figura B.6 apresenta a distribuição ao longo do dia é possível notar um erro maior no início e no final do dia, dando um indício da necessidade de correção do ângulo zenital para o melhor desempenho do modelo. Ao isolarmos os modelos trimestralmente, podemos notar que o segundo trimestre, Figura B.8, e terceiro trimestre, Figura B.9, do ano possuíram um desempenho melhor, comparados ao primeiro, Figura B.7, e quarto trimestre, Figura B.10, podendo investigar a possibilidade de segmentar o modelo em estações visando especializar a ANN. Avaliando os modelos em meses distintos e em estações diferentes como maio, Figura B.11, julho, Figura B.12, e novembro Figura B.13, os modelos apresentam resultados similares para os dois horizontes de predição, com destaque para dia 16 de março de 2019 e 27 de julho de 2019 com maiores desvios, e para o intervalo de 15 de novembro de 2019 a 23 de novembro de 2019 onde todos os modelos tiveram não só erro de predição, como tiveram dificuldade de acompanhar a linha de GHI, parecendo que o modelo se perdeu no intervalo. Deixando estes pontos abertos para investigação ao longo da continuidade da pesquisa.

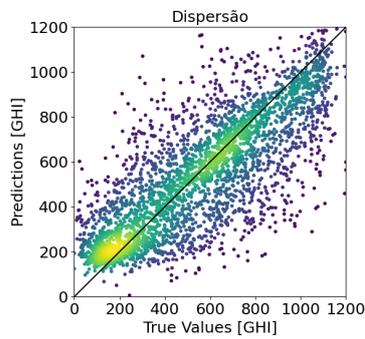
B.4 CONSIDERAÇÕES FINAIS

Os resultados encontrados no estudo preliminar suportam a continuidade da pesquisa. As bases de dados estão disponíveis e são factíveis de utilizar no estudo proposto, de fato os dados foram coletados e manipulados para a condução do estudo preliminar. Com relação a proposta de usar dados de satélite para predição de geração de energia PV, foi apresentado que os dados possuem relação o que possibilita uma otimização do modelo para obter um melhor desempenho, não só baseado no estudo, mas na revisão da literatura, foram encontradas linhas de pesquisa sobre o assunto. E por fim, mesmo com um modelo simplificado foram encontrados resultados promisso-

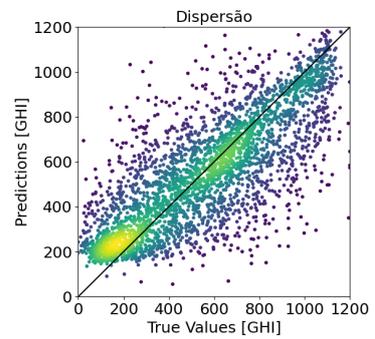
Figura B.3 – Distribuição da predição de GHI pela ANN x GHI da estação terrestre, com horizonte de predição de 30 minutos.



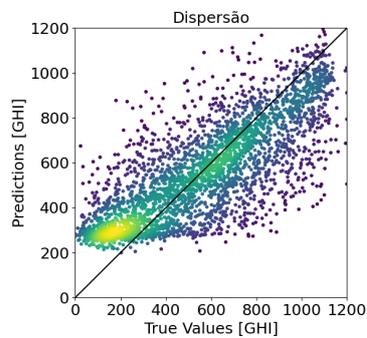
(a) Modelo M07.



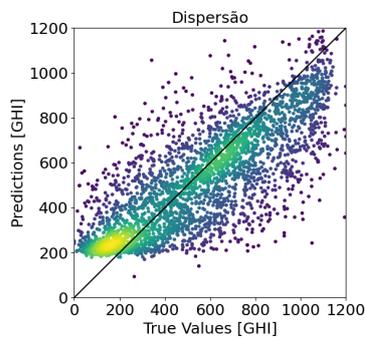
(b) Modelo M08.



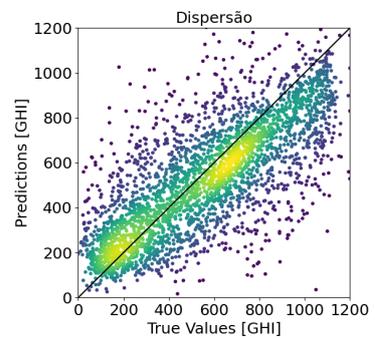
(c) Modelo M09.



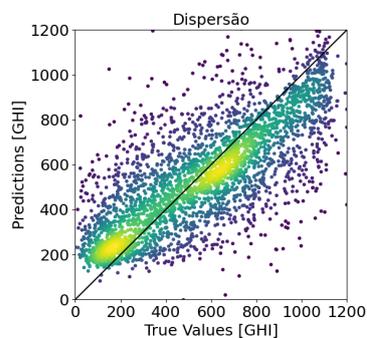
(d) Modelo M10.



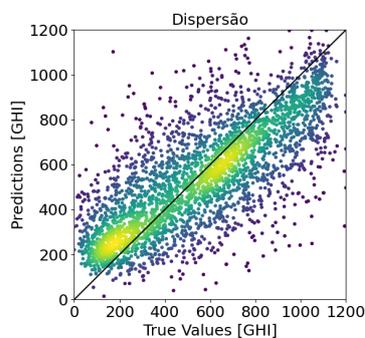
(e) Modelo M11.



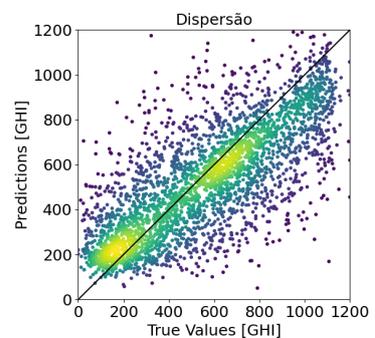
(f) Modelo M01.



(g) Modelo M02.



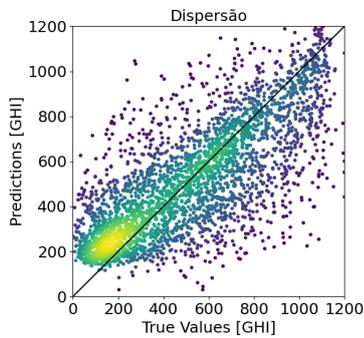
(h) Modelo M03.



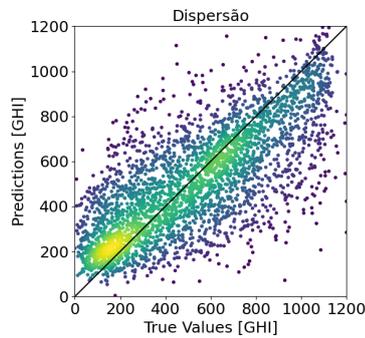
(i) Modelo M04.

res, após o estudo preliminar fica evidente a relação e o potencial de predição com a metodologia proposta.

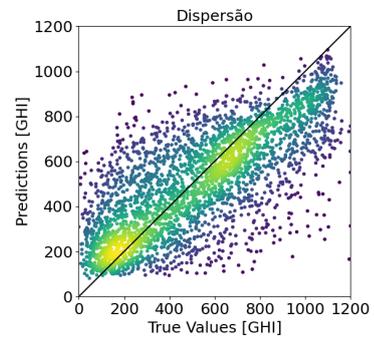
Figura B.4 – Distribuição da predição de GHI pela ANN x GHI da estação terrestre, com horizonte de predição de 60 minutos.



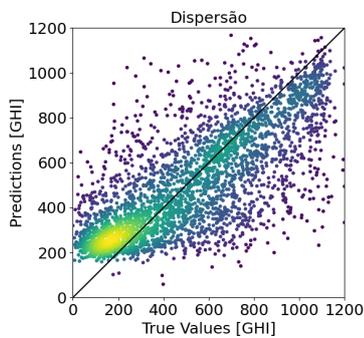
(a) Modelo M12.



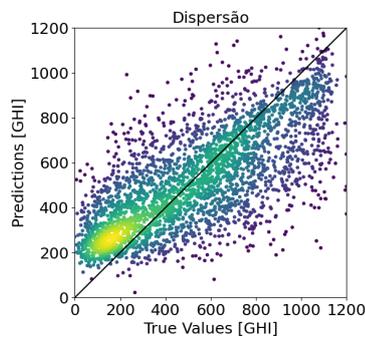
(b) Modelo M13.



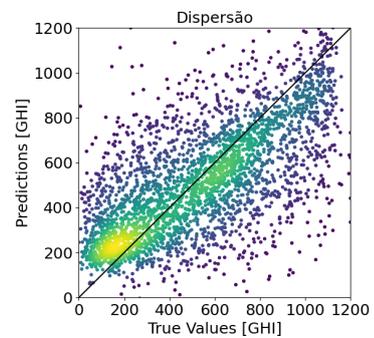
(c) Modelo M14.



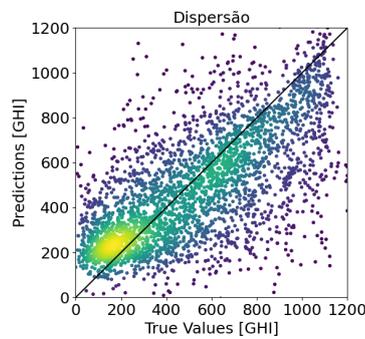
(d) Modelo M15.



(e) Modelo M16.



(f) Modelo M05.



(g) Modelo M06.

Figura B.5 – Histograma de distribuição do erro de predição.

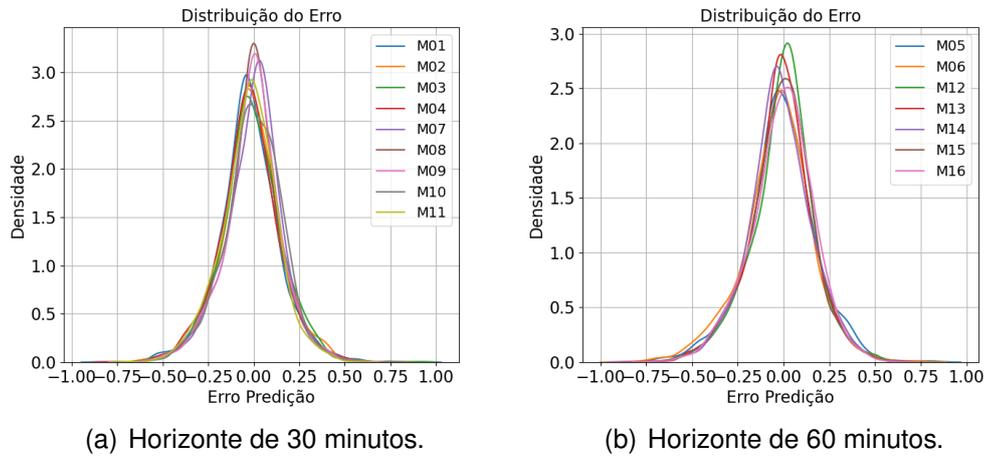


Figura B.6 – Predição de GHI pela ANN e GHI da estação terrestre, média horária do ano de 2019 com passo de 1h.

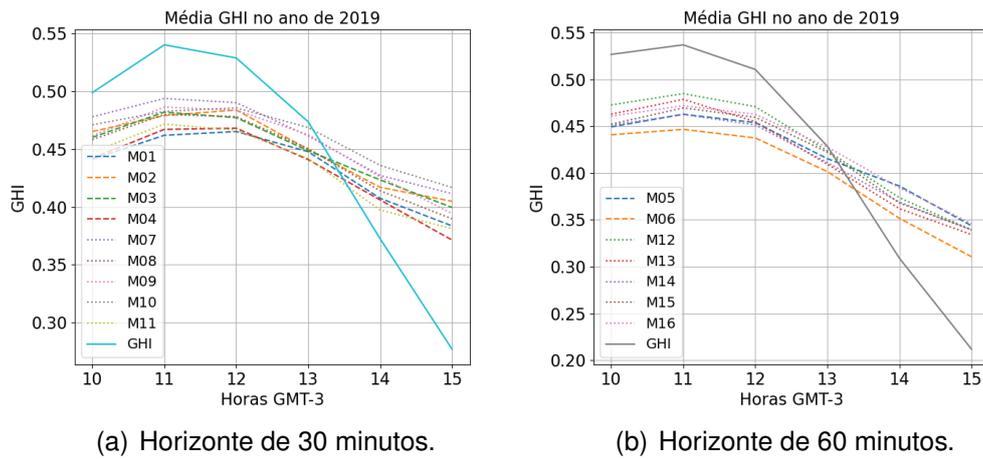


Figura B.7 – Predição de GHI pela ANN e GHI da estação terrestre, média horária do primeiro trimestre do ano de 2019 com passo de 1h.

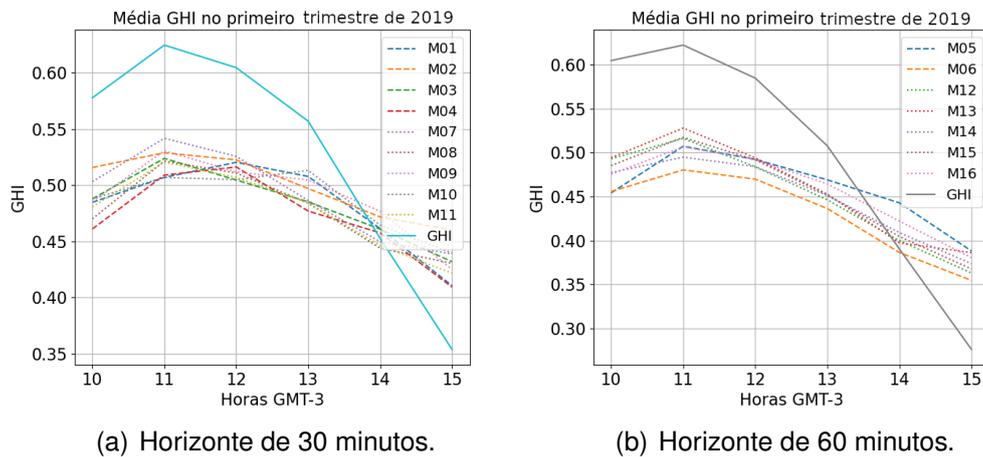
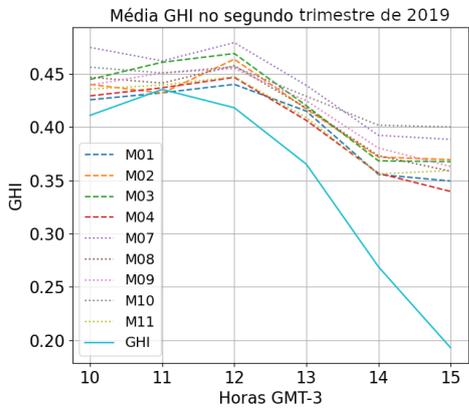
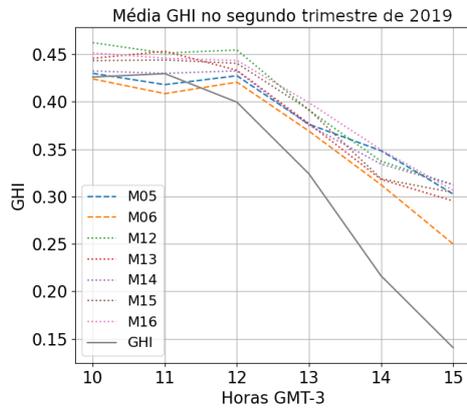


Figura B.8 – Predição de GHI pela ANN e GHI da estação terrestre, média horária do segundo trimestre do ano de 2019 com passo de 1h.

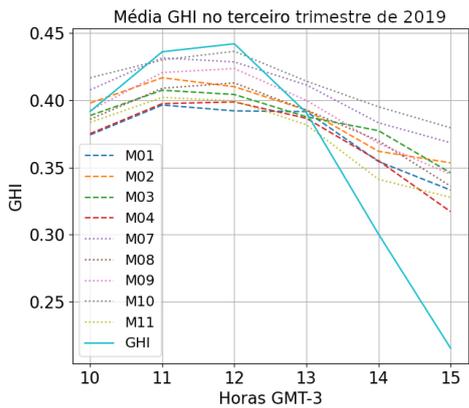


(a) Horizonte de 30 minutos.

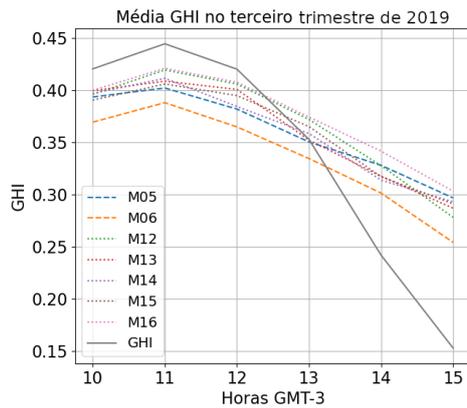


(b) Horizonte de 60 minutos.

Figura B.9 – Predição de GHI pela ANN e GHI da estação terrestre, média horária do terceiro trimestre do ano de 2019 com passo de 1h.

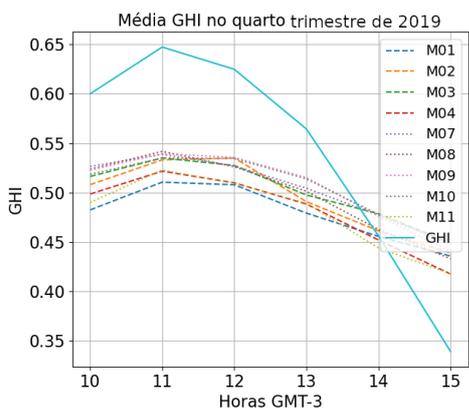


(a) Horizonte de 30 minutos.

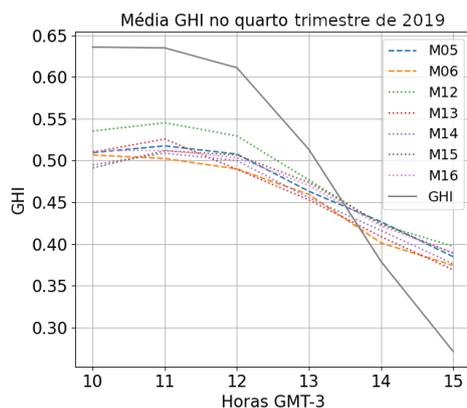


(b) Horizonte de 60 minutos.

Figura B.10 – Predição de GHI pela ANN e GHI da estação terrestre, média horária do quarto trimestre do ano de 2019 com passo de 1h.

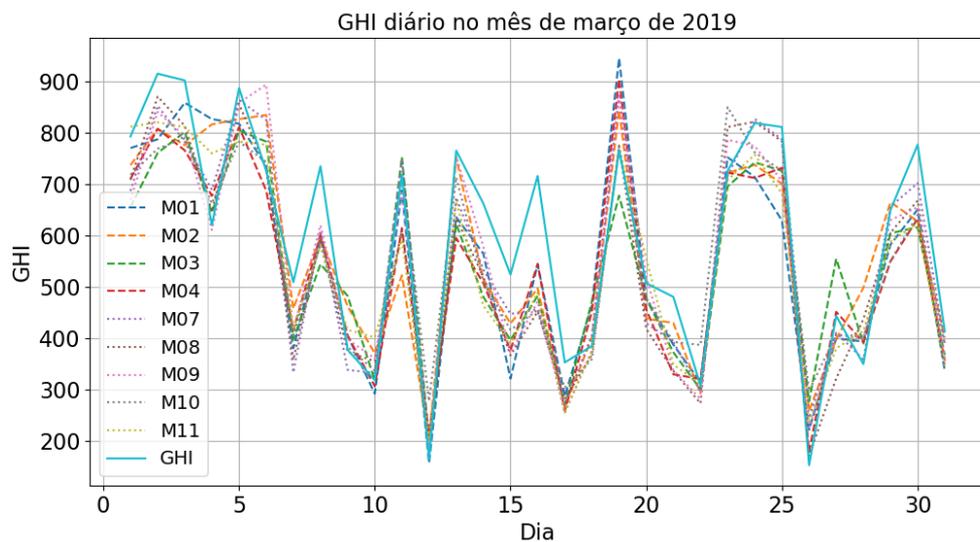


(a) Horizonte de 30 minutos.

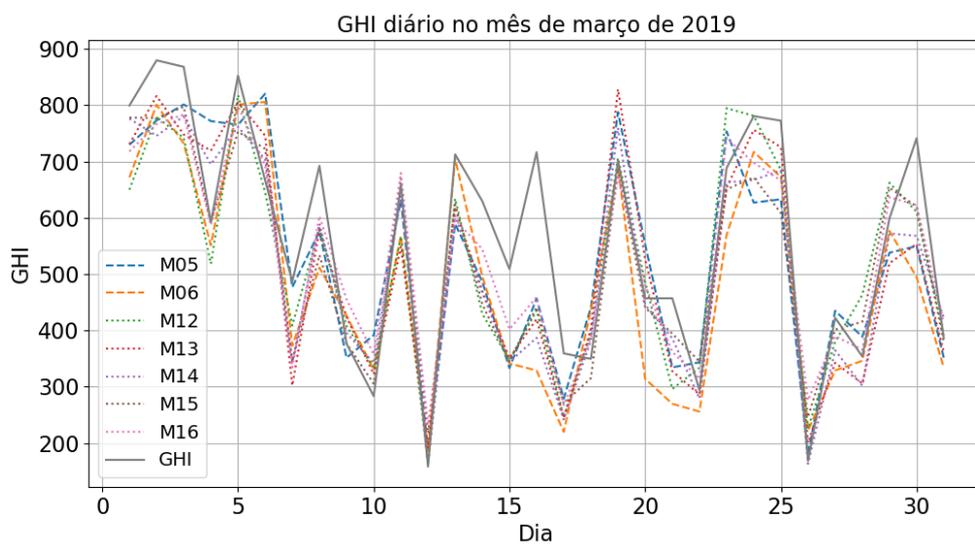


(b) Horizonte de 60 minutos.

Figura B.11 – Predição de GHI pela ANN e GHI da estação terrestre, média diária do mês de março de 2019.

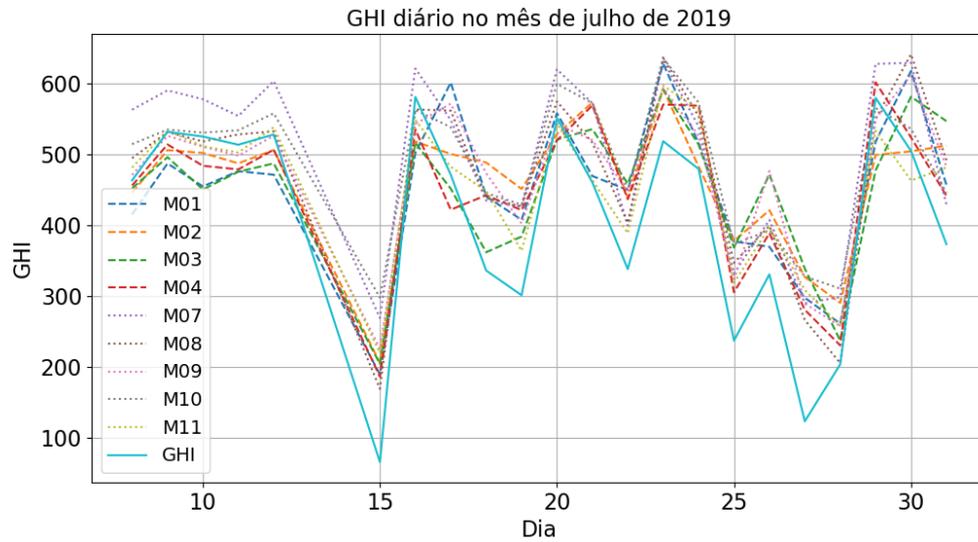


(a) Horizonte de 30 minutos.

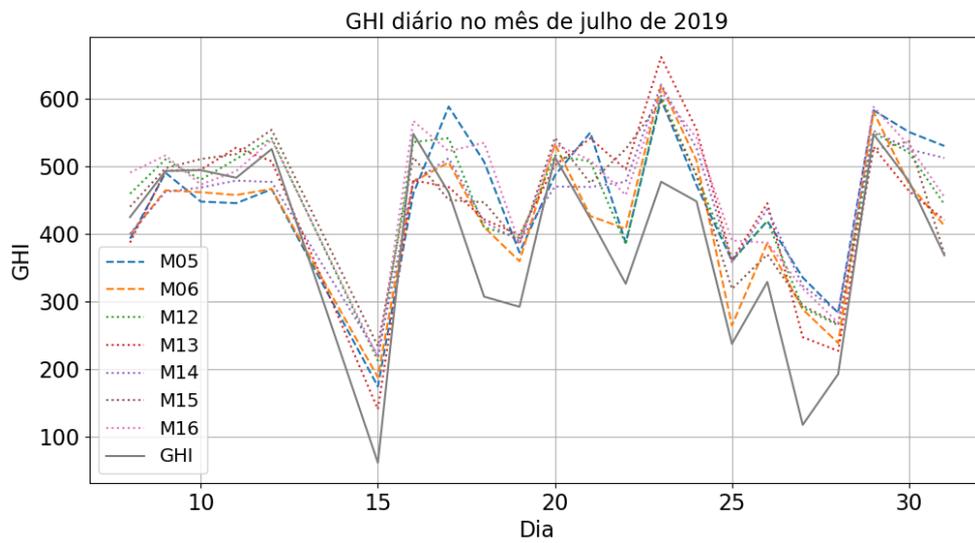


(b) Horizonte de 60 minutos.

Figura B.12 – Predição de GHI pela ANN e GHI da estação terrestre, média diária do mês de julho de 2019.

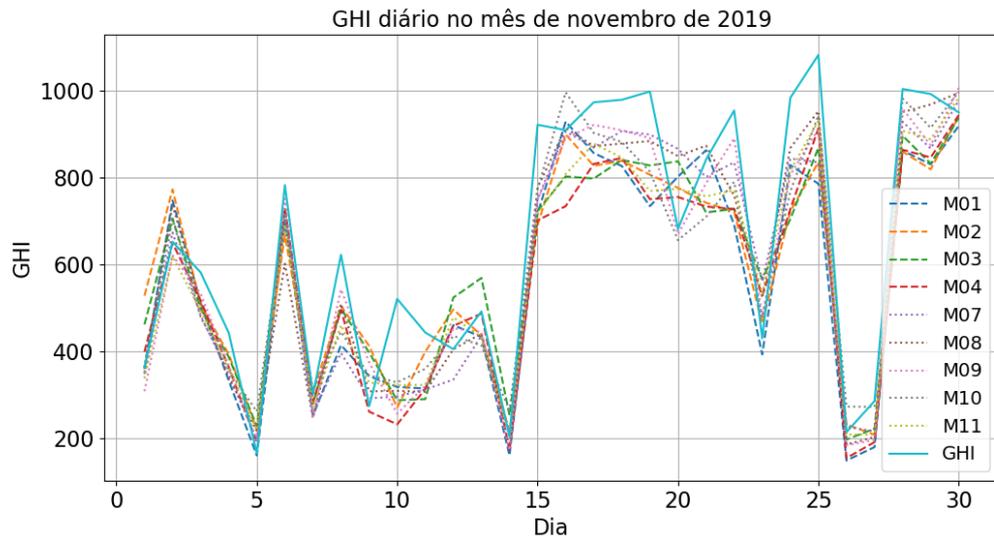


(a) Horizonte de 30 minutos.

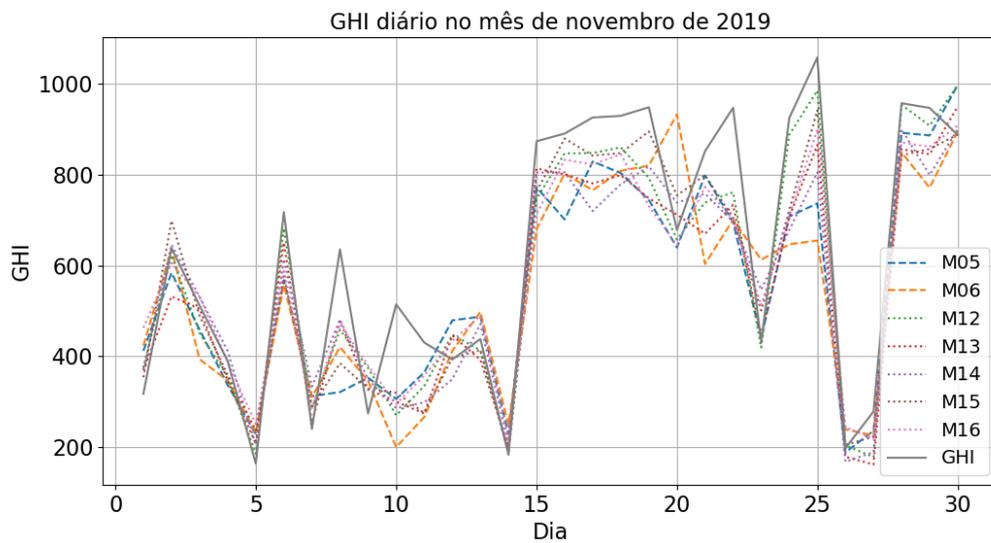


(b) Horizonte de 60 minutos.

Figura B.13 – Predição de GHI pela ANN e GHI da estação terrestre, média diária do mês de novembro de 2019.



(a) Horizonte de 30 minutos.



(b) Horizonte de 60 minutos.

APÊNDICE C – APRESENTAÇÃO DOS RESULTADOS

C.1 MELHORES RESULTADOS DE CADA CONFIGURAÇÃO

Tabela C.1 – Configuração e RMSE dos melhores 4 resultados para cada conjunto em um horizonte de 30 min.

#	RMSE W/m^2	Convoluções	Filtros	Kernel	Fusão Dados	Camadas MLP	Dropout	Neurônios	Parâmetros treinados	Épocas	Tempo de treino (s)
Configuração: Imagem											
1	182.27	1	32	3	-	1	0.15	64	93665	683	58.30
2	185.22	1	32	5	-	1	0.15	128	196193	609	58.15
3	185.27	1	32	3	-	2	0.15	64	97825	597	52.17
4	185.99	1	32	3	-	1	0.15	256	339809	611	54.09
Configuração: Imagem, DOY, Min											
1	128.21	1	16	3	16	2	0.15	128	35137	664	58.28
2	128.34	1	16	3	16	2	0.15	256	86977	686	60.07
3	128.95	1	16	3	0	2	0.15	256	236465	754	71.98
4	129.93	1	16	5	0	2	0.15	256	246705	641	62.63
Configuração: UV, DOY, Min											
1	101.08	-	-	-	-	2	0.15	256	67073	1718	73.54
2	101.22	-	-	-	-	2	0.15	128	17153	1739	74.72
3	102.93	-	-	-	-	1	0.15	256	1281	1745	73.51
4	103.78	-	-	-	-	1	0.15	128	641	1745	72.31
Configuração: AT, UV, RH, WS, DOY, Min											
1	101.54	-	-	-	-	2	0.15	128	17537	1707	72.99
2	101.62	-	-	-	-	2	0.15	256	67841	1662	71.02
3	102.40	-	-	-	-	1	0.15	128	1025	1739	72.20
4	102.58	-	-	-	-	1	0.15	256	2049	1740	72.43

Tabela C.2 – Continuação: configuração e RMSE dos melhores 4 resultados para cada conjunto em um horizonte de 30 min.

#	RMSE W/m^2	Convoluções	Filtros	Kernel	Fusão Dados	Camadas MLP	Dropout	Neurônios	Parâmetros treinados	Épocas	Tempo de treino (s)
Configuração: AT, RH, WS, DOY, Min											
1	180.95	-	-	-	-	2	0.15	256	67585	1733	103.01
2	189.54	-	-	-	-	2	0.15	128	17409	1741	101.70
3	195.06	-	-	-	-	2	0.15	64	4609	1725	101.60
4	202.14	-	-	-	-	1	0.15	256	1793	1750	100.49
Configuração: AT, RH, DOY, Min											
1	188.41	-	-	-	-	2	0.15	256	67329	1748	103.77
2	194.44	-	-	-	-	2	0.15	128	17281	1740	101.98
3	195.82	-	-	-	-	2	0.15	64	4545	1749	103.10
4	208.86	-	-	-	-	1	0.15	256	1537	1744	99.17
Configuração: Imagem, UV, DOY, Min											
1	94.36	1	16	5	0	2	0.15	256	246961	583	55.77
2	94.54	1	16	3	0	2	0.15	256	236721	582	51.87
3	94.93	1	16	3	0	2	0.15	64	51249	609	53.71
4	95.36	1	16	5	0	2	0.15	128	115121	579	55.17
Configuração: Imagem, AT, UV, RH, WS, DOY, Min											
1	94.55	1	16	3	0	2	0.15	256	237489	580	51.31
2	94.64	1	16	5	0	1	0.15	256	181937	805	76.83
3	95.36	1	16	5	0	2	0.15	128	115505	581	55.83
4	95.60	1	16	5	0	2	0.15	256	247729	602	58.64
Configuração: Imagem, AT, RH, WS, DOY, Min											
1	124.39	1	16	3	16	2	0.15	256	87745	646	65.37
2	125.05	1	16	5	16	2	0.15	256	97985	698	74.89
3	125.71	1	32	3	16	2	0.15	128	51569	604	60.93
4	126.45	1	32	3	0	2	0.15	256	406881	630	64.94
Configuração: Imagem, AT, RH, DOY, Min											
1	123.30	1	16	3	16	2	0.15	256	87489	614	61.47
2	125.65	1	32	3	16	2	0.15	128	51441	636	64.02
3	126.59	1	16	5	16	2	0.15	256	97729	658	70.49
4	126.92	1	32	3	16	2	0.15	256	103537	598	60.65

Tabela C.3 – Configuração e RMSE dos melhores 4 resultados para cada conjunto em um horizonte de 60 min.

#	RMSE W/m^2	Convoluções	Filtros	Kernel	Fusão Dados	Camadas MLP	Dropout	Neurônios	Parâmetros treinados	Épocas	Tempo de treino (s)
Configuração: Imagem											
1	205.85	1	32	5	-	1	0.15	128	196193	599	62.50
2	206.90	1	32	5	-	1	0.15	64	114145	629	65.39
3	207.18	2	16	3	-	1	0.15	64	18529	636	66.65
4	208.10	1	32	3	-	1	0.15	64	93665	678	65.46
Configuração: Imagem, DOY, Min											
1	130.87	1	16	3	16	2	0.15	256	86977	712	70.06
2	137.88	1	32	5	16	2	0.15	256	123505	604	64.98
3	138.33	1	16	5	0	2	0.15	256	246705	650	68.85
4	139.01	1	32	3	16	2	0.15	128	51185	651	64.32
Configuração: UV, DOY, Min											
1	126.38	-	-	-	-	2	0.15	256	67073	1655	72.08
2	126.48	-	-	-	-	2	0.15	128	17153	1740	74.79
3	126.86	-	-	-	-	2	0.15	64	4481	1749	75.22
4	132.38	-	-	-	-	1	0.15	256	1281	1951	82.00
Configuração: AT, UV, RH, WS, DOY, Min											
1	131.67	-	-	-	-	2	0.15	256	67841	1736	74.24
2	131.76	-	-	-	-	2	0.15	128	17537	1747	75.37
3	133.93	-	-	-	-	1	0.15	256	2049	1734	72.37
4	134.97	-	-	-	-	1	0.15	128	1025	1723	71.58
Configuração: AT, RH, WS, DOY, Min											
1	188.56	-	-	-	-	2	0.15	256	67585	1745	74.84
2	195.08	-	-	-	-	2	0.15	128	17409	1744	74.59
3	200.66	-	-	-	-	2	0.15	64	4609	1714	72.74
4	207.82	-	-	-	-	1	0.15	256	1793	1741	72.03

Tabela C.4 – Continuação: configuração e RMSE dos melhores 4 resultados para cada conjunto em um horizonte de 60 min.

#	RMSE W/m^2	Convoluções	Filtros	Kernel	Fusão Dados	Camadas MLP	Dropout	Neurônios	Parâmetros treinados	Épocas	Tempo de treino (s)
Configuração: AT, RH, DOY, Min											
1	188.04	-	-	-	-	2	0.15	256	67329	1749	75.62
2	192.97	-	-	-	-	2	0.15	128	17281	1746	76.29
3	201.53	-	-	-	-	2	0.15	64	4545	1747	75.24
4	214.56	-	-	-	-	1	0.15	256	1537	2873	119.06
Configuração: Imagem, UV, DOY, Min											
1	120.33	1	16	5	0	2	0.15	256	246961	594	58.51
2	121.35	1	32	3	0	2	0.15	256	406369	567	52.21
3	121.55	1	16	3	0	2	0.15	128	104881	571	52.80
4	121.68	1	16	5	0	2	0.15	128	115121	576	55.74
Configuração: Imagem, AT, UV, RH, WS, DOY, Min											
1	121.24	1	16	5	0	2	0.15	128	115505	574	55.10
2	121.51	1	16	3	0	2	0.15	128	105265	567	49.11
3	121.87	1	32	5	16	2	0.15	256	124529	575	56.30
4	121.89	1	32	3	16	2	0.15	256	104049	571	50.85
Configuração: Imagem, AT, RH, WS, DOY, Min											
1	134.40	1	32	5	16	2	0.15	128	72049	586	57.64
2	135.09	1	32	5	0	2	0.15	256	427361	590	57.98
3	135.71	1	16	5	16	2	0.15	256	97985	623	60.46
4	135.91	1	16	3	16	2	0.15	128	35521	624	55.04
Configuração: Imagem, AT, RH, DOY, Min											
1	135.24	1	16	3	16	2	0.15	128	35393	692	59.81
2	135.61	1	32	3	16	2	0.15	128	51441	601	53.18
3	136.16	1	16	5	16	2	0.15	128	45633	613	58.20
4	137.00	1	32	3	16	2	0.15	256	103537	595	52.63

Tabela C.5 – Configuração e RMSE dos melhores 4 resultados para cada conjunto em um horizonte de 120 min.

#	RMSE W/m^2	Convoluções	Filtros	Kernel	Fusão Dados	Camadas MLP	Dropout	Neurônios	Parâmetros treinados	Épocas	Tempo de treino (s)
Configuração: Imagem											
1	234.94	1	16	5	-	1	0.15	64	57137	707	59.07
2	235.28	1	16	5	-	1	0.15	128	98225	628	52.77
3	236.05	1	16	3	-	1	0.15	64	46897	681	52.45
4	237.63	2	32	5	-	1	0.15	64	78401	595	54.69
Configuração: Imagem, DOY, Min											
1	148.80	1	16	3	16	2	0.15	256	86977	746	58.08
2	150.96	1	32	3	0	2	0.15	256	406113	608	48.55
3	151.03	1	16	5	0	2	0.15	128	114993	708	61.07
4	151.97	1	32	5	0	2	0.15	256	426593	640	56.46
Configuração: UV, DOY, Min											
1	149.14	-	-	-	-	2	0.15	128	17153	1897	77.94
2	149.94	-	-	-	-	2	0.15	256	67073	1855	75.93
3	150.38	-	-	-	-	2	0.15	64	4481	1839	75.91
4	154.31	-	-	-	-	1	0.15	256	1281	1923	76.28
Configuração: AT, UV, RH, WS, DOY, Min											
1	161.38	-	-	-	-	2	0.15	256	67841	1916	73.61
2	164.00	-	-	-	-	2	0.15	128	17537	1926	74.28
3	170.15	-	-	-	-	2	0.15	64	4673	1896	72.08
4	171.24	-	-	-	-	1	0.15	256	2049	1923	71.37
Configuração: AT, RH, WS, DOY, Min											
1	187.18	-	-	-	-	2	0.15	256	67585	1928	81.14
2	191.12	-	-	-	-	2	0.15	128	17409	1926	81.66
3	198.21	-	-	-	-	2	0.15	64	4609	1913	80.94
4	218.29	-	-	-	-	1	0.15	256	1793	1918	78.88

Tabela C.6 – Continuação: configuração e RMSE dos melhores 4 resultados para cada conjunto em um horizonte de 120 min.

#	RMSE W/m^2	Convoluções	Filtros	Kernel	Fusão Dados	Camadas MLP	Dropout	Neurônios	Parâmetros treinados	Épocas	Tempo de treino (s)
Configuração: AT, RH, DOY, Min											
1	192.48	-	-	-	-	2	0.15	256	67329	1927	80.71
2	194.20	-	-	-	-	2	0.15	128	17281	1916	79.98
3	206.65	-	-	-	-	2	0.15	64	4545	1921	80.39
4	225.59	-	-	-	-	1	0.15	256	1537	1928	73.07
Configuração: Imagem, UV, DOY, Min											
1	143.86	1	16	3	0	2	0.15	128	104881	631	49.89
2	144.81	1	16	5	0	2	0.15	256	246961	609	54.99
3	145.37	1	32	5	0	2	0.15	256	426849	615	55.41
4	145.58	1	32	3	16	2	0.15	256	103281	600	48.77
Configuração: Imagem, AT, UV, RH, WS, DOY, Min											
1	144.56	1	32	5	0	2	0.15	256	427617	583	52.85
2	145.16	1	16	3	0	2	0.15	128	105265	599	47.42
3	145.49	1	32	5	16	2	0.15	256	124529	568	50.39
4	146.01	1	16	5	0	2	0.15	64	61681	582	51.75
Configuração: Imagem, AT, RH, WS, DOY, Min											
1	150.15	1	16	3	16	2	0.15	128	35521	708	56.33
2	150.30	1	32	3	0	2	0.15	256	406881	648	51.32
3	151.71	1	32	5	16	2	0.15	128	72049	601	54.81
4	151.91	1	16	5	16	2	0.15	128	45761	617	55.40
Configuração: Imagem, AT, RH, DOY, Min											
1	149.54	1	16	3	16	2	0.15	256	87489	767	62.60
2	150.24	1	32	5	0	2	0.15	256	427105	607	56.52
3	151.72	1	32	5	16	2	0.15	256	124017	601	55.96
4	152.29	1	16	3	0	2	0.15	128	105009	721	58.76

Tabela C.7 – Configuração e RMSE dos melhores 4 resultados para cada conjunto em um horizonte de 180 min.

#	RMSE W/m^2	Convoluções	Filtros	Kernel	Fusão Dados	Camadas MLP	Dropout	Neurônios	Parâmetros treinados	Épocas	Tempo de treino (s)
Configuração: Imagem											
1	244.06	1	16	5	-	1	0.15	256	180401	618	51.80
2	244.14	1	16	3	-	1	0.15	128	87985	678	55.59
3	245.08	1	16	5	-	1	0.15	64	57137	646	53.79
4	245.39	1	16	3	-	2	0.15	64	51057	625	50.96
Configuração: Imagem, DOY, Min											
1	154.82	1	32	5	0	2	0.15	256	426593	702	63.09
2	155.02	1	16	3	16	2	0.15	256	86977	774	65.67
3	155.12	1	16	3	0	2	0.15	128	104753	617	51.04
4	155.42	1	32	5	0	2	0.15	128	212961	603	52.59
Configuração: UV, DOY, Min											
1	154.43	-	-	-	-	2	0.15	64	4481	2145	87.27
2	155.41	-	-	-	-	2	0.15	128	17153	2151	87.65
3	156.42	-	-	-	-	2	0.15	256	67073	2110	86.72
4	161.77	-	-	-	-	1	0.15	256	1281	2166	83.37
Configuração: AT, UV, RH, WS, DOY, Min											
1	164.72	-	-	-	-	2	0.15	256	67841	2157	87.52
2	165.82	-	-	-	-	2	0.15	128	17537	2166	87.64
3	167.32	-	-	-	-	2	0.15	64	4673	2166	87.09
4	192.13	-	-	-	-	1	0.15	256	2049	2166	81.62
Configuração: AT, RH, WS, DOY, Min											
1	187.71	-	-	-	-	2	0.15	128	17409	2161	104.57
2	188.88	-	-	-	-	2	0.15	256	67585	2142	103.10
3	211.59	-	-	-	-	2	0.15	64	4609	2165	104.05
4	218.76	-	-	-	-	1	0.15	256	1793	2172	100.69

Tabela C.8 – Continuação: configuração e RMSE dos melhores 4 resultados para cada conjunto em um horizonte de 180 min.

#	RMSE W/m^2	Convoluções	Filtros	Kernel	Fusão Dados	Camadas MLP	Dropout	Neurônios	Parâmetros treinados	Épocas	Tempo de treino (s)
Configuração: AT, RH, DOY, Min											
1	188.92	-	-	-	-	2	0.15	256	67329	2150	101.85
2	192.02	-	-	-	-	2	0.15	128	17281	2162	103.38
3	216.60	-	-	-	-	2	0.15	64	4545	2165	102.79
4	223.30	-	-	-	-	1	0.15	256	1537	2160	100.12
Configuração: Imagem, UV, DOY, Min											
1	150.33	1	32	3	0	2	0.15	128	192609	627	57.39
2	152.17	1	32	5	0	2	0.15	128	213089	594	53.79
3	153.33	1	16	5	0	2	0.15	128	115121	639	57.12
4	153.35	1	16	3	0	2	0.15	256	236721	624	56.55
Configuração: Imagem, AT, UV, RH, WS, DOY, Min											
1	153.40	1	32	5	0	2	0.15	256	427617	611	56.15
2	154.74	1	32	3	0	2	0.15	256	407137	634	56.95
3	154.77	1	32	3	0	2	0.15	128	192993	618	55.97
4	155.34	1	32	5	16	2	0.15	256	124529	613	56.66
Configuração: Imagem, AT, RH, WS, DOY, Min											
1	154.50	1	32	3	16	2	0.15	128	51569	685	55.77
2	155.67	1	32	3	0	2	0.15	256	406881	676	55.66
3	155.68	1	16	5	0	2	0.15	128	115377	714	61.60
4	155.98	1	16	3	16	2	0.15	256	87745	731	58.84
Configuração: Imagem, AT, RH, DOY, Min											
1	154.87	1	32	3	0	2	0.15	128	192737	646	51.65
2	155.70	1	32	5	16	2	0.15	256	124017	671	58.39
3	157.06	1	32	3	16	2	0.15	128	51441	696	55.07
4	157.28	1	32	5	0	2	0.15	256	427105	613	53.61

C.2 DISTRIBUIÇÃO DOS RESULTADOS PARA CADA PARÂMETRO DA REDE

Abaixo são apresentadas as distribuições do RMSE para cada um dos parâmetros de treinamento da rede. Nestes gráficos é possível avaliar a influência de um determinado parâmetro na distribuição do erro.

Figura C.1 – Distribuição do resultado para a Configuração Img no horizonte de 30 min.

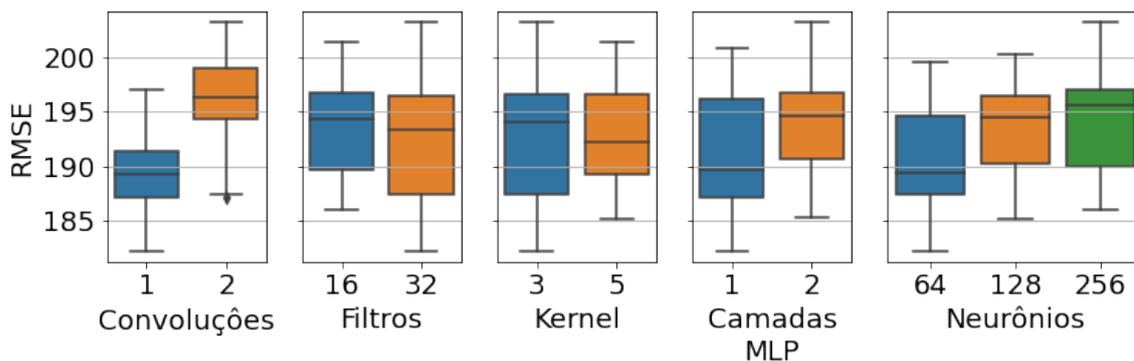


Figura C.2 – Distribuição do resultado para a Configuração Img no horizonte de 60 min.

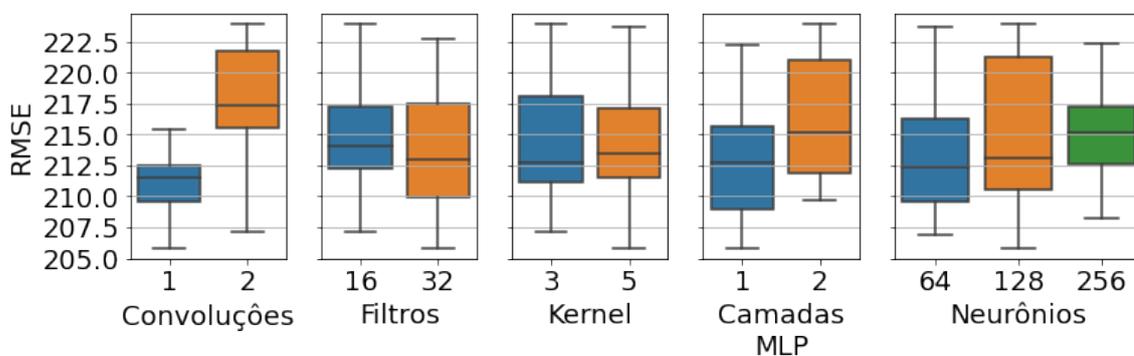


Figura C.3 – Distribuição do resultado para a Configuração Img no horizonte de 120 min.

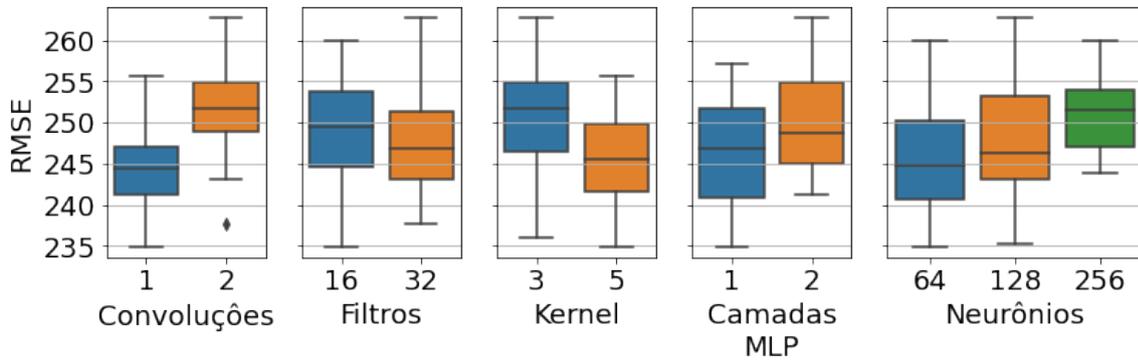


Figura C.4 – Distribuição do resultado para a Configuração Img no horizonte de 180 min.

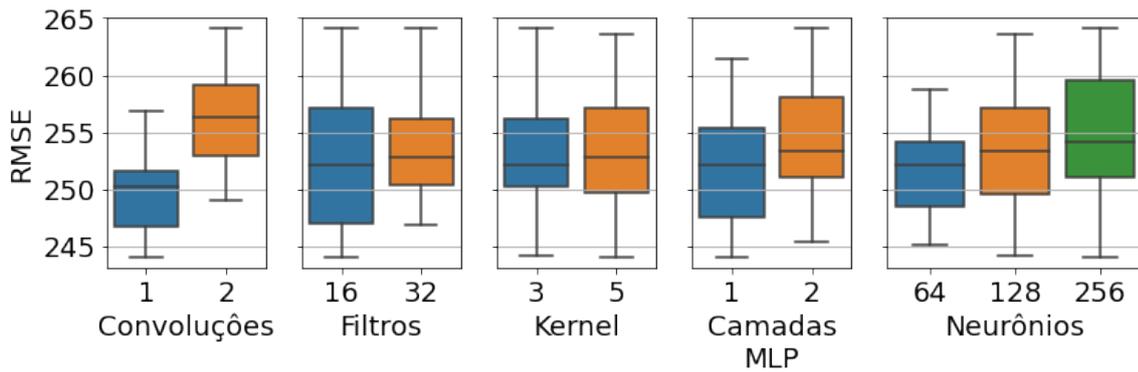


Figura C.5 – Distribuição do resultado para a Configuração Img+t no horizonte de 30 min.

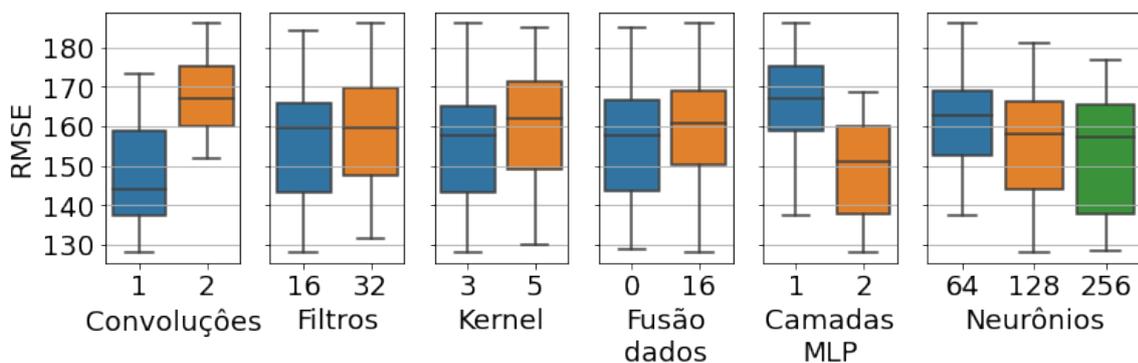


Figura C.6 – Distribuição do resultado para a Configuração Img+t no horizonte de 60 min.

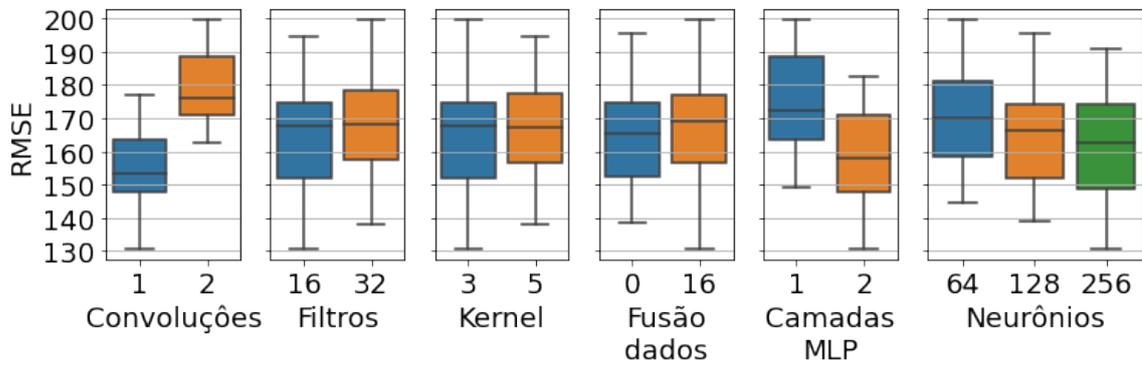


Figura C.7 – Distribuição do resultado para a Configuração Img+t no horizonte de 120 min.

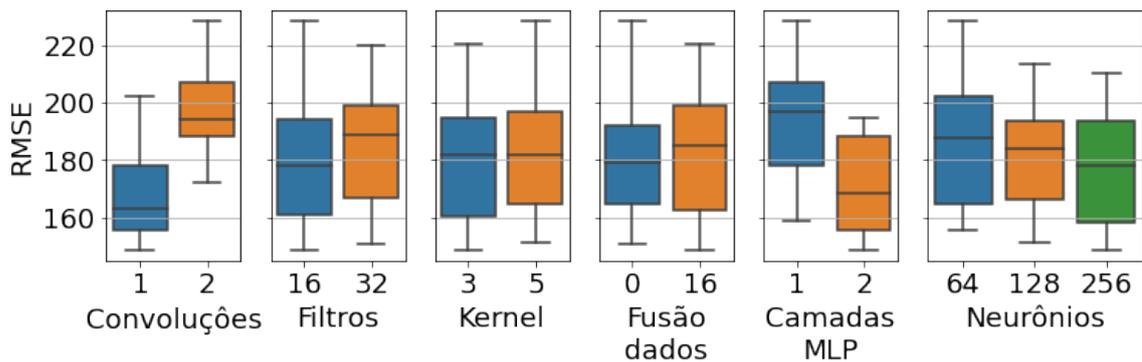


Figura C.8 – Distribuição do resultado para a Configuração Img+t no horizonte de 180 min.

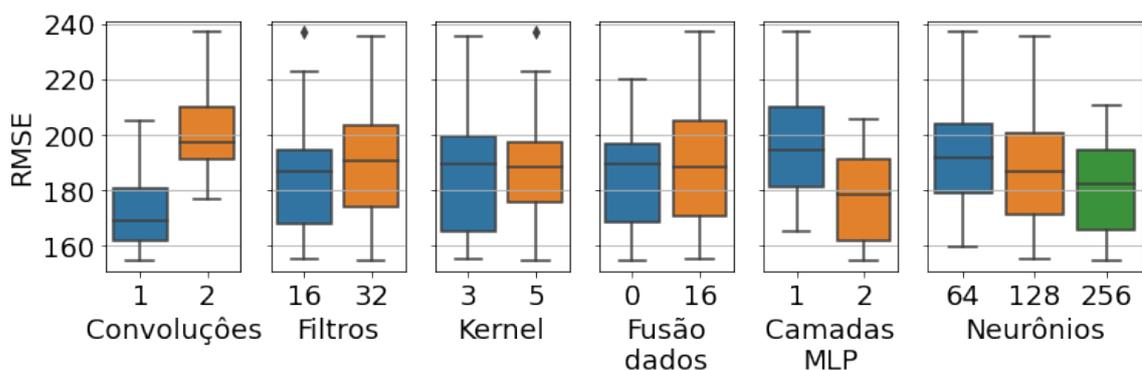


Figura C.9 – Distribuição do resultado para a Configuração UV+t, UV+AT+RH+WS+t, AT+RH+WS+t e AT+RH+t no horizonte de 30 min.

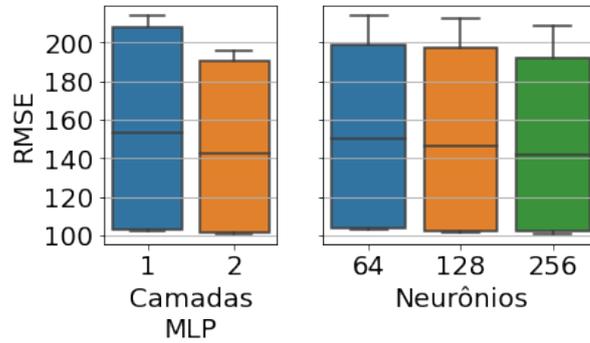


Figura C.10 – Distribuição do resultado para a Configuração UV+t, UV+AT+RH+WS+t, AT+RH+WS+t e AT+RH+t no horizonte de 60 min.

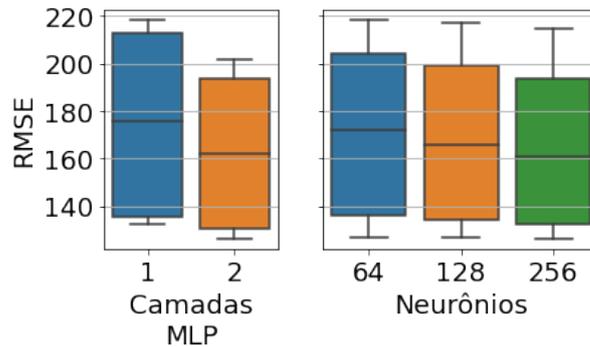


Figura C.11 – Distribuição do resultado para a Configuração UV+t, UV+AT+RH+WS+t, AT+RH+WS+t e AT+RH+t no horizonte de 120 min.

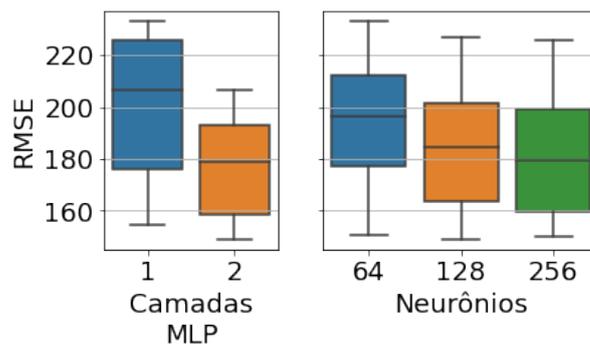


Figura C.12 – Distribuição do resultado para a Configuração UV+t, UV+AT+RH+WS+t, AT+RH+WS+t e AT+RH+t no horizonte de 180 min.

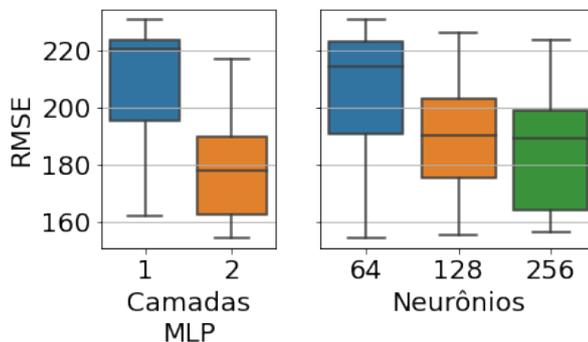


Figura C.13 – Distribuição do resultado para a Configuração Img+UV+t, Img+UV+AT+RH+WS+t, Img+AT+RH+WS+t, Img+AT+RH+t no horizonte de 30 min.

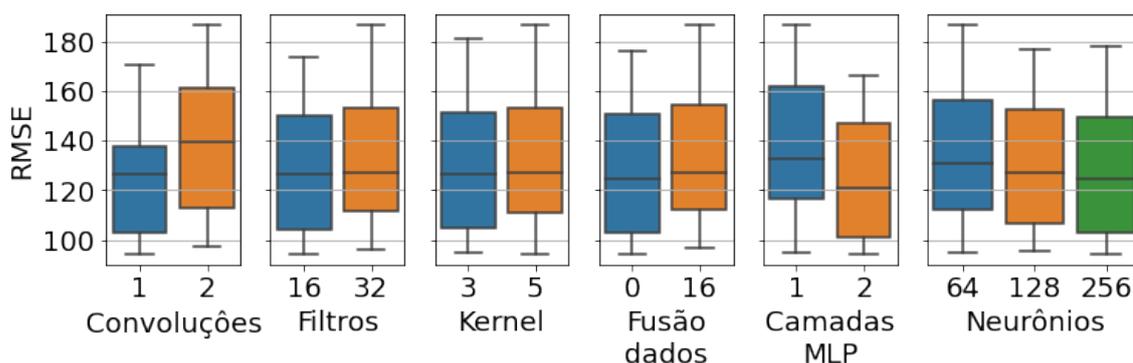


Figura C.14 – Distribuição do resultado para a Configuração Img+UV+t, Img+UV+AT+RH+WS+t, Img+AT+RH+WS+t, Img+AT+RH+t no horizonte de 60 min.

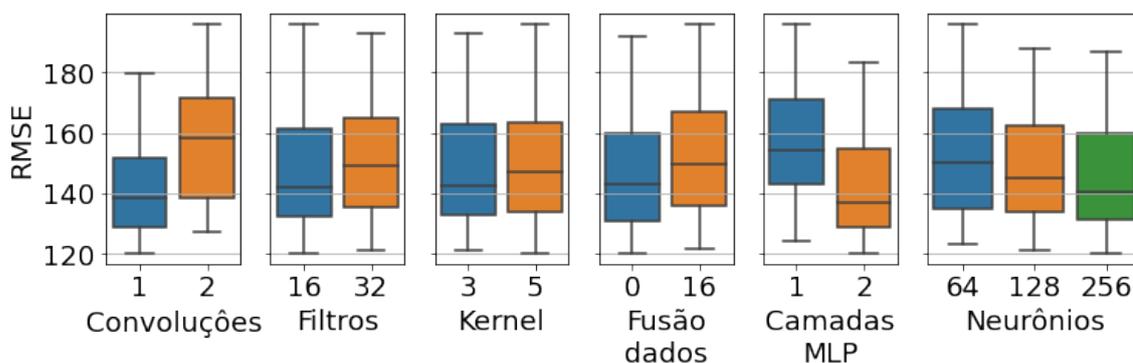


Figura C.15 – Distribuição do resultado para a Configuração Img+UV+t , Img+UV+AT+RH+WS+t , Img+AT+RH+WS+t , Img+AT+RH+t no horizonte de 120 min.

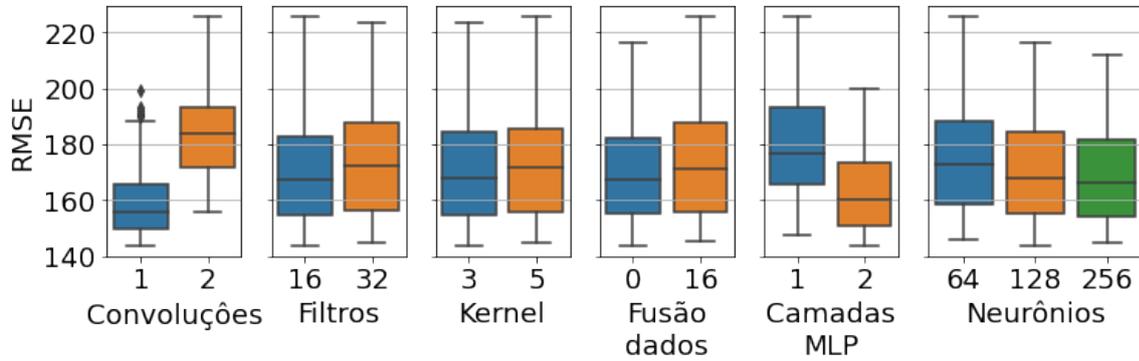


Figura C.16 – Distribuição do resultado para a Configuração Img+UV+t , Img+UV+AT+RH+WS+t , Img+AT+RH+WS+t , Img+AT+RH+t no horizonte de 180 min.

