

Universidade Federal de São Carlos – UFSCar
Programa de Pós-Graduação em Ciência da Computação
Eduarth Dapper Heinen

**Redes Neurais Recorrentes e *XGBoost*
Aplicados à Previsão de Radiação Solar no
Horizonte de Curto Prazo**

São Carlos - SP

Agosto/2018

Eduarth Dapper Heinen

Redes Neurais Recorrentes e *XGBoost* Aplicados à Previsão de Radiação Solar no Horizonte de Curto Prazo

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação. Área de Concentração: Inteligência Artificial.

Orientador: Prof. Dr. Ricardo A. S. Fernandes

São Carlos - SP

Agosto/2018

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

“Na verdade, a compreensão de que a natureza de todas as coisas muda constantemente é a chave para a verdadeira felicidade, pois significa que por pior que seja a situação, ela mudará.”
(Daisaku Iideka)

Resumo

Estimativas precisas da geração e consumo de eletricidade são necessárias para promover a integração de fontes de energia renováveis à malha elétrica, orientando o equilíbrio do suprimento de eletricidade através dos ciclos e flutuações destes recursos. Neste cenário, diversos experimentos têm desenvolvido métodos de previsão de radiação solar, mas poucos destes aplicam Redes Neurais Recorrentes e seu potencial de modelar séries temporais. Esta dissertação revisa alguns destes estudos e apresenta um experimento comparando Redes Neurais Recorrentes, *XGBoost* e modelos de persistência dos índices de Céu Claro e Claridade em estimativas de irradiância solar. A precisão destes modelos é verificada em observações minuto a minuto de Denver e Las Vegas, re-amostradas em frequências de 5 minutos, 30 minutos, 1 hora e média diária. O modelo *XGBoost* apresenta os melhores resultados em todos os horizontes de previsão, com nRMSE entre 12.9% em resoluções de 5 minutos, e 21.2% em média diária. Modelos de persistência dos índices de Céu Claro e Claridade demonstram precisão comparável, com nRMSE de 14.2% a 22.5% em previsões até 1 hora à frente. RNNs superam modelos de persistência em previsões de irradiância média diária, alcançando nRMSE de 21.3%.

Palavras-chaves: Redes neurais artificiais, RNN, Radiação solar, Previsão, Seleção de atributos, XGBoost, Modelos de Persistência, Índice de Céu Claro, Índice de Claridade

Abstract

Precise estimates of energy production and consumption are essential to promote the integration of renewable energy sources to the electrical grid, guiding power supply balance through cycles and fluctuations inherent to these resources. In this scenario, many experiments have developed solar radiation forecasting methods, but not so many of those apply Recurrent Neural Networks and their potential of modeling time series. This dissertation reviews some of the studies and outlines an experiment comparing Recurrent Neural Networks, *XGBoost* and persistence of Clear Sky and Clearness indexes in solar irradiance forecasts. Model precision is verified in minute observations from Denver and Las Vegas, re-sampled in resolutions of 5 minutes, 30 minutes, 1 hour and daily mean. XGBoost gives the best results in all forecast horizons, with nRMSE between 12.9% at 5 minute resolutions, and 21.2% at daily mean. Persistence models for Clear Sky and Clearness indexes show comparable precision, with nRMSE from 14.2% to 22.5% in forecasts up to 1 hour ahead. RNNs outperform persistence in forecasts of daily mean irradiance, reaching nRMSE 21.3%.

Key-words: Artificial neural networks, RNN, Solar radiation, Forecasts, Feature selection, XGBoost, Persistence Models, Clear Sky Index, Clearness Index

Lista de ilustrações

Figura 1	– Mapa de palavras-chave destacando termos suprimidos na <i>query</i> de busca.	29
Figura 2	– Redes de Citações, antes e depois da seleção por título e resumo. . . .	30
Figura 3	– Modelo de Céu Claro desenvolvido por Pedro e Coimbra (2012), cuja superfície corresponde à expectativa de produção a cada dia do ano τ_Y e hora do dia τ_D	33
Figura 4	– Representação de um Neurônio Artificial. Fonte: Haykin (2009).	37
Figura 5	– Funções de ativação Limiar (<i>Threshold</i>) e Sigmóide. Fonte: Haykin (2009).	37
Figura 6	– Diagramas e regiões de decisão dos três neurônios artificiais <i>perceptron</i> combinados para simular o operador OU-exclusivo (XOR).	38
Figura 7	– Arquitetura de um Multilayer Perceptron.	39
Figura 8	– Conexões entre neurônios ocultos em arquiteturas de Redes Neurais Recorrentes (RNNs).	40
Figura 9	– Componentes da irradiância solar decompostos por Transformada de <i>Wavelets</i> . Fonte: Cao e Cao (2005)	42
Figura 10	– Exemplo de árvore de regressão.	43
Figura 11	– Resultados das previsões obtidas para o experimento inicial.	47
Figura 12	– Representação gráfica do modelo de Céu Claro elaborado para averiguar a expectativa de irradiância máxima correspondente às horas do dia e dias do ano, antes e após regularização por RBF.	49
Figura 13	– Proporções dos conjuntos resultantes do pré-processamento e re-amostragem dos dados para a cidade de Las Vegas.	50
Figura 14	– Pré-processamento, união e re-amostragem dos conjuntos de dados. . .	51
Figura 15	– Atributos selecionados por RFE, com ênfase aos responsáveis por 50% da importância total do sub-conjunto.	52
Figura 16	– Procedimento de treinamento e avaliação dos modelos.	55
Figura 17	– Raiz do Erro Médio Quadrático Normalizada (nRMSE) médio através dos conjuntos de dados, para cada modelo e horizonte.	56
Figura 18	– Estimativas do modelo XGBoost entre os dias 8 e 12 de junho de 2015, comparadas à irradiância observada.	56
Figura 19	– Estimativas do modelo de persistência do índice de céu claro, entre os dias 8 e 12 de junho de 2015, comparadas à irradiância observada. . . .	57
Figura 20	– Estimativas do modelo de persistência do índice de claridade, entre os dias 8 e 12 de junho de 2015, comparadas à irradiância observada. Seguidas da projeção do dia 10 de junho, entre as 8 e 16 horas.	58

Figura 21 – Estimativas do modelo RNN, entre os dias 8 e 12 de junho de 2015,
comparadas à irradiância observada. 59

Lista de tabelas

Tabela 1 – Lista dos 10 documentos mais citados.	31
Tabela 2 – Atributos melhor classificados por cada método de seleção.	46
Tabela 3 – Erros (nRMSE e Erro Absoluto Médio (MAE)) calculados para os distintos períodos de análise do experimento inicial.	47
Tabela 4 – Primeiros 5 atributos selecionados em cada resolução através dos conjuntos.	53
Tabela 5 – nRMSE e MAPE de todos os modelos e horizontes examinados para a cidade de Las Vegas. Valores decimais representam porcentagens (0.25 = 25%)	67
Tabela 6 – Atributos melhor classificados por cada método de seleção.	67
Tabela 7 – nRMSE e MAPE de todos os modelos e horizontes examinados para a cidade de Denver. Valores decimais representam porcentagens (0.25 = 25%)	68

Glossário

Erro Absoluto Médio (MAE) do inglês *Mean Absolute Error* é determinado pela Equação 1, sendo \hat{y}_i a estimativa, y_i o valor observado e N o número de exemplos.

$$MAE = \sum_{i=1}^N |y_i - \hat{y}_i| \quad (1)$$

Erro Percentual Absoluto Médio (MAPE) do inglês *Mean Absolute Percentage Error* é determinado pela Equação 2, sendo \hat{y}_i a estimativa, y_i o valor observado e N o número de exemplos.

$$MAPE = \frac{100\%}{N} \times \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (2)$$

Raiz do Erro Médio Quadrático (RMSE) do inglês, *Root Mean Squared Error*, é determinada pela Equação 3, sendo \hat{y}_i a estimativa, y_i o valor observado e N o número de exemplos.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (3)$$

Raiz do Erro Médio Quadrático Normalizada (nRMSE) do inglês, *Normalized Root Mean Squared Error*, é determinada pela Equação 4, sendo \hat{y}_i a estimativa, y_i o valor observado e N o número de exemplos. O termo de normalização \bar{y} frequentemente refere-se à média das observações na amostra, contudo difere entre autores.

$$nRMSE = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2}}{\bar{y}} \quad (4)$$

Siglas

ANN Redes Neurais Artificiais, do inglês *Artificial Neural Networks*

CNN Rede Neural Convolutacional, do inglês *Convolutional Neural Networks*

DHI Irradiância Horizontal Difusa, do inglês *Difuse Horizontal Irradiance*

DNI Irradiância Normal Direta, do inglês *Direct Normal Irradiance*

GHI Irradiância Global Horizontal, do inglês *Global Horizontal Irradiance*

LM Algoritmo de Levenberg Marquardt

MAE Erro Absoluto Médio, do inglês *Mean Absolute Error*

MAPE Erro Percentual Absoluto Médio, do inglês *Mean Absolute Percentage Error*

MLP Multilayer Perceptron

nRMSE Raiz do Erro Médio Quadrático Normalizada, do inglês *Normalised Root Mean Squared Error*

NWP Previsões Numéricas do Clima, do inglês *Numerical Weather Prediction*

RBF Função de *Kernel* de Base Radial, do inglês *Radial Basis Function*

RFE Eliminação Recursiva de Atributos, do inglês *Recursive Feature Elimination*

RMSE Raiz do Erro Médio Quadrático, do inglês *Root Mean Squared Error*

RNN Rede Neural Recorrente, do inglês *Recurrent Neural Networks*

XGBoost *eXtreme Gradient Boosting*

XOR OU-exclusivo, do inglês *eXclusive OR*

Sumário

1	Introdução	21
1.1	Mudança Climática, Combustíveis Fósseis e Geração de Energia	21
1.2	Motivação e Justificativa	22
1.3	Hipótese e Objetivos	24
1.4	Organização da Dissertação	24
2	Fundamentos sobre Radiação Solar	25
2.1	Componentes da Radiação Solar Incidente	25
2.2	Modelos e Índices de Céu Claro e Claridade	25
2.2.1	Modelos e Índice de Céu Claro	26
2.2.2	Radiação Extraterrestre e Índice de claridade	27
2.3	Previsão Numérica do Clima	27
3	Revisão da Literatura	29
3.1	Seleção da Bibliografia e Metanálise	29
3.2	Modelos de Previsão de Irradiância ou Produção Fotovoltaica	31
3.2.1	Modelos de Persistência	31
3.2.2	Modelos Probabilísticos de Séries Temporais	33
3.2.3	Técnicas de Aprendizado de Máquina	36
3.2.3.1	Multilayer Perceptron (MLP)	38
3.2.3.2	Redes Neurais Recorrentes (RNN)	40
3.2.3.3	Extreme Gradient Boosting (XGboost)	43
4	Metodologia e Resultados	45
4.1	Experimento Inicial	45
4.1.1	Pré-processamento e Seleção de Atributos	45
4.1.2	Experimentos e Resultados	46
4.2	Metodologia Proposta após Experimentos Iniciais	48
4.2.1	Conjuntos de Dados	48
4.2.2	Seleção de Atributos	51
4.2.3	Experimentos	53
4.3	Resultados e Discussão	55
5	Considerações Finais	61
	Referências	63

APÊNDICE A Medidas de Erro para todos os Modelos e Horizontes 67

1 Introdução

Este capítulo foi organizado em três seções: a primeira discute o problema de pesquisa, tornando clara a importância da substituição de combustíveis fósseis por fontes renováveis na produção de eletricidade; a segunda seção traz as motivações e justificativas que permeiam essa dissertação, principalmente a capacidade de antecipar e responder às variações de recursos fotovoltaicos; na sequência, são apresentadas as hipóteses e objetivos; e, por fim, a organização da dissertação em termos de seus capítulos.

1.1 Mudança Climática, Combustíveis Fósseis e Geração de Energia

Durante as Convenções das Nações Unidas sobre Mudanças Climáticas, realizadas anualmente desde 1995, governantes e membros da comunidade científica discutem a interferência humana no clima, bem como suas consequências. Destas convenções resultaram o Protocolo de Kyoto, o Acordo de Cancún e o Acordo de Paris, este último assinado em dezembro de 2015 na 21ª Conferência das Partes. O novo acordo estabelece um compromisso entre 197 países, com o objetivo de impedir que a temperatura média global se eleve 2°C quando comparada a períodos anteriores à revolução industrial (UNFCCC, 2015). Manter o aquecimento abaixo deste limite deve reduzir o risco de a população mundial enfrentar escassez de comida e água, eventos climáticos extremos e as adversidades sociais e econômicas causadas por estes fenômenos (PACHAURI et al., 2014).

O Painel Intergovernamental sobre Mudança Climática (IPCC) é o órgão da ONU responsável por elaborar revisões da evidência científica sobre mudança climática. Seu último relatório sumariza 1184 cenários que contemplam emissões, concentrações de diversos gases e uso do solo até 2100. Destes cenários são extraídas 4 projeções representativas das medidas necessárias para impedir que concentrações de Gases do Efeito Estufa (GEE) ultrapassem determinados patamares. Somente em uma delas o aquecimento é contido abaixo de 2°C, sendo necessário reduzir entre 40% a 70% das emissões, aumentar três a quatro vezes a parcela de energia produzida por fontes limpas e capturar grandes quantidades de carbono da atmosfera, alcançando emissões anuais nulas ou negativas em 2100 (PACHAURI et al., 2014).

Ademais, ao comparar estimativas recentes, pressupondo a adoção irrestrita de todas as políticas do acordo, com a trajetória média das projeções do IPCC em que o aquecimento é controlado, Rogelj et al. (2016) encontraram uma diferença de 13 ± 3 Gt CO₂-eq – gases equivalentes em termos de aquecimento da atmosfera a mesma quantidade

de CO₂ – anuais até 2030. Portanto, a trajetória de emissões não somente desvia da meta do acordo, como é consistente com o aquecimento médio de 2,6 a 3,1°C em 2100.

Emissões antropogênicas de GEE escalaram de níveis próximos de zero em 1750, até ultrapassarem 49 Gt CO₂-eq anuais em 2010. Somente no período entre 1970 e 2010, foram produzidas metade de todas as emissões provenientes de atividades humanas na história (PACHAURI et al., 2014). Durante este período, o desenvolvimento econômico foi sustentado pela crescente oferta de energia, suprida majoritariamente por combustíveis fósseis. De maneira que a infraestrutura de produção de eletricidade, transporte e processos industriais estabelecida é dependente destes combustíveis e intensamente poluente (HÖÖK; TANG, 2013).

A IEA (*International Energy Agency*) estima que, em 2014, a produção de energia elétrica e o aquecimento doméstico foram responsáveis por 42% das emissões globais provenientes de combustíveis fósseis (IEA, 2016a). Especificamente, nos Estados Unidos, em 2015, 35% das emissões de CO₂ resultaram da geração de energia, consequência do alto consumo de combustíveis fósseis no setor elétrico, que emprega carvão mineral, gás natural e petróleo como fontes de 82% da energia produzida (EPA, 2017). No Brasil, em virtude da predominância de usinas hidroelétricas, a produção de energia elétrica provocou somente 8% das emissões por queima de combustíveis fósseis em 2010. No mesmo período, a principal fonte de GEE foi a conversão de áreas de floresta para uso agropecuário, responsável por 42% do CO₂ emitido (BRASIL, 2016).

Portanto, reduzir a intensidade de carbono na geração de energia elétrica é parte fundamental dos esforços para conter o aquecimento, bem como são necessários avanços em eficiência e mudanças de hábitos dos consumidores para que a demanda energética mundial possa ser reduzida. Nos cenários do IPCC, em que o objetivo do acordo é cumprido, a parcela de fontes de baixo carbono (renováveis, nuclear e bioenergia com captura de carbono) no suprimento de eletricidade escala de aproximadamente 25-30% atualmente, para 80-85% em 2050, cessando por completo a geração por combustíveis fósseis até 2100 (PACHAURI et al., 2014; IRENA, 2018). Porém, alcançar estes objetivos, requer rápida expansão da capacidade produtiva e adequação dos sistemas de energia para responder aos ciclos e variações destas fontes.

1.2 Motivação e Justificativa

Em 2015, Fontes de Energia Renováveis (RES, do inglês *Renewable Energy Sources*) representaram mais da metade da capacidade de geração instalada, ultrapassando pela primeira vez na história a construção de usinas fósseis e nucleares combinadas. No ano seguinte, RES compreendiam 24% da produção mundial de energia, compostas principalmente de energia hidroelétrica (70%), seguida da eólica (16%), bioenergia (9%) e

fotovoltaica (5%) [IEA \(2016b\)](#).

A energia fotovoltaica, especificamente, apresentou reduções de 40% a 75% nos custos de instalação entre 2010 e 2015. E, embora o LCOE¹ de instalações fotovoltaicas continue superior ao de hidroelétricas, termoelétricas modernas e usinas a gás com turbinas de ciclo combinado, a capacidade de geração fotovoltaica, equivalente a 225 GW em 2014, cresceu 49 GW em 2015, 70-75 GW em 2016, 94 GW em 2017, e ainda deve triplicar até 2020 [IEA \(2016b\)](#), [IEA \(2017\)](#), [IRENA \(2018\)](#).

Contudo, o principal obstáculo à extensa adoção de fontes renováveis é a integração destas aos sistemas elétricos convencionais. Em contraste com usinas baseadas em combustíveis fósseis – instalações de larga escala, capazes de ajustar seu rendimento aos picos e declínios de demanda – a variabilidade inerente aos recursos renováveis desafia a operação e estabilidade dos sistemas elétricos de potência. A produção de energia fotovoltaica, por exemplo, acompanha a radiação solar que, por sua vez, está sujeita à localização geográfica, condições climáticas e época do ano. Embora parte destes aspectos possa ser determinada com precisão, ainda resta uma significativa margem de incerteza, causada principalmente pelas condições meteorológicas ([EDENHOFER et al., 2011](#)).

Ademais, sistemas elétricos que contenham grandes proporções de RES precisam ser capazes de responder aos ciclos e flutuações destes recursos. Para isso, se fazem necessárias medidas para equilibrar o suprimento de energia sem restringir a produção renovável, como: coordenar usinas para que acompanhem a variação das fontes renováveis, por exemplo, reduzindo o rendimento de usinas a gás nos momentos de maior potencial fotovoltaico; acionar usinas de apoio ou transferir energia elétrica de outras regiões durante os períodos improdutivos; e ampliar a capacidade em épocas de alta disponibilidade ([IEA, 2016b](#)).

Previsões em horizontes distintos reduzem as incertezas e auxiliam a operação dos sistemas elétricos de potência. Estimativas da produção e consumo para o dia seguinte, por exemplo, permitem alocar usinas capazes de suprir a demanda da região, assim como estimativas em horizontes de algumas horas alertam para flutuações da produção e permitem tomar medidas para evitar a interrupção do suprimento. Motivados pela renovação do setor energético e pelas limitações mencionadas, grupos de pesquisa têm se dedicado à previsão da radiação solar para distintos horizontes e localidades ([INMAN; PEDRO; COIMBRA, 2013](#); [DIAGNE et al., 2013](#); [ANTONANZAS et al., 2016](#)).

¹ *Levelized Cost of Electricity*: proporção entre produção estimada e custo de manutenção e combustível, ao longo da vida útil de uma usina.

1.3 Hipótese e Objetivos

O intuito desta dissertação é examinar o desempenho de Redes Neurais Recorrentes em previsões de radiação solar. Estas arquiteturas diferem de outras Redes Neurais Artificiais pela capacidade de armazenar informações de exemplos anteriores, reproduzindo as correlações entre os dados ao longo do tempo (PASCANU; MIKOLOV; BENGIO, 2013). Entretanto, essa arquitetura de redes neurais é raramente aplicada à previsão de séries temporais de produção fotovoltaica ou radiação solar.

Os modelos são treinados e avaliados utilizando dados provenientes das cidades de Denver e Las Vegas nos Estados Unidos, re-amostrados em resoluções de 5 minutos, 30 minutos, 1 hora e média diária. Neste sentido, combinou-se as configurações supramencionadas como modelos de previsão baseados em Rede Neural Recorrente (RNN), *Extreme Gradient Boosting* (XGBoost) e à persistência dos índices de Céu Claro e de Claridade. Também foi examinado o efeito da seleção dos atributos mais informativos, bem como a inclusão de estimativas de cobertura do céu obtidas de Previsões Numéricas do Clima (NWP) e dos índices de Céu Claro e de Claridade. Com estes experimentos pretende-se verificar as seguintes hipóteses:

1. Redes Neurais Recorrentes devem estimar as flutuações da radiação solar incidente com precisão comparável a de outros modelos e experimentos na literatura;
2. A seleção de subconjuntos pode revelar atributos mais informativos, permitindo utilizar conjuntos menores sem reduzir a precisão e apoiando a exploração do problema.

Após realizar a limpeza de dados (*data cleaning*), anexar índices de Céu Claro e de Claridade e reamostrar os conjuntos de dados, estes são submetidos à seleção de atributos, realizada pela combinação do modelo XGBoost a um algoritmo de Eliminação Recursiva de Atributos (RFE). Na sequência, os subconjuntos de atributos foram submetidos como entradas aos modelos de previsão, a saber: RNN e o próprio XGBoost. Os resultados obtidos por estes modelos serão comparados aos modelos de persistência dos índices de Céu Claro e de Claridade.

1.4 Organização da Dissertação

O Capítulo 2 apresenta os conceitos fundamentais relacionados à Radiação Solar, como: composição da Radiação Solar Global Horizontal, previsões NWP e índices de Céu Claro e Claridade. No Capítulo 3 é realizada a revisão da literatura, no tocante aos modelos de predição baseados em aprendizado de máquina. Na sequência, o Capítulo 4 descreve a preparação dos dados e os experimentos, seguidos pela discussão dos resultados. Por fim, o Capítulo 5 destina-se a apresentar as conclusões da pesquisa.

2 Fundamentos sobre Radiação Solar

Este capítulo apresenta conceitos envolvidos na previsão de radiação solar e produção de energia fotovoltaica. Descreve inicialmente os componentes da Irradiância Global Horizontal, que contém tanto a radiação que atinge diretamente a superfície de um painel fotovoltaico, como a parcela espalhada pela atmosfera. Descreve, em seguida, modelos e índices de Céu Claro e de claridade, assim como o processo que resulta nestes índices. E, por fim, apresenta [Previsões Numéricas do Clima \(NWP\)](#) elaboradas por redes globais de observação meteorológica.

2.1 Componentes da Radiação Solar Incidente

Medições de satélites indicam que 1360 W/m^2 ([KOPP; LEAN, 2011](#)) de energia em forma de radiação solar alcançam a atmosfera a todo instante ([INMAN; PEDRO; COIMBRA, 2013](#)). Enquanto atravessa a atmosfera, parte desta radiação é refletida de volta ao espaço ou se dispersa pelas partículas suspensas no ar. Portanto, a radiação disponível para produção de energia é determinada pela união de dois componentes: [Irradiância Normal Direta \(DNI\)](#) que corresponde aos raios que atingem perpendicularmente uma superfície a nível do solo; e [Irradiância Horizontal Difusa \(DHI\)](#) que descreve a radiação dispersa na atmosfera que ainda alcança esta superfície. A combinação destas é dada pela Equação 2.1, sendo θ o ângulo de zênite:

$$GHI = DHI + DNI \cdot \cos \theta \quad (2.1)$$

Métodos preditivos de irradiância frequentemente utilizam [Irradiância Global Horizontal \(GHI\)](#) como objetivo, por ser um indicador direto do potencial de geração de energia. As irradiâncias direta e difusa, por sua vez, assim como o ângulo de zênite, podem ser usados como atributos informativos ([INMAN; PEDRO; COIMBRA, 2013](#)).

2.2 Modelos e Índices de Céu Claro e Claridade

Índices de Céu Claro e Claridade descrevem a proporção entre a expectativa de irradiância máxima e o valor observado a nível do solo em determinado momento. Assim, expressam a atenuação imposta pelas condições climáticas a cada observação. O que diferencia estes índices é o método empregado para estimar a irradiância máxima: modelos de irradiância em céu claro ou a irradiância extraterrestre. Em virtude destes índices

acompanharem os ciclos diário e anuais do recurso solar, são utilizados para remover a sazonalidade de séries temporais de irradiância, como também na normalização dos dados.

2.2.1 Modelos e Índice de Céu Claro

Uma vez que a maioria dos aspectos determinantes da produção fotovoltaica podem ser estimados com precisão – como por exemplo a área dos painéis, a eficiência de conversão e os movimentos da terra ao redor do sol – a variabilidade do recurso solar é causada principalmente pelas condições climáticas e atmosféricas (MARQUEZ; COIMBRA, 2013; PEDRO; COIMBRA, 2012). Por este motivo, Modelos de Céu Claro buscam determinar a influência das condições climáticas, pela análise de variáveis meteorológicas ou de séries temporais de produção fotovoltaica, elaborando estimativas que reproduzem condições ideais. Estas são combinadas às previsões de satélite para gerar novos conjuntos de dados ou preencher períodos ausentes, utilizadas como parâmetro de normalização de séries temporais de radiação conforme previamente mencionado, ou em Métodos de Persistência (INMAN; PEDRO; COIMBRA, 2013), descritos no Capítulo 3.

Tipicamente, a elaboração destes modelos utiliza variáveis meteorológicas combinadas com geometria solar e modelos de transmissividade atmosférica (RTM, do inglês *Radiative Transfer Models*) (ANTONANZAS et al., 2016). Neste grupo encontram-se o modelo de Bird e Hulstrom (1981) que examina a profundidade ótica dos aerossóis suspensos na atmosfera e concentrações de vapor de água e ozônio e o modelo *Solis* (MUELLER et al., 2004) que emprega RTM, funções de Beer-Lambert e integração por meio do espectro solar para obter os componentes da Irradiância Global Horizontal.

Além desta abordagem, Pedro e Coimbra (2012) propõem um modelo derivado unicamente das medições de produção de uma usina fotovoltaica, elaborado a partir de um mapeamento tridimensional do rendimento médio a cada hora e dia do ano, ao longo dos anos disponíveis. Então, uma superfície é ajustada ao relevo deste mapeamento, de maneira que seu formato reproduz a variação diária e anual da irradiância. Assim, as estimativas do modelo correspondem aos pontos nesta superfície.

Partindo de um determinado modelo de Céu Claro, o respectivo índice de Céu Claro (k_t) é obtido pela Equação 2.2, sendo I_t a irradiância observada e I_t^{CS} a estimativa em Céu Claro no instante t .

$$k_t = \frac{I_t}{I_t^{CS}} \quad (2.2)$$

2.2.2 Radiação Extraterrestre e Índice de claridade

A Radiação Extraterrestre descreve a radiação que atinge o topo da atmosfera em um dia e hora específicos. É determinada pela constante solar ($I_{CS} = 1360W/m^2$) em função do ângulo de zênite θ e do fator de correção da excentricidade da órbita da terra E_0 , como indicado pela Equação 2.3. Assim, descreve a variação causada pelas relações astronômicas, uma vez que E_0 corrige a distância solar ao longo do ano e o ângulo de zênite carrega informações sobre o ciclo diário e sobre a latitude onde é realizada a medição (IQBAL, 1983).

$$I_{t+1}^{EX} = I_{CS} \cdot E_0 \cdot \cos \theta_{t+1} \quad (2.3)$$

Portanto, a Radiação Extraterrestre é capaz de estabelecer a influência dos componentes determinísticos da irradiância, de modo que a variabilidade remanescente é causada exclusivamente pelas influências atmosféricas e meteorológicas. O índice de claridade (K_t), proporção entre a Radiação Extraterrestre (I^{EX}) e a Irradiância observada à nível do solo (I_t), equivale à variação causada por estes fatores não determinísticos da irradiância (TOVAR-PESCADOR, 2008), conforme expresso por meio da Equação 2.4.

$$K_t = \frac{I}{I_t^{EX}} \quad (2.4)$$

Calcular o fator de correção da excentricidade (E_0) e o ângulo de zênite (θ_Z) requer apenas variáveis astronômicas. E_0 expressa a variação entre a distância média da terra ao sol (r_0) e a distância presente, determinado pela Equação 2.5:

$$E_0 = (r_0/r)^2 = 1.000110 + 0.034221 \cos \Gamma + 0.001280 \sin \Gamma + 0.000719 \cos 2\Gamma + 0.000077 \sin 2\Gamma \quad (2.5)$$

em que Γ representa o ângulo diário, sendo d_n o n -ésimo dia, expresso por:

$$\Gamma = 2\pi(d_n - 1)/365 \quad (2.6)$$

O Ângulo de Zênite, por sua vez, é uma variável presente em diversos conjuntos de dados. Embora possa ser determinado em função da latitude, horário e dia do ano (IQBAL, 1983).

2.3 Previsão Numérica do Clima

Previsões Numéricas do Clima (NWP) são produzidas por modelos matemáticos dos fenômenos atmosféricos, alimentados por uma rede global de observação meteorológica. Estes modelos partem do pressuposto que estimativas precisas dependem apenas de

informação suficiente sobre as condições iniciais e sobre as leis da física que governam o clima (INMAN; PEDRO; COIMBRA, 2013).

Modelos NWP dividem-se, de acordo com a extensão geográfica e resolução espacial das estimativas, em Globais ou de Mesoescala. A elaboração destes modelos se inicia pela divisão do seu domínio espacial em uma matriz. Cada posição desta matriz armazena a condição climática, discretizada em um conjunto de variáveis, de uma região que pode variar de $3 \times 3km$ (HRRR) a $28 \times 28km$ (*Global Forecast System*). Então, um sistema de equações é capaz de identificar e reproduzir fenômenos climáticos de grande escala, simultaneamente causadores das observações em pequena escala e compostos por elas (DIAGNE et al., 2013; INMAN; PEDRO; COIMBRA, 2013).

Devido à resolução das observações discretizadas ultrapassar dezenas de quilômetros, modelos NWP têm dificuldades em determinar precisamente a posição e o formato de formações de nuvens. Contudo, fornecem estimativas precisas dos padrões meteorológicos abrangentes, além de horizontes de previsão mais distantes e informações sobre diversos fenômenos climáticos distintos.

Além disso, estimativas destes modelos são regularmente publicados por órgãos como a NOAA (*US National Oceanic and Atmospheric Administration*). A agência disponibiliza previsões formuladas por modelos NWP¹ globais, como *Global Forecast System* com resolução espacial de $28 \times 28km$ e previsões a cada 6 horas até 16 dias a frente; e modelos de Mesoescala, como *Rapid Update Cycle (RUC)* que abrangem o território dos Estados Unidos, com resolução de $13 \times 13km$ e estimativas horárias até 18 horas a frente; e *High Resolution Rapid Refresh (HRRR)* que deriva previsões horárias com resolução de $3 \times 3km$ das estimativas do modelo RUC. (INMAN; PEDRO; COIMBRA, 2013).

¹ NOAA Numerical Weather Prediction: <https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/numerical-weather-prediction>

3 Revisão da Literatura

3.1 Seleção da Bibliografia e Metanálise

O processo de seleção de referências bibliográficas iniciou pela consulta da base de dados Scopus. Esta inclui mais de 23 mil jornais científicos e oferece consultas por *query* de busca, combinando palavras-chave, ou suprimindo artigos que contenham-nas em campos específicos como título e resumo. Após exportar os resultados de uma pesquisa inicial – contendo os termos: *solar, radiation, forecasting, forecast, fotovoltaic, power, energy e prediction* – empregou-se a ferramenta *VOS Viewer* (ECK; WALTMAN, 2014) para elaborar redes de citações e mapas de palavras-chave. Estes mapeamentos auxiliaram a exclusão de termos que não correspondem ao âmbito da pesquisa, como por exemplo *agriculture, drought, stars e heat flux*, em destaque na Figura 1.

As redes de citações, por sua vez, revelavam a existência de um conjunto predominante, fortemente conectado por suas referências, circundado por documentos isolados, apresentados na Figura 2a. No centro deste aglomerado encontram-se estudos sobre previsão de irradiância e produção fotovoltaica, como Mellit e Pavan (2010), Bacher, Madsen e Nielsen (2009) e Voyant et al. (2013), representados por esferas de maior diâmetro, proporcionais ao número de conexões com outros artigos da amostra. Assim, a presença dos artigos isolados demonstra que estes contêm os termos consultados, porém sua distância e separação sugere de que possuem objetivos distintos dos artigos centrais e, consequentemente, não correspondem ao tema desta pesquisa.

Refinando os termos de consulta e exclusão, foi possível limitar os resultados a 599 documentos. Estes foram então selecionados pela concordância de seus títulos e resumos

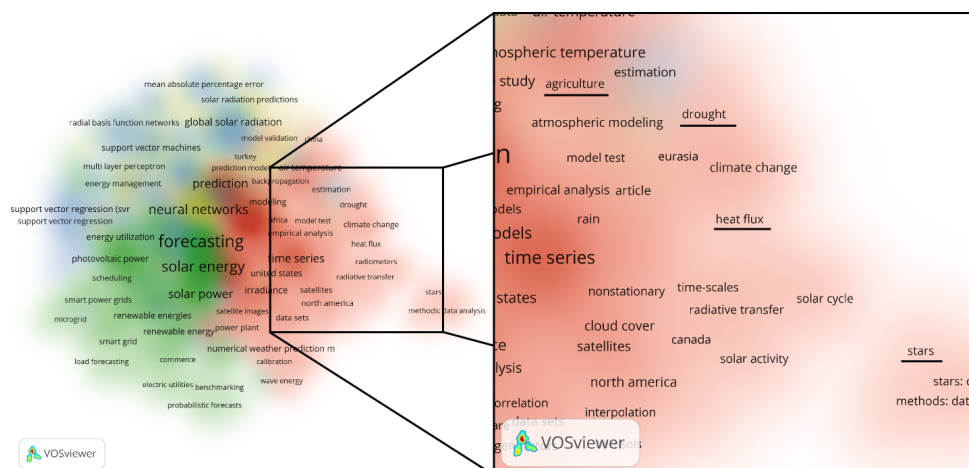


Figura 1 – Mapa de palavras-chave destacando termos suprimidos na *query* de busca.

com três critérios de inclusão, a saber: **(1)** O estudo descreve um método preditivo da radiação solar ou produção fotovoltaica; **(2)** apresenta uma revisão da literatura em previsão de radiação solar ou produção fotovoltaica; **(3)** contribui com novos métodos, informações sobre o problema ou fornece novos conjuntos de dados. Após esse novo refinamento, resultou deste processo um conjunto de 307 documentos sendo, 239 deles correspondentes a experimentos preditivos, 14 revisões da literatura e 52 contribuições aos objetivos da pesquisa. A exclusão de artigos não relacionados pode ser observada na rede de citações da Figura 2b, com número reduzido de resultados isolados e conexões distantes.

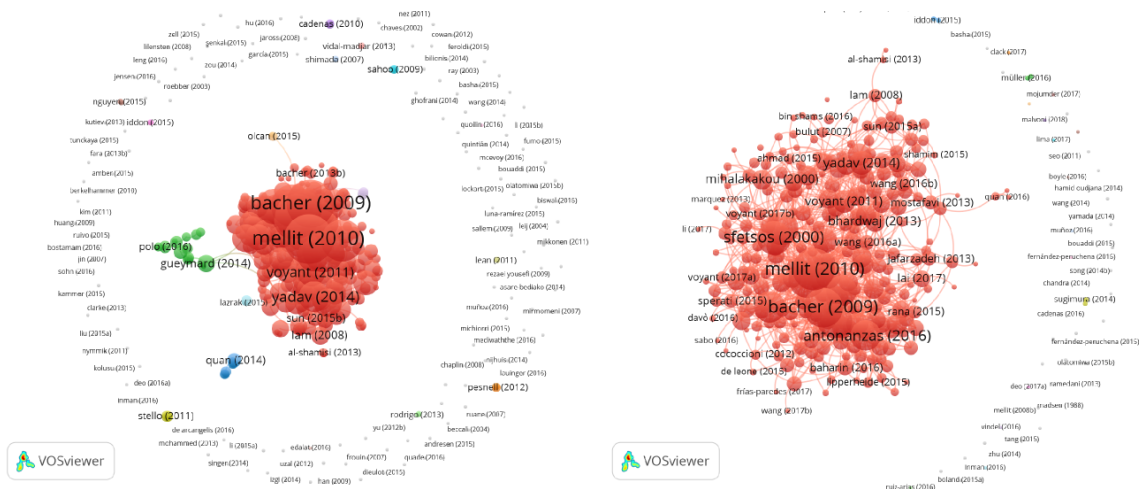


Figura 2 – Redes de Citações, antes e depois da seleção por título e resumo.

Após selecionar artigos que efetivamente compõem a amostra da bibliografia em estimativas de irradiância e produção fotovoltaica, a metanálise deste conjunto foi realizada por meio da ferramenta *Bibliometrix* (ARIA; CUCCURULLO, 2017), a qual revela autores, periódicos e demais documentos de maior relevância, bem como estatísticas de crescimento do número de publicações e palavras-chave específicas. Com base nessa análise pode-se verificar que publicações na área cresceram 48.4% entre 2010 e 2016, e que a 4ª palavra-chave mais mencionada é *neural networks*, superada por *solar radiation*, *forecasting* e *solar energy*. Além disso, a ferramenta determina os 10 documentos mais citados, apresentados na Tabela 1.

Embora não tenha sido possível extrair todos os documentos, foi selecionado um subconjunto priorizando artigos mais relevantes de acordo com a análise bibliométrica e que empregam redes neurais artificiais em previsões no curto prazo.

Tabela 1 – Lista dos 10 documentos mais citados.

Referência	Citações
Mellit e Pavan (2010)	317
Bacher, Madsen e Nielsen (2009)	313
Inman, Pedro e Coimbra (2013)	261
Reikard (2009)	221
Sfetsos e Coonick (2000)	204
Chen et al. (2011)	190
Diagne et al. (2013)	176
Pedro e Coimbra (2012)	174
Paoli et al. (2010)	157
Yadav e Chandel (2014)	139

3.2 Modelos de Previsão de Irradiância ou Produção Fotovoltaica

Motivados pela operação eficiente dos sistemas elétricos de potência, alguns grupos de pesquisa têm se dedicado à previsão da radiação solar e produção fotovoltaica para distintos horizontes, localidades e extensões geográficas. Estes experimentos podem ser divididos em três categorias: **(1)** métodos de Persistência presumem que parte das condições se sustenta entre a última leitura e o horizonte de previsão; **(2)** métodos de Análise Probabilística de Séries Temporais realizam previsões por meio da análise da correlação às leituras anteriores; e **(3)** técnicas de Aprendizado de Máquina, como Redes Neurais, produzem estimativas de acordo com os relacionamentos inferidos dos exemplos de treinamento.

3.2.1 Modelos de Persistência

Modelos de Persistência apoiam-se na suposição de que, entre o momento da leitura e o horizonte de previsão, parte das condições climáticas, atmosféricas e de irradiância se mantêm inalteradas. Em sua forma mais simples, modelos de *Naive Persistence* produzem estimativas considerando que a irradiância, ou qualquer que seja a variável dependente, sustenta exatamente os mesmos valores entre o instante da estimativa e o horizonte de previsão. Conseqüentemente, agem como indicadores da relação entre complexidade do modelo e precisão das estimativas, de modo que técnicas mais sofisticadas justificam-se somente caso demonstrem maior precisão (INMAN; PEDRO; COIMBRA, 2013).

Abordagens de *Smart Persistence* distinguem-se por empregar variáveis que acompanham os ciclos solares ou agregam informações meteorológicas, como índices de Claridade e Céu Claro. Estimativas de persistência do índice de Céu Claro, por exemplo, correspondem à expectativa de irradiância máxima no horizonte de previsão, atenuada

pelas condições atmosféricas na última leitura, como descrito pelas Equações 3.1 e 3.2:

$$k_t = \frac{I_t}{I_t^{CS}} \quad (3.1)$$

$$\hat{I}_{t+h} = I_{t+h}^{CS} \cdot k_t \quad (3.2)$$

sendo \hat{I}_{t+h} a estimativa da irradiância no horizonte h , I_{t+h}^{CS} é a expectativa de irradiância em céu claro e k_t é o índice de Céu Claro na última observação. A Persistência do índice de Claridade, por sua vez, substitui a expectativa em céu claro (I_{t+h}^{CS}) por Irradiância Extra-terrestre (I_{t+h}^{EX}) e índice de Céu Claro (k_t) por índice de Claridade (K_t) (ANTONANZAS et al., 2016). Avaliando a precisão destes modelos em observações horárias de irradiância, Inman, Pedro e Coimbra (2013) constataram erro absoluto (e_{abs}) inferior a 5% em dias predominantemente ensolarados, variando entre 20% e 60% em dias nublados. Considerando estes resultados, em conjunto com os gráficos que retratam a correspondência entre a irradiância prevista e observada, os autores deduzem que rápidas flutuações causadas pela passagem de nuvens são responsáveis pela maior parte dos erros destes modelos.

Entretanto, experimentos como os apresentados por Bacher, Madsen e Nielsen (2009), Pedro e Coimbra (2012) e Lipperheide, Bosch e Kleissl (2015), sugerem aprimoramentos aos modelos de persistência. Bacher, Madsen e Nielsen (2009) elaboraram um mapeamento da produção fotovoltaica máxima a cada dia do ano e hora do dia $f_{max}(x, y)$. Então, a superfície deste mapeamento é suavizada, ponderando as estimativas em Céu Claro $k_{(x,y)}$ em proporção aos pesos determinados pela regressão quantílica das leituras adjacentes em uma Função de Kernel de Base Radial (RBF) (gaussiana). Este processo resulta em um modelo do rendimento máximo da usina, que é robusto à presença de *outliers*. Os autores ainda utilizaram como parâmetro de normalização dos dados e na elaboração de estimativas a persistência da última observação diurna e da média diária dos índices de Céu Claro. A precisão destes modelos, avaliada em observações de 21 sistemas fotovoltaicos na Dinamarca durante o ano 2006, alcança Raiz do Erro Médio Quadrático Normalizada (nRMSE) (normalizado por rendimento médio 248W/h) entre 50% e 200% em horizontes de até 6 horas e entre 100% e 250% em horizontes de 18 a 30 horas à frente.

Pedro e Coimbra (2012) também derivam um modelo de Céu Claro da produção fotovoltaica mapeada a cada dia e hora. Porém, ao invés de regularizar este mapeamento, utilizam uma superfície poligonal contínua, observada na Figura 3. Desta forma, estimativas em Céu Claro podem ser obtidas acessando a posição nesta superfície que corresponde ao instante desejado. Previsões horárias destes modelos, em dados provenientes de uma usina de 1 MW na Califórnia, entre novembro de 2009 e agosto de 2011, obtiveram nRMSE

(determinado pela Equação 3.3) de 0.19.

$$nRMSE = \sqrt{\frac{\sum_{t=1}^m (P_t - \hat{P}_t)^2}{\sum_{t=1}^m P_t^2}} \quad (3.3)$$

sendo \hat{P}_t a estimativa e P_t a produção fotovoltaica no momento t .

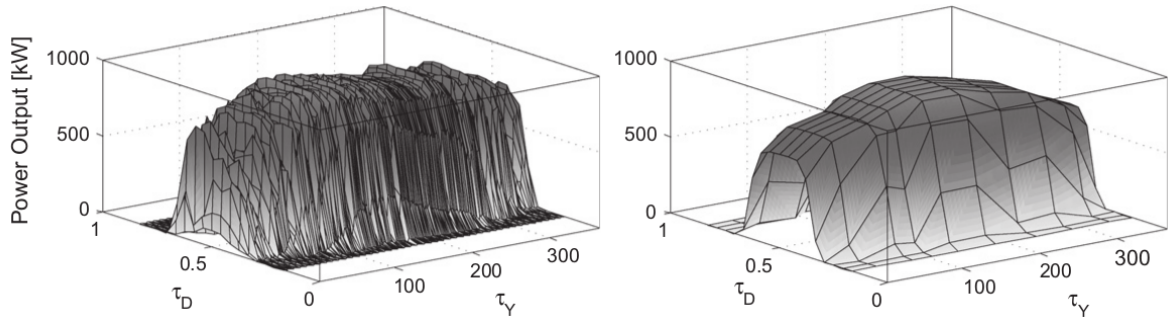


Figura 3 – Modelo de Céu Claro desenvolvido por [Pedro e Coimbra \(2012\)](#), cuja superfície corresponde à expectativa de produção a cada dia do ano τ_Y e hora do dia τ_D .

Por fim, o estudo de [Lipperheide, Bosch e Kleissl \(2015\)](#) examina a Persistência de Velocidade das Nuvens em previsões instantâneas (segundo a segundo) e horizontes de até 3 minutos de uma usina fotovoltaica de 48 MW em Boulder City, Nevada. Este método produz mapas da potência de cada inversor fotovoltaico e os desloca na direção e velocidade da movimentação das nuvens. A produção nas células cobertas pelo mapa é calculada presumindo que a potência deslocada se mantém. Enquanto os painéis descobertos têm seu rendimento calculado por *Ramp Persistence*, em que a variação de rendimento observada no último segundo é estendida até o horizonte de previsão, ou seja, $\hat{P}_{t+1} = P_t + \Delta[P_t - P_{t-1}]$. Além destes, adota como critério de comparação a persistência da variação da irradiância na área dos painéis, que escala o rendimento atual pela diferença entre irradiância presente e estimada por um modelo de Céu Claro de [Ineichen e Perez \(2002\)](#). Em horizontes de 20 e 60 segundos, a Persistência de Velocidade das Nuvens apresenta *nRMSE* de 2.7% e 7.3%, superando o modelo de Céu Claro por 16.2% e 10.6%, respectivamente.

3.2.2 Modelos Probabilísticos de Séries Temporais

Modelos Auto-Regressivos (*AR*), Auto-Regressivos de Médias Móveis (*ARMA*) e Auto-Regressivos Integrados de Médias Móveis (*ARIMA*) são exemplos de Modelos Probabilísticos para Análise de Séries Temporais. Estes métodos buscam explorar a dependência temporal entre observações em séries estacionárias – que demonstram forte correlação entre leituras sucessivas e comportamento semelhante independente do período examinado –

como a irradiância solar (INMAN; PEDRO; COIMBRA, 2013). Modelos auto-regressivos de m -ésima ordem, como o apresentado na Equação 3.4:

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_m z_{t-m} + \epsilon_t \quad (3.4)$$

descrevem a observação atual (z_t) por meio da combinação linear das observações anteriores ($z_{t-1} \dots z_{t-m}$) aos parâmetros ($\phi_1 \dots \phi_m$) que caracterizam sua correlação, acompanhada de um termo de erro (ϵ_t) aleatório e independente, com média 0 e variância constante, frequentemente referido como processo de ruído branco (*white noise*) ou puramente randômico. Os parâmetros ϕ , por sua vez, são ajustados pela análise das funções de autocorrelação (*ACF*) e autocorrelação parcial (*PACF*) ou por algoritmos como *Recursive Least Squares (RLS)*.

Em Yule et al. (1927), os autores empregam como analogia do funcionamento destes modelos a movimentação harmônica de um pêndulo perturbado por ervilhas lançadas em sua direção. Nesta demonstração, são feitos z registros da posição do pêndulo e o termo de erro (ϵ) resulta das ervilhas que eventualmente o atingem. A análise da movimentação do pêndulo revela características como peso, comprimento do cabo e densidade do meio em que se encontra. Similarmente, a modelagem de séries temporais aproximadamente estacionárias – como a temperatura de fornalhas industriais, cotações no mercado de ações e ciclos da radiação solar – auxilia a compreensão dos sistemas que compõem estes fenômenos (BISGAARD; KULAHCI, 2011).

Experimentos utilizando modelos *AR* para previsão de irradiância e produção fotovoltaica incluem o estudo dirigido por Boland (2008). Este corrobora a suposição de que a radiação solar pode ser dividida em dois componentes: um determinístico, caracterizado pelos movimentos astronômicos; e outro estocástico que resulta das flutuações impostas pelas condições atmosféricas. Portanto, os ciclos diários e anual são determinados pela Transformada de Fourier e sua influência subtraída da variação dos dados, produzindo uma série residual que encerra o componente estocástico. Então, um modelo auto-regressivo de primeira ordem *AR(1)* é utilizado para reproduzir esta série residual, alcançando nRMSE 23.3% em horizontes de 1 hora.

O estudo de Bacher, Madsen e Nielsen (2009) compara modelos *AR*, *ARX* (auto-regressivo com entradas exógenas) e Linear em estimativas de produção de 21 sistemas fotovoltaicos, na Dinamarca, durante o ano 2006. Os conjuntos de dados utilizados compreendem leituras do rendimento ou estimativas derivadas de previsões *NWP* e são normalizados em proporção do índice de Céu Claro descrito na seção anterior. O modelo *ARX* é formulado acrescentando ao modelo *AR* da produção fotovoltaica (p) um termo

referente à previsão derivada de [NWP](#):

$$\hat{p}_{t+h} = m + \phi_1 p_t + \phi_2 p_{t-s} + \beta_1 \hat{p}_{t+h}^{NWP} + \epsilon_{t+h} \quad (3.5)$$

assumindo \hat{p}_{t+h}^{NWP} como estimativa em função de *NWP* no instante $t + h$, p_t e p_{t-s} observações anteriores relevantes e $m, \phi_1, \phi_2, \beta_1$ parâmetros ajustados à série por um filtro adaptativo de Mínimos Quadrados Recursivos (*RLS*). Assim, alcança os melhores resultados entre os três modelos: *nRMSE* (RMSE normalizado por rendimento médio $248W/h$) entre 50% e 100% em horizontes até 6 horas, e entre 50% e 150% em horizontes de 18 à 30 horas à frente.

Modelos *ARMA* combinam modelos de Médias Móveis (MA, do inglês *Moving Average*) aos modelos auto-regressivos. Assim como processos *AR* reproduzem o componente determinístico da irradiância – a dependência temporal entre as observações anteriores – modelos *MA*(n) examinam seu componente estocástico por meio da combinação linear do ruído (ϵ_{t-i} $i = 1, 2, \dots, n$) aos parâmetros (θ_i $i = 1, 2, \dots, n$), como na Equação 3.6 em que $w_t = \nabla z_t = z_t - z_{t-1}$, ou seja, a diferença entre cada observação e aquela que a antecede.

$$w_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_n \epsilon_{t-n} \quad (3.6)$$

Então, o processo *ARMA*(p, q) resulta da união dos modelos *AR*(p) e *MA*(q), como descrito pela Equação 3.7 ([INMAN; PEDRO; COIMBRA, 2013](#); [BISGAARD; KULAHCI, 2011](#)).

$$w_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} \quad (3.7)$$

O experimento dirigido por [David et al. \(2016\)](#) aplica modelos *ARMA* e *GARCH* (do inglês, *generalized autoregressive conditional heteroskedasticity*) para estimar irradiância global horizontal em frequências de 10 minutos a 1 hora e dados provenientes de 6 localizações. Após normalizar os dados em razão do índice de Céu Claro, removendo os ciclos diário e anual, os autores aplicam três métodos preditivos: um *ensemble* de modelos de persistência; um modelo *ARMA* combinado a um algoritmo *RLS* (*Recursive Least Squares*) que determina recursivamente os coeficientes de regressão; e um modelo *GARCH* que associa faixas de confiança às previsões do modelo *ARMA*, de acordo com a variabilidade dos erros anteriores, produzindo estimativas probabilísticas. Comparando os resultados médios entre todas as localidades, os modelos de persistência obtém *nRMSE* entre 22.3 e 32.5% e *rMAE* (em proporção da irradiância média global) entre 12 e 20.9% para horizontes de 10 a 60 minutos, enquanto o modelo *ARMA* recursivo alcança *nRMSE* entre 20.8 e 30.3% e *rMAE* entre 12.5 e 21.4% nos mesmos horizontes. As previsões do modelo

GARCH são avaliadas por CPRS (*Continuous Ranked Probability Score*), que considera a precisão das estimativas pontuais e a confiabilidade dos intervalos, com resultados entre 9.5 e 15.2%.

Processos $ARIMA(p, d, q)$ aplicam processos $ARMA(p, q)$ à d -ésima subtração dos dados: $\nabla^d z_t = (z_t - z_{t-1}) - (z_{t-1} - z_{t-2}) - \dots - (z_{t-d+1} - z_{t-d})$. Desta forma, são capazes de modelar séries temporais não-estacionárias pois, embora as leituras não apresentem expressiva correlação entre si, cada grau de subtração resulta em uma série com menor variabilidade, revelando um processo estacionário subjacente (BISGAARD; KULAHCI, 2011).

Reikard (2009) emprega um modelo $ARIMA(1, 0, 0)(1, 1, 0)$ composto de dois termos de regressão: da última leitura de radiação solar i_{t-1} ; e da subtração da leitura atual pelo mesmo horário no dia anterior $i_t - i_{t-24}$, caracterizando a variação dos ciclos diários. Este é comparado aos modelos AR , *Unobserved Components Model* (que considera as relações trigonométricas do ciclo diário) e uma Rede Neural Artificial treinada por *Backpropagation*. Conjuntos de dados de seis localidades diferentes – três coletados de 1987 à 1990 e três entre 2007 e 2008 – normalizados por logaritmo natural $\ln(i_t)$, são utilizados na avaliação dos modelos. Os resultados do modelo ARIMA superam os outros em horizontes de 15 e 30 minutos e de 1 a 4 horas, alcançando Erro Percentual Absoluto Médio (MAPE) entre 29.9% em Phoenix e 41.1% em Denver. Contudo, em horizontes de 5 minutos é superado pelo modelo AR, com MAPE de 13%.

3.2.3 Técnicas de Aprendizado de Máquina

Redes Neurais Artificiais (RNAs) são sistemas de reconhecimento de padrões inspirados nos neurônios e sinapses do cérebro humano (INMAN; PEDRO; COIMBRA, 2013). São compostas por unidades simples de processamento, organizadas em arquiteturas especializadas e submetidas a um processo de treinamento. Durante este processo, ajustam a força das sinapses para minimizar o erro de suas respostas aos exemplos apresentados, aperfeiçoando sua habilidade de identificar padrões e mapear os relacionamentos entre os dados (PAOLI et al., 2010).

Um neurônio artificial k , representado pela Figura 4, combina os sinais de entrada $x_{1\dots m}$ aos pesos sinápticos ajustados durante o treinamento $w_{k1\dots km}$ e um termo de *bias* b_k , por meio da Equação 3.8.

$$v_k = \sum_{j=1}^m w_{kj} x_j + b_k \quad (3.8)$$

O potencial de ativação v_k do neurônio é transferido para o domínio de uma função de ativação $\varphi(\cdot)$ que pode apresentar o formato de uma curva suave, como funções

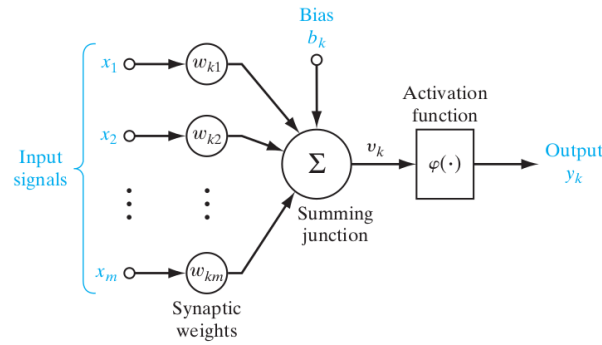


Figura 4 – Representação de um Neurônio Artificial. Fonte: Haykin (2009).

sigmóide, ou restringir os resultados a valores específicos, como funções de limiar (*Threshold*), apresentadas na Figura 5. Por fim, a resposta y_k do neurônio é transmitida à camada subsequente na arquitetura.

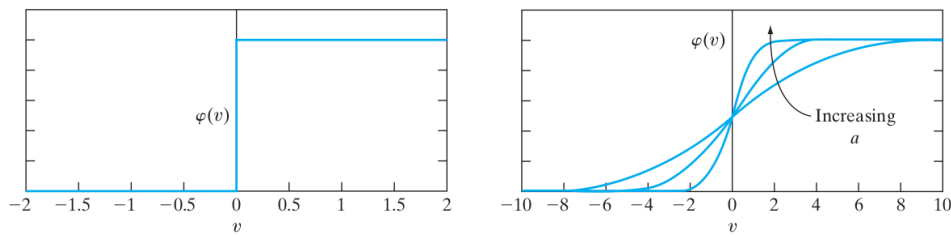


Figura 5 – Funções de ativação Limiar (*Threshold*) e Sigmóide. Fonte: Haykin (2009).

Individualmente, neurônios artificiais apresentam capacidade limitada de identificar relacionamentos. Por exemplo, considere reproduzir o comportamento de um operador **OU-exclusivo (XOR)** que aceita tuplas binárias (x_1, x_2) com valores diferentes. Um neurônio isolado – que combina dois sinais de entrada e um termo de *bias* aos respectivos pesos sinápticos, classificando-os de acordo com uma função de limiar (*threshold*) – estabelece uma fronteira de decisão linear capaz de separar exemplos em classes distintas \mathcal{C}_1 e \mathcal{C}_2 , como os neurônios N_1 e N_2 da Figura 6. Contudo, as regiões de decisão do operador **XOR** não são linearmente separáveis e, conseqüentemente, simular seu comportamento requer um terceiro neurônio N_3 unindo os resultados dos neurônios anteriores (HAYKIN, 2009).

A combinação de neurônios para simular o operador **XOR** evidencia um aspecto determinante das **ANNs**: neurônios simples, organizados em estruturas especializadas, adquirem a capacidade de aprender problemas complexos e generalizar os padrões identificados às novas observações (HAYKIN, 2009). Arquiteturas de Redes Neurais consolidadas incluem *Multilayer Perceptrons (MLPs)*, formados por camadas de neurônios adaptados para detectar características específicas dos exemplos; **Redes Neurais Convolucionais (CNNs)** que esquadriham regiões cada vez menores de uma imagem, reconhecendo formas e objetos; e **RNNs** capazes de armazenar informações sobre exemplos anteriores e identificar relacionamentos ao longo do tempo.

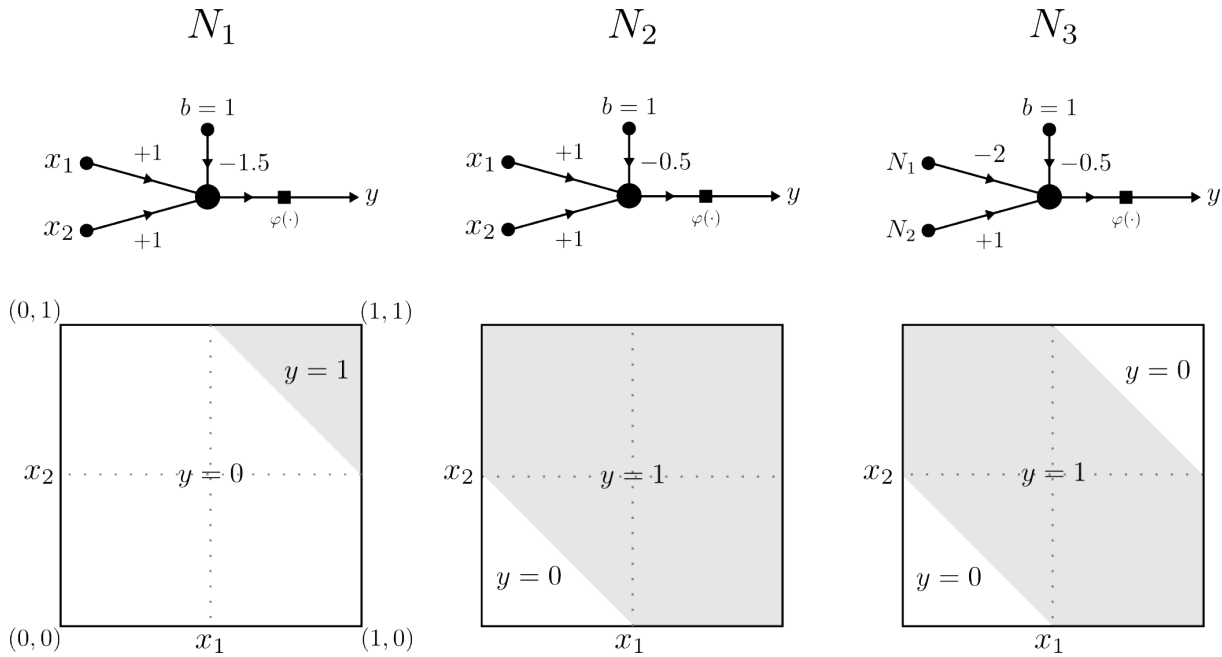


Figura 6 – Diagramas e regiões de decisão dos três neurônios artificiais *perceptron* combinados para simular o operador **XOR**.

Na sequência, são apresentadas breves descrições de **MLPs** e **RNNs**, acompanhadas de experimentos aplicando cada arquitetura à previsão de radiação solar ou produção de energia fotovoltaica. Ao final desta seção também é descrito o modelo **XGBoost**, utilizado na seleção de atributos e, posteriormente, comparado aos outros métodos preditivos durante os experimentos.

3.2.3.1 Multilayer Perceptron (MLP)

MLPs são arquiteturas de Redes Neurais *Feedforward*, organizadas em camadas e conectadas de modo que o resultado de cada neurônio seja transmitido à camada seguinte, como na Figura 7. Conseqüentemente, a hipótese sobre determinado exemplo resulta da contribuição de diversos neurônios que, por meio do treinamento, transformam-se em detectores das características marcantes dos dados (HAYKIN, 2009).

O treinamento desta arquitetura é dividido em duas etapas: durante o passo *forward*, um exemplo é apresentado à rede neural, o sinal de entrada é transmitido e modificado pelos neurônios ocultos e a resposta é consolidada pela última camada; então, no passo *backward*, os pesos sinápticos de cada neurônio são ajustados, da última à primeira camada, pelo algoritmo de treinamento, convencionalmente empregado o algoritmo *Backpropagation*. Este algoritmo encontra o gradiente descendente do vetor de pesos sinápticos, ou a direção de mudança dos pesos que reduz o erro, em respeito à contribuição do neurônio ao erro total da rede. Na primeira iteração, o erro individual de cada neurônio na última camada é determinado em proporção ao erro da estimativa, ajustando o vetor de pesos sinápticos. Então, o gradiente é transmitido à penúltima camada, calculada a

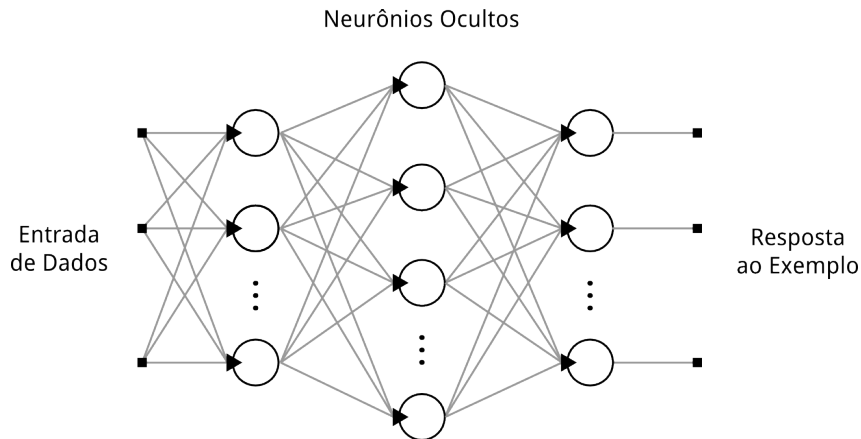


Figura 7 – Arquitetura de um Multilayer Perceptron.

contribuição dos neurônios desta camada ao erro da camada que os sucede e seus pesos sinápticos corrigidos proporcionalmente, antes de transmitir o erro desta camada para a antepenúltima e assim por diante, até chegar à primeira (HAYKIN, 2009).

Experimentos avaliando o desempenho desta arquitetura em previsões de irradiância incluem o estudo dirigido por Voyant et al. (2013), que compara o desempenho de modelos ARMA, MLP e híbrido MLP+ARMA. Também avalia o efeito da normalização por índices de Céu Claro e de claridade, combinados a médias móveis, com o intuito de remover a sazonalidade dos dados. Argumentando que uma metodologia uniforme de avaliação é necessária para comparar experimentos distintos, os autores propõem: testar os modelos em longos períodos e localizações distintas, utilizar Raiz do Erro Médio Quadrático (RMSE) e nRMSE como indicadores de precisão, analisar a taxa de erro a cada estação do ano e comparar os resultados com modelos de persistência, como exemplares da proporção entre precisão e complexidade dos modelos. Os melhores resultados apresentados são de nRMSE 23.4% da MLP aplicada à média diária, 27.3% da MLP em estimativas da irradiância 24 horas a frente, 14.9% do modelo MLP+ARMA em horizontes de 1 hora e 20.2% da MLP em estimativas de 5 minutos a frente.

O estudo de Mellit e Pavan (2010) emprega uma MLP treinada pelo Algoritmo de Levenberg Marquardt (LM) em estimativas da irradiância incidente, transformadas em produção fotovoltaica, nas próximas 24 horas. Este algoritmo de aprendizado difere do *Backpropagation* pois aproxima o erro dos neurônios com uma expressão de segunda ordem e mantém alterações do vetor de pesos somente caso reduzam o erro do modelo. São apresentados conjuntos de exemplos contendo data, hora e médias diárias de irradiância e temperatura do ar, entre julho de 2008 e maio de 2009. Estimativas de produção fotovoltaica são produto das previsões de irradiância do modelo MLP-LM, área dos painéis e de coeficientes representando eficiência de conversão de energia e balanço do sistema. A validação dos modelos é realizada somente em 4 dias selecionados, apresentando nRMSE entre 32.98% e 75.40%.

Redes Neurais Artificiais também foram empregadas na análise comparativa de [Pedro e Coimbra \(2012\)](#), assim como métodos de persistência, modelo ARIMA, k-NN (do inglês *k-nearest neighbors*) e ANN otimizada por algoritmos genéticos. Utilizando somente informações locais, os modelos produzem estimativas da geração de energia elétrica de uma usina na Califórnia, em horizontes 1 e 2 horas a frente. Para remover a sazonalidade dos dados sem recorrer às informações meteorológicas, foi elaborado um mapa tridimensional da produção ao longo dos anos, de maneira que sua superfície representa a expectativa de produção média a cada instante. Normalizar as leituras em razão desta expectativa remove os ciclos determinísticos anual e diário da radiação solar, restando somente a variação causada pelas condições meteorológicas. Ao final dos testes, os resultados em horizontes de 1 e 2 horas foram: modelos de persistência alcançaram nRMSE de 19,27% e 28,86%, o modelo ARIMA obteve 18,95% e 25,89%, o modelo k-NN 20,90% e 29,14%, a RNA treinada com o LM atingiu nRMSE de 15,82% e 25,61%, e a RNA otimizada com algoritmo genético obteve os melhores resultados com 13,07% e 18,71% ([PEDRO; COIMBRA, 2012](#)).

3.2.3.2 Redes Neurais Recorrentes (RNN)

Redes Neurais Recorrentes são arquiteturas de ANNs que apresentam conexões sinápticas entre neurônios ocultos submetidas a um atraso temporal, como ilustrado pela Figura 8. Estas conexões buscam representar implicitamente a estrutura temporal das observações, por meio de mudanças de estado (matriz de pesos sinápticos) da rede e do *feedback* (retroalimentação) dos estados ou respostas anteriores, como uma memória de curto prazo. Conseqüentemente, durante o treinamento, além de elaborar representações que traduzam os padrões de entrada em respostas adequadas, estes neurônios devem produzir mapeamentos das relações temporais através da sequência de exemplos ([ELMAN, 1990](#)).

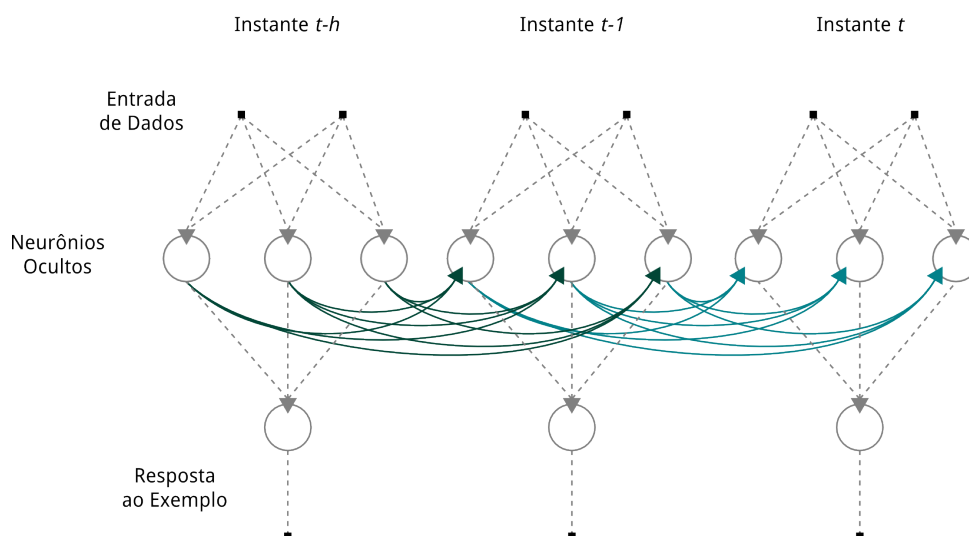


Figura 8 – Conexões entre neurônios ocultos em arquiteturas de RNNs.

Identificar relacionamentos temporais requer que o aprendizado ajuste os pesos sinápticos não somente em proporção ao erro atual, como também ao erro nas estimativas anteriores. Com este propósito, o algoritmo *Truncated Backpropagation Through Time* ($BPTT(h)$) estende conexões recorrentes, calculando o *backpropagation* de h estados e respostas anteriores destes neurônios como se formassem uma rede com h camadas sequenciais (em outras palavras, na direção contrária das conexões destacadas na Figura 8). O ajuste da matriz de pesos atual resulta, portanto, da combinação das correções determinadas para h estados anteriores (HAYKIN, 2009).

Neste aspecto, RNNs assemelham-se às Redes Neurais Profundas, pois seu aprendizado exige propagar gradientes pelas numerosas camadas. Característica que também dá origem aos problemas denominados: *vanishing gradients*, ou a progressiva diminuição dos gradientes, até que não produzam alterações significativas; e o cenário oposto, *exploding gradients*, em que os gradientes crescem exponencialmente a cada iteração (PASCANU; MIKOLOV; BENGIO, 2013).

Três experimentos que aplicam RNNs à previsão de radiação solar foram identificados na amostra selecionada da literatura, a saber: o modelo *NARX* apresentado em Ahmad, Anderson e Lie (2015), *Elman Networks* avaliadas por Sfetsos e Coonick (2000) e o estudo de Cao e Cao (2005), que aplica transformada de *wavelets* às séries temporais de radiação solar para gerar os atributos apresentados às RNNs.

Modelos *NARX* (do inglês, *Nonlinear Auto-Regressive with eXogenous inputs*) assim como os modelos *ARX* apresentados na seção 3.2.2, consideram q observações anteriores da série temporal ($a_{t-1} \dots a_{t-q}$) e p observações de uma variável externa ($b_{t-1} \dots b_{t-p}$) ao elaborar estimativas. *Nonlinear*, entretanto, refere-se à substituição dos coeficientes de regressão probabilística por uma MLP com conexões de *feedback*. Modelo treinado para identificar a correlação com o horizonte de previsão da sequência de observações da série temporal e dos estados anteriores da rede, reapresentados à camada de entrada junto de cada novo exemplo (HAYKIN, 2009).

Ahmad, Anderson e Lie (2015) avaliam os modelos *NARX*, MLP, *ARMA* e persistência *naive* da irradiância solar horária, em horizontes 24 horas à frente em 16 cidades na Nova Zelândia. A princípio, 12 subconjuntos de atributos, restritos à cidade de Auckland e agrupados por sua correlação com a irradiância, são apresentados ao modelo *NARX*. O subconjunto contendo temperatura máxima e mínima, pressão atmosférica, umidade relativa, ângulos de zênite e azumite e precipitação é selecionado por apresentar os melhores resultados. Então, utilizando 3 anos de observações ainda restritas à mesma cidade, o modelo *NARX* apresenta nRMSE (normalizado em proporção da amplitude da irradiância $I_{max} - I_{min}$) de 4.9%, seguido dos modelos *AR* com 6.5%, Persistência com 9% e MLP com 9.6%. Estimativas para as 16 cidades são comparadas à irradiância observada em gráficos para dias selecionados.

Elman Networks são arquiteturas recorrentes que armazenam o estado da camada oculta em uma camada de contexto, com o mesmo número de neurônios, reapresentando-o junto aos dados de entrada na iteração seguinte (ELMAN, 1990). O estudo de Sfetsos e Coonick (2000) compara *Elman Networks*, Redes RBF, Sistemas *Neuro-Fuzzy* e ANNs treinadas por *Backpropagation* e Algoritmo de Levenberg Marquardt (LM) em previsões de irradiância solar média horária. Leituras abrangendo 63 dias do ano de 1996, coletadas na ilha de Córsega na França, são utilizadas na verificação dos modelos. Entre os melhores resultados predominam abordagens baseadas em Redes Neurais Artificiais, com *Elman Networks*, Sistemas *Neuro-Fuzzy* e ANNs treinadas por LM, obtendo RMSE de 32.4, 31.93 e 29.82 W/m^2 , respectivamente, em conjuntos contendo somente radiação solar; e RMSE de 33.94, 30.97 e 27.58 W/m^2 ao acrescentar variáveis meteorológicas aos dados. Entretanto, o treinamento da arquitetura recorrente leva, em média, 20.860 segundos, comparado aos 654 segundos dos Sistemas *Neuro-Fuzzy* e 83 segundos necessários ao modelo treinado por LM.

Cao e Cao (2005) empregam Transformada *Wavelet* para obter componentes da radiação solar em diversos domínios no plano tempo-frequência. Observações diárias de irradiância entre os anos 1995 e 2000, provenientes de uma estação meteorológica em Xan-gai, são decompostas em uma série de baixa frequência a_3 (aproximações), subsequentemente detalhada por três séries de alta frequência d_3 , d_2 e d_1 (detalhes), apresentadas na Figura 9. De maneira que sua combinação reconstitua a série original. Cada componente é, então, apresentado a uma *Recurrent Backpropagation Network* (RBPN), arquitetura que reinterpreta as respostas da camada de saída à entrada da rede. Em comparação ao conjunto de dados original (não decomposto por *Wavelet*) que resulta em RMSE de 3.82 MJ/m^2 , estimativas do modelo RBPN recebendo como entrada a série a_3 apresentam RMSE de 0.92 MJ/m^2 .

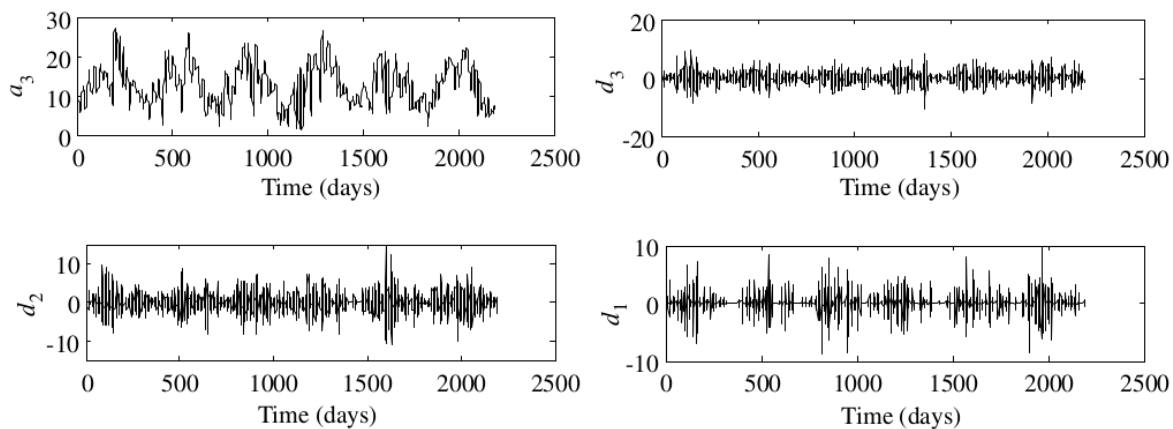


Figura 9 – Componentes da irradiância solar decompostos por Transformada de *Wavelets*. Fonte: Cao e Cao (2005)

3.2.3.3 Extreme Gradient Boosting (XGboost)

Modelos **XGBoost** constituem *ensembles* de árvores de regressão semelhantes à apresentada na Figura 10, que formulam hipóteses sobre os exemplos agregando as respostas de uma assembleia de preditores simples. O conjunto de árvores de regressão que compõem esta assembleia é elaborado em duas etapas: **(1)** uma árvore T_0 é construída adicionando à sua estrutura a ramificação do atributo F_i que mais aperfeiçoa o preditor a cada etapa e, definida sua estrutura, são calculados os valores ótimos para as folhas l ; **(2)** a inclusão de novas árvores $T_{1...N}$ ao *ensemble* é orientada pelo erro residual do modelo, de maneira que novos preditores complementem as imperfeições dos anteriores (FRIEDMAN, 2002).

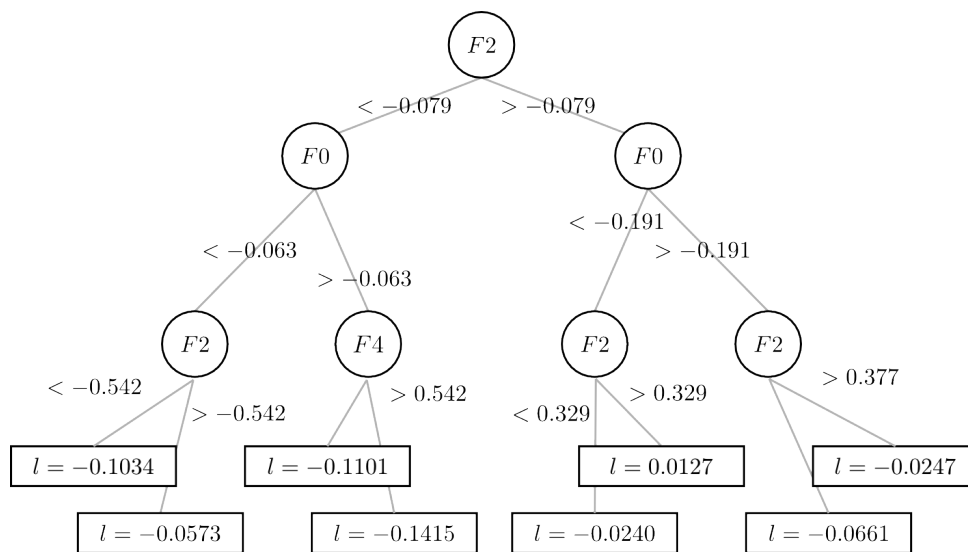


Figura 10 – Exemplo de árvore de regressão.

Modelos **XGBoost**, contudo, diferem de outras técnicas de *Gradient Boosting* pois empregam um algoritmo sensível à dispersão na busca por ramificações, que torna a complexidade computacional do modelo linear ao número de observações não ausentes. Assim como integram otimizações do uso de recursos que permitem calcular paralelamente a aptidão dos atributos (CHEN; GUESTRIN, 2016).

Somente um experimento aplicando modelos **XGBoost** à previsão de irradiância solar foi identificado na literatura correlata. O estudo publicado por Urraca et al. (2016) utiliza observações da irradiância solar diária, provenientes de 38 estações meteorológicas na Espanha, acompanhadas de temperatura, umidade relativa, velocidade do vento, precipitação e índice de Claridade, no treinamento e verificação de um modelo **XGBoost**. Após remover *outliers* (1% das observações) e estabelecer os parâmetros do modelo por meio de um algoritmo de Busca Aleatória, este alcança **nRMSE** de 12.34%. A contribuição individual dos atributos à precisão das estimativas também é examinada. Sendo índice de Claridade, umidade relativa e temperatura máxima responsáveis por 62%, 20% e 3.8% do aprimoramento observado.

4 Metodologia e Resultados

Experimentos com modelos preditivos de radiação solar ocorreram em duas etapas. Na primeira, foram avaliadas as estimativas de um modelo [RNN](#) da [GHI](#) incidente, em resoluções de 5 minutos, na cidade de Portland entre 2014 e 2016. Na segunda etapa, foram comparados os modelos de Persistência dos índices de Céu Claro e Claridade, [XGBoost](#) e [RNN](#) em observações de [GHI](#) em Las Vegas e Denver, minuto a minuto entre 2012 e 2016, re-amostradas em frequências de 5 minutos, 30 minutos, 1 hora e média diária. A [Seção 4.1](#) descreve brevemente os dados, a seleção de atributos, o modelo e os resultados do experimento inicial. A [Seção 4.2](#), por sua vez, caracteriza o experimento principal, dividindo-o em Conjunto de Dados, Seleção de Atributos e Experimentos. Por fim, a [Seção 4.1](#) apresenta e discute os resultados.

4.1 Experimento Inicial

A precisão de uma [RNN](#) foi examinada em estimativas da [GHI](#) observada em uma estação em Portland (45°N , 122°O)¹ durante os anos de 2014 a 2016. Os dados obtidos contém: irradiâncias global horizontal, direta e difusa (Wh/m^2), temperatura do ar, velocidade e direção do vento e produção de energia dos painéis, em intervalos de 5 minutos. Após acrescentar observações horárias da cobertura do céu no aeroporto de Portland, obtidas do *National Oceanic and Atmospheric Administration* (NOAA), os dados foram submetidos a um processo de remoção de *outliers*, normalização das leituras, seleção de atributos e, por fim, utilizados no treinamento e avaliação de um modelo [RNN](#).

4.1.1 Pré-processamento e Seleção de Atributos

São considerados *outliers* leituras rotuladas incertas ou ausentes nos dados, ou que apresentam desvio padrão 3 vezes superior ao da amostra. Por este critério, foram removidas 3.516 (2,92%) das 120.498 observações entre 1º de janeiro de 2014 e 31 de dezembro de 2016. Leituras válidas foram, então, ordenadas em períodos contendo 3 e 6 leituras consecutivas (15 e 30 minutos) de todos os atributos disponíveis.

Durante a normalização, a média dos atributos é centralizada em 0 e a escala das observações reduzida para que estejam contidas no intervalo $[-1, 1]$. Esta etapa é necessária pois a presença de atributos com magnitudes diferentes, como velocidade do vento (m/s) e irradiância global horizontal (Wh/m^2), por exemplo, pode causar a sub ou superestimação de sua relevância pelos modelos preditivos.

¹ [University of Oregon – Solar Radiation Monitoring Laboratory](#)

Utilizando a ferramenta Weka (HALL et al., 2009), foram aplicados 3 algoritmos de seleção de atributos aos conjuntos de dados: *CFS Subset* – seleciona os atributos que apresentam maior correlação com o objetivo e menor redundância; *ReliefF* – compara uma observação aleatória a k exemplos da mesma classe e k de classes diferentes, aprimorando estimativas da relevância em função da similaridade aos membros da mesma classe e distinção aos outros (ROBNIK-ŠIKONJA; KONONENKO, 1997); e *Principal Component Analysis (PCA)* – transforma atributos correlacionados em um número igual ou inferior de componentes principais independentes e representativos da maior variação possível dos dados, entretanto, não assegura que preservem a mesma relevância se transportados de volta ao espaço de atributos original (VOYANT et al., 2017). A Tabela 2 apresenta os 5 atributos com melhor classificação pelos métodos de seleção.

Tabela 2 – Atributos melhor classificados por cada método de seleção.

CFS Subset	ReliefF	PCA
15 minutos		
GHI t_{-1}	DNI t_{-1}	Horário t_{+1}
Cobertura Céu t_{-3}	DHI t_{-1}	Dia
Voltagem t_{-1}	DNI t_{-2}	Cobertura Céu t_{-3}
Voltagem t_{-1}	DNI t_{-3}	Horário t_{-2}
Dia	DHI t_{-3}	DHI t_{-2}
30 minutos		
GHI t_{-1}	DNI t_{-1}	Horário t_{+1}
Cobertura Céu t_{-6}	DNI t_{-2}	Cobertura Céu t_{-5}
Voltagem t_{-1}	DHI t_{-1}	Dia
	DNI t_{-3}	Temp. painéis t_{-5}
	DNI t_{-4}	Temp. painéis t_{-5}

Apesar dos resultados divergentes, alguns atributos se repetem de maneira consistente, como: irradiâncias direta, difusa e global horizontal, cobertura do céu, dia e horário. A relevância das observações de Cobertura do Céu, apesar de sua resolução temporal estender-se por diversas leituras dos outros atributos, sugere que contenham informações significativas sobre as condições climáticas no momento da estimativa. Dia e Horário, por sua vez, representam o componente determinístico da energia solar, referente aos movimentos astronômicos e aproximado pelos índices de Céu Claro ou Claridade. A predominância de leituras recentes como t_{-1} e t_{-2} pode ser motivada pelo horizonte de previsão de 5 minutos.

4.1.2 Experimentos e Resultados

Inicialmente, o modelo RNN foi treinado em uma amostra contendo os anos de 2014 e 2015 e avaliado nos períodos de 12 a 15 de julho e 11 a 14 de dezembro de 2016.

Considerando a classificação dos atributos na etapa anterior, os dados de entrada foram limitados a: 3 ou 6 observações mais recentes de irradiância (15 e 30 minutos); cobertura do céu mais recente; e dia e horário da estimativa. Parâmetros de configuração da RNN, como o tamanho da camada *hidden* (18 neurônios), a função de ativação *softsign* e o algoritmo de treinamento *RMSprop*, foram determinados empiricamente buscando melhores resultados.

No conjunto de observações em intervalos de 15 minutos, o modelo alcança **nRMSE** 50,1% em julho e 123,1% em dezembro, e 68,0% e 128,6% em intervalos de 30 minutos dos mesmos períodos. A Figura 11 apresenta a média e a variação das estimativas comparadas ao valor observado, como também a cobertura do céu a cada hora. Esta projeção evidencia variação e erro significativos no início dos dias, além de “degraus” coincidentes com as mudanças de cobertura do céu.

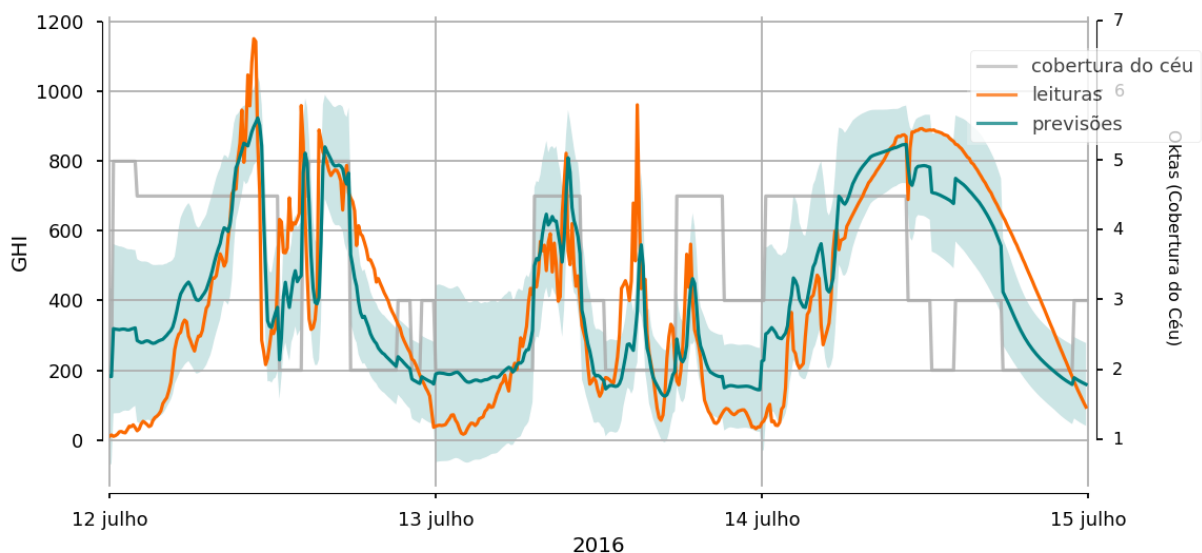


Figura 11 – Resultados das previsões obtidas para o experimento inicial.

Seguindo este experimento, foram preparados dois conjuntos de dados – em horários restritos das 7 às 18 horas e contendo as 24 horas do dia – com o intuito de explorar as causas do erro e variação constatados. O período utilizado na validação dos modelos também foi ampliado para todo o ano de 2016 e, junto da **nRMSE**, é calculado o **Erro Absoluto Médio (MAE)** das estimativas. A Tabela 3 apresenta os resultados dos novos experimentos.

Tabela 3 – Erros (**nRMSE** e **MAE**) calculados para os distintos períodos de análise do experimento inicial.

Período	Entrada	nRMSE	MAE
5 – 19h	3×GHI (15 min.) e cob. nuvens	83.76%	146.56 W/m^2
5 – 19h	6×GHI (30 min.) e cob. nuvens	78.25%	139.95 W/m^2
7 – 18h	3×GHI (15 min.) e cob. nuvens	68.79%	156.4 W/m^2
24h	3×GHI (15 min.) e cob. nuvens	305%	314.8 W/m^2

Estes resultados sugerem que: **(a)** o período noturno não contribui para a precisão do modelo; **(b)** observações em intervalos de 30 minutos produzem melhores estimativas, contrariando os resultados em dias selecionados; **(c)** o aprimoramento dos resultados utilizando o período restrito entre 7 e 18h não justifica desconsiderar períodos produtivos durante a manhã e início da noite, especialmente no verão.

4.2 Metodologia Proposta após Experimentos Iniciais

Nesta segunda etapa de experimentos, foram avaliados os modelos de Persistência dos índices de Céu Claro e de Claridade, [XGBoost](#) e [RNN](#), em observações de Las Vegas e Denver para horizontes de 5 minutos, 30 minutos, 1 hora e média diária. Esta seção descreve os seguintes componentes destes experimentos: Conjunto de Dados, incluindo modelo de Céu Claro, pré-processamento e re-amostragem das observações; Seleção de Atributos; e, por fim, Experimentos e Resultados.

4.2.1 Conjuntos de Dados

As observações de Portland foram substituídas pelos conjuntos de Las Vegas (36.06°N, 115.8°O) e Denver (39.75°N, 104.62°O), publicados pelo *National Renewable Energy Laboratory (NREL)*². A principal justificativa para a substituição é a ausência do ângulo de zênite solar no conjunto de Portland, necessário para determinar a radiação extraterrestre e o índice de Claridade. Aproximar seus valores requer calcular a declinação e horário solares, processo que, embora praticável, é de difícil validação e ainda presume-se que possa introduzir ruído aos modelos. Já os conjuntos de dados publicados pelo *NREL*, contêm observações minuto a minuto de irradiância e de variáveis meteorológicas e astronômicas, como o ângulo de zênite e azimute, velocidade e direção do vento e temperatura, abrangendo mais de 30 estações meteorológicas e fotovoltaicas nos Estados Unidos.

Durante o pré-processamento dos conjuntos de Denver e Las Vegas, constatou-se que menos de 0.025% dos exemplos em ambos os conjuntos constituem *outliers*, identificados por [GHI](#) negativo, [DNI](#) inferior a -999 , temperaturas abaixo de -99° ou ângulo de zênite superior a 89.95° . Esta última restrição se faz necessária, pois conforme o ângulo de zênite θ – distância do sol ao ponto mais alto da esfera celeste – se aproxima de 90° , reduz a irradiância extraterrestre I^{EX} , tendendo a 0 (zero) e causando índices de clareza $K \rightarrow \infty$, conforme mostrado por meio das equações 4.1 e 4.2:

$$\lim_{\cos \theta \rightarrow 0} I^{EX} = I_{CS} \cdot E_0 \cdot \cos \theta = 0 \quad (4.1)$$

² [Measurement and Instrumentation Data Center \(MIDC\) – NREL](#)

$$\lim_{I^{EX} \rightarrow 0} K = \frac{GHI}{I^{EX}} = \infty \quad (4.2)$$

Isto ocorre, pois mesmo após o pôr-do-sol ($\theta > 90^\circ$), ainda se observam valores de **GHI** pela contribuição da radiação difusa.

Foram acrescentados aos conjuntos a irradiância extraterrestre e o índice de claridade (calculados de acordo a Seção 2.2.2), estimativas da irradiância em Céu Claro, o respectivo índice de Céu Claro (produzidos pelo modelo descrito à seguir) e estimativas da cobertura de nuvens, provenientes do modelo NWP *Rapid Update Cycle (RUC)*³. Informações extraídas de NWP fora limitadas à cobertura de nuvens, pois afetam diretamente a radiação incidente, assim como se mostrou inviável examinar a importância de todas as variáveis envolvidas na construção destes modelos.

O modelo de irradiância em céu claro elaborado consiste em um mapeamento da expectativa máxima de **GHI** a cada hora e dia do ano, apresentado na Figura 12, semelhante aos descritos por Pedro e Coimbra (2012) e Bacher, Madsen e Nielsen (2009). Neste modelo, o eixo x representa as horas do dia, o eixo y são os dias do ano e a superfície descreve a média entre os valores máximos de **GHI** observados a cada dia e hora (x, y), entre 2012 e 2014.

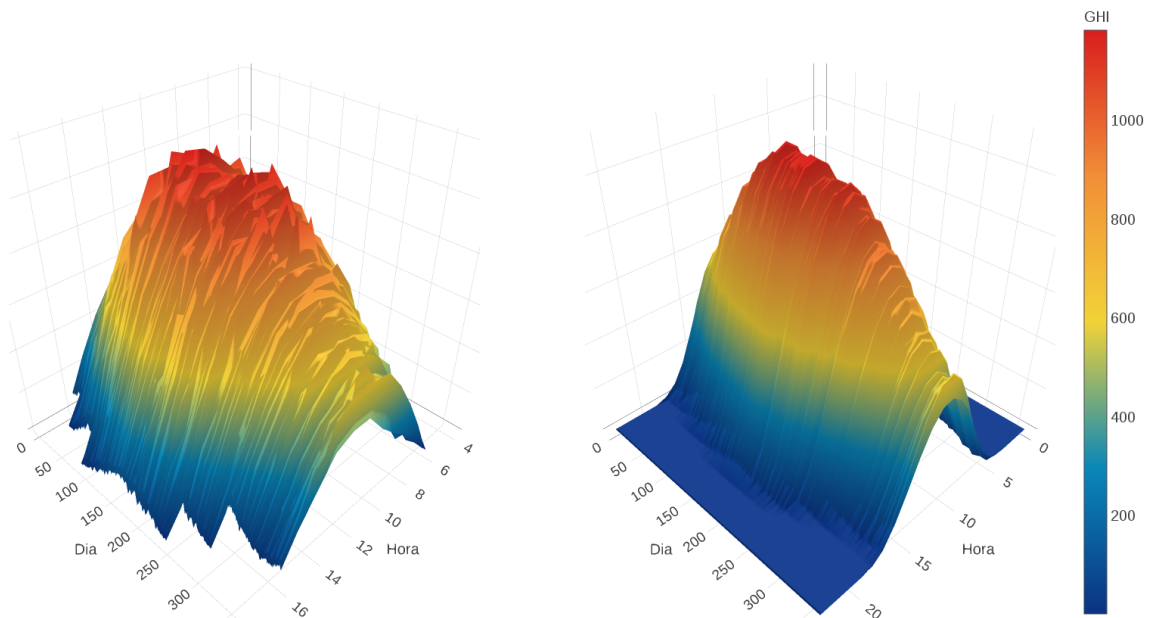


Figura 12 – Representação gráfica do modelo de Céu Claro elaborado para averiguar a expectativa de irradiância máxima correspondente às horas do dia e dias do ano, antes e após regularização por RBF.

³ Earth System Research Laboratory - Rapid Update Cycle (RUC) – NOAA

Uma **RBF** regulariza esta superfície, preenchendo leituras ausentes e atenuando os picos e oscilações observados na primeira parte da Figura 12. A superfície seguinte resulta deste processo, sendo que estimativas da irradiância em condições ideais, no dia e hora (x, y) , são obtidas acessando a posição correspondente nesta superfície.

Unir o conjunto de estimativas NWP da cobertura de nuvens horária aos conjuntos de observações minuto a minuto, incluindo os índices de Céu Claro e Clareza, exigiu que fossem re-amostradas em resoluções compatíveis, especificamente: 5 minutos, 30 minutos, 1 hora e média diária. Nos casos em que a frequência ultrapassa a resolução original dos dados, como ao re-amostrar estimativas NWP ou índices de Céu Claro horários em intervalos de 5 ou 30 minutos, os intervalos ausentes foram preenchidos pela interpolação linear dos valores.

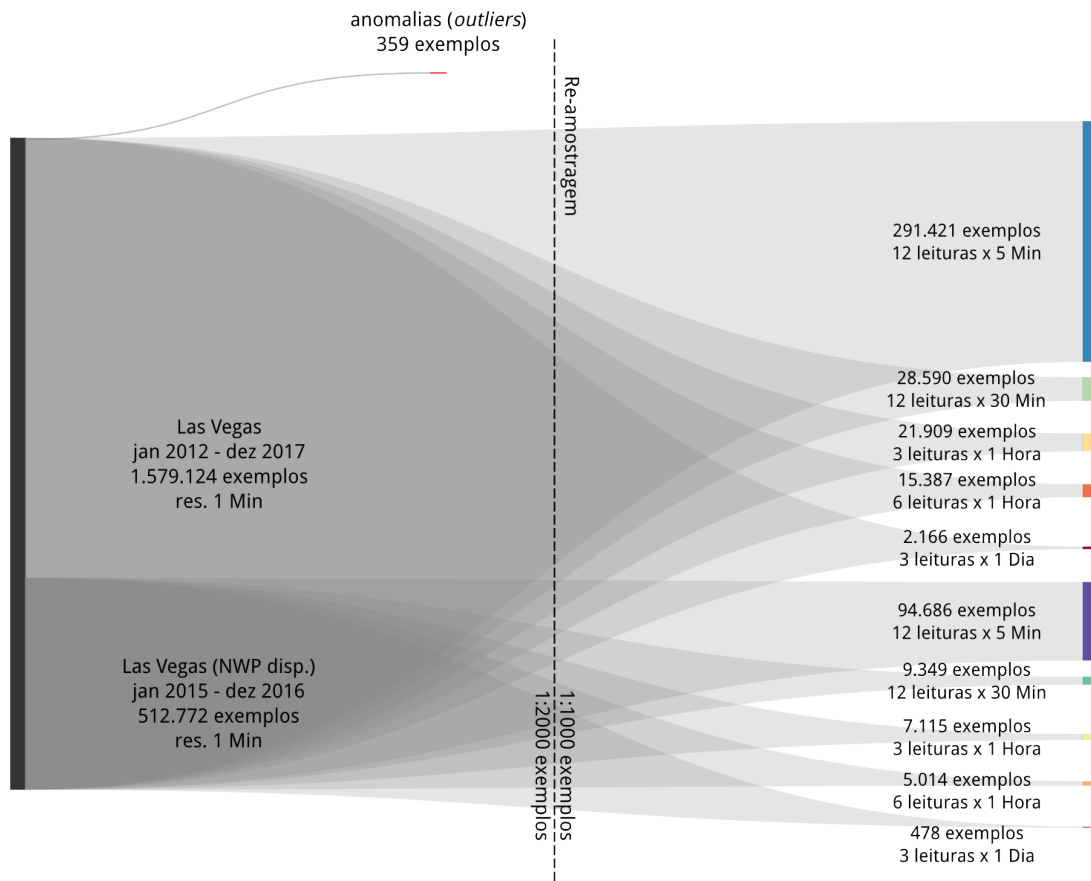


Figura 13 – Proporções dos conjuntos resultantes do pré-processamento e re-amostragem dos dados para a cidade de Las Vegas.

Este processo resultou em 20 conjuntos de dados para cada cidade, em proporções semelhantes às apresentadas na Figura 13, que descreve os conjuntos de Las Vegas. Metade destes contém estimativas NWP da cobertura de nuvens e compreendem o período limitado aos anos de 2015 e 2016, devido à disponibilidade de NWP. Cada exemplo integrante destes conjuntos descreve uma série temporal contendo: 12 leituras consecutivas de 5 minutos, 12 leituras de 30 minutos, 6 leituras de 1 hora, 3 leituras de 1 hora

ou 3 leituras de média diária. Em oposição aos conjuntos do primeiro experimento, nos quais as primeiras observações do dia sucedem as do dia anterior de maneira contínua, apesar da ausência do período noturno, as leituras são separadas dia a dia durante o pré-processamento. Conseqüentemente, exemplos em resoluções de 1 hora abrangem menos observações consecutivas (3 ou 6 ao invés de 12), caso contrário, excedem as leituras disponíveis nos dias de inverno. Exemplos em resoluções de 1 dia, por outro lado, apresentam somente 3 observações consecutivas com intuito de conservar o maior número possível de amostras.

Na seqüência, o infográfico da Figura 14 apresenta todas as etapas do pré-processamento e união dos conjuntos de dados.

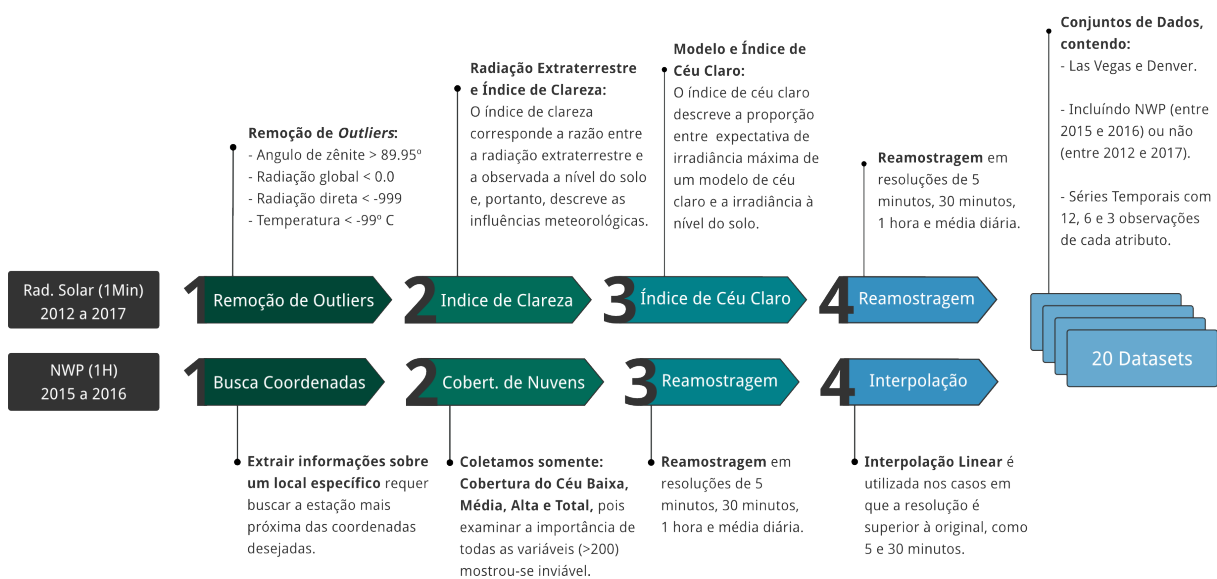


Figura 14 – Pré-processamento, união e re-amostragem dos conjuntos de dados.

4.2.2 Seleção de Atributos

Um algoritmo de **Eliminação Recursiva de Atributos (RFE)** foi utilizado na seleção de sub-conjuntos de variáveis informativas. A cada iteração deste algoritmo, o atributo menos relevante, conforme a classificação de um modelo preditivo, é removido do sub-conjunto. Este processo favorece atributos informativos em conjunto em vez de individualmente relevantes, diminuindo o ruído causado pelo aprendizado de características irrelevantes e a possibilidade de *overfitting* do modelo, usual reflexo da inexistência de exemplos de treinamento suficientes para distinguir associações relevantes entre as variáveis disponíveis (GUYON et al., 2002).

O modelo **Extreme Gradient Boosting (XGBoost)** determina quais atributos são eliminados, utilizando como critério de classificação o número de ramificações dos seus valores por meio das árvores do *ensemble*. Foram apresentadas amostras aleatórias de cada conjunto de dados e a importância dos atributos escalada em proporção do **nRMSE**

do modelo durante cada época de treinamento. A Figura 15 apresenta a classificação dos atributos para os conjuntos de Denver, revelando a predominância de um grupo pequeno de variáveis, principalmente em intervalos mais curtos. Característica que motivou a restringir a seleção aos atributos responsáveis por 50% da importância total do conjunto.

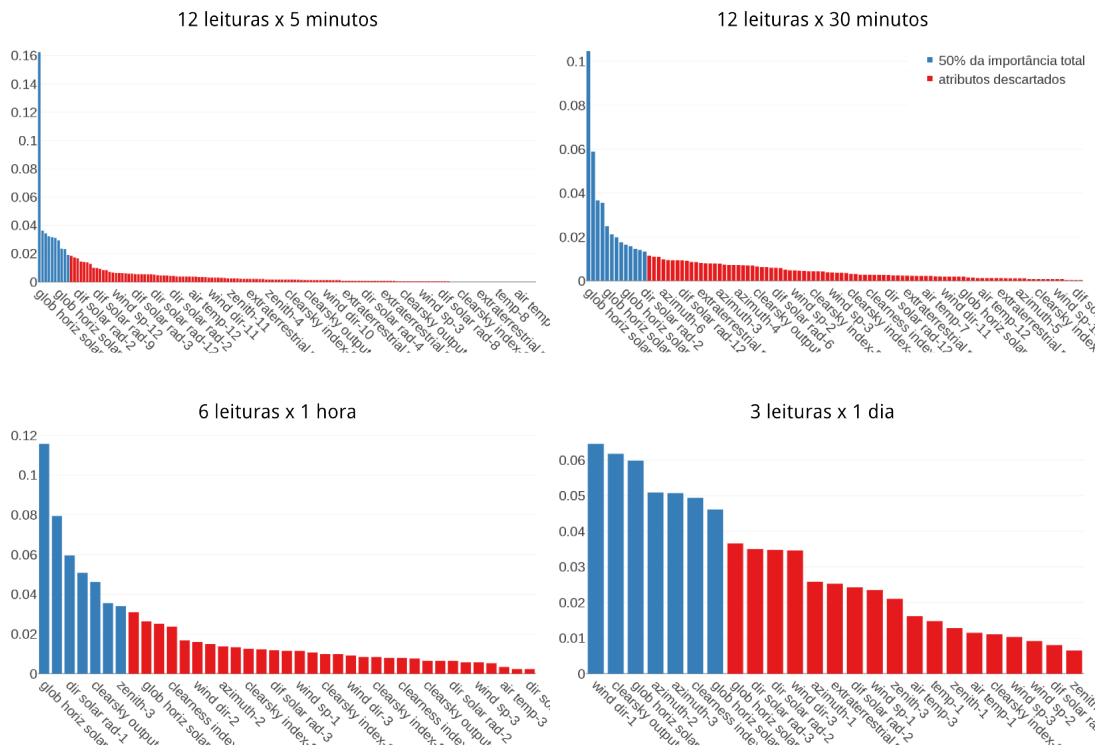


Figura 15 – Atributos selecionados por RFE, com ênfase aos responsáveis por 50% da importância total do sub-conjunto.

Em concordância ao experimento inicial, alguns atributos repetem-se consistentemente entre os sub-conjuntos selecionados. A Tabela 4 apresenta os 5 atributos melhor classificados em cada resolução, exibindo um único atributo quando os resultados coincidem, para os conjuntos de Denver, Denver (NWP), Las Vegas e Las Vegas (NWP).

Prevalece a **Irradiância Global Horizontal (GHI)** como atributo mais informativo em todos os conjuntos e horizontes, acompanhada das irradiâncias difusa (**DHI**) e direta (**DNI**), provavelmente devido à correlação com o objetivo. Índices de Céu Claro e Claridade, como também expectativas de irradiância em Céu Claro – representados por *clearsky-index*, *clearness-index* e *clearsky-output* – demonstram maior relevância em intervalos curtos, enquanto variáveis astronômicas como ângulos de zênite e azimute, e também velocidade e direção do vento, são mais informativas em horizontes distantes. Somente em frequências diárias as observações NWP de cobertura de nuvens demonstram importância, possível consequência de sua resolução temporal (1 hora) e espacial (13km).

Tabela 4 – Primeiros 5 atributos selecionados em cada resolução através dos conjuntos.

Leituras × Frequência	1º atributo	2º atributo	3º atributo	4º atributo	5º atributo
12×5 min.	GHI _{t-1}	GHI _{t-12} , DHI _{t-1} , clearsky- output _{t-1}	DHI _{t-1} , DNI _{t-1}	GHI _{t-3} , GHI _{t-12} , DNI _{t-1}	GHI _{t-6} , DNI _{t-1} , clearsky- output _{t-1} , clearness- index _{t-2}
12×30 min.	GHI _{t-1}	DHI _{t-1}	DNI _{t-1}	clearsky- output _{t-1} , clearness- index _{t-1} , zenith _{t-1} , azimuth _{t-1}	GHI _{t-1} , GHI _{t-2} , zenith _{t-1}
3×1 hora	GHI _{t-1}	azimuth _{t-3}	DHI _{t-1} , DNI _{t-1} , zenith _{t-1}	DHI _{t-1} , zenith _{t-1}	clearsky- output _{t-1}
6×1 hora	GHI _{t-1}	DNI _{t-1} , azimuth _{t-5} , azimuth _{t-6}	DHI _{t-1} , DNI _{t-1}	DHI _{t-1} , DNI _{t-1} , azimuth _{t-6}	azimuth _{t-5} , azimuth _{t-3} , clearsky- output _{t-1}
3×1 dia	GHI _{t-1} , GHI _{t-3} , wind-dir _{t-1}	GHI _{t-1} , azimuth _{t-1} , clearsky- output _{t-1} , high-cloud- cover _{t-1}	GHI _{t-1} , DHI _{t-3} , wind- speed _{t-3} , azimuth _{t-3}	clearsky- output _{t-1} , azimuth _{t-1} , azimuth _{t-2} , azimuth _{t-3}	DHI _{t-1} , medium- cloud- cover _{t-2} , azimuth _{t-2} , azimuth _{t-3}

4.2.3 Experimentos

A precisão das estimativas de uma [RNN](#) simples a dos modelos de persistência dos índices de Céu Claro e Claridade foi comparada, bem como do modelo [XGBoost](#), em virtude da acurácia que demonstrou durante a seleção de atributos. Esta arquitetura recorrente apresenta conexões entre os neurônios ocultos e a camada de entrada, de maneira idêntica às camadas de contexto das *Elman Networks*.

Estimativas de persistência do índice de Céu Claro, assim como às descritas na seção [3.2.1](#), correspondem à expectativa de irradiância em condições climáticas ótimas no horizonte de previsão, atenuada pelas condições climáticas presentes. Esta atenuação, por sua vez, é representada pelo índice de Céu Claro k_t , razão da irradiância observada I_t pela expectativa I_t^{CS} obtida do modelo de Céu Claro apresentado na seção [4.2.1](#). A persistência do índice de Claridade difere unicamente na origem das estimativas de irradiância máxima, utilizando a irradiância extraterrestre, calculada em relação ao ângulo de zênite pela

Equação 2.3. Os modelos RNN empregam uma única camada oculta contendo neurônios em número igual a 80% do número de atributos, proporção acima da qual não foi possível notar melhoria das estimativas. Empregou-se também funções de ativação *Exponential Linear Unit (ELU)* com o intuito de reduzir o efeito de *vanishing gradients* (CLEVERT; UNTERTHINER; HOCHREITER, 2015). Por sua vez, o modelo XGBoost é composto de 100 árvores de regressão com profundidade máxima de 3 níveis.

O procedimento de treinamento e verificação dos modelos, representado pela Figura 16, é composto das seguintes etapas:

1. Os conjuntos de dados são divididos pelo algoritmo *K-Fold Cross Validation* em 10 trechos. A cada iteração, 9 destes são utilizados no treinamento, alternando o trecho utilizado para verificar o modelo durante 10 iterações. Como o modelo é sempre avaliado em exemplos que desconhece, é possível verificar sua capacidade de generalização ou *overfitting*;
2. A contribuição dos índices de Céu Claro e Claridade é avaliada acrescentando somente um, ou outro, ou ambos os atributos aos conjuntos de treinamento;
3. Após o treinamento e validação dos modelos RNN e XGBoost, os atributos eleitos como ramificações do *ensemble* são apresentados a um novo modelo RNN-sf (*selected features*), armazenando a taxa erro ($e_{rnn,i}$, $e_{xgb,i}$, $e_{rnn-se,i}$ e $e_{rnn-rfe,i}$) de cada modelo após a validação;
4. Por fim, os sub-conjuntos informativos, responsáveis por 50% da importância total dos atributos selecionados por RFE, são apresentados a um terceiro modelo RNN-rfe.

Modelos de Persistência, por sua vez, não requerem treinamento e não apresentam *overfitting*, pois resultam do índice de Céu Claro ou Claridade aplicado à expectativa de irradiância máxima no horizonte. Todavia, são avaliados utilizando os mesmos conjuntos em um experimento dedicado.

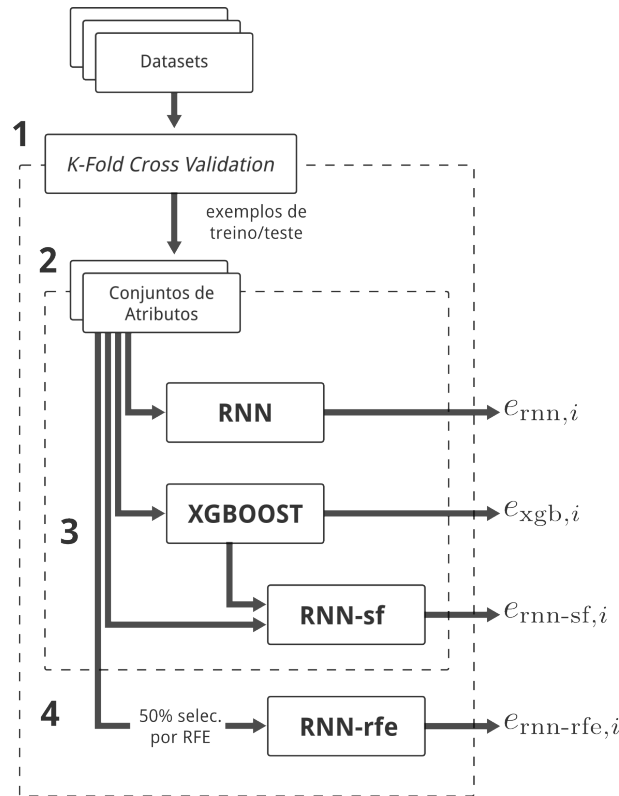


Figura 16 – Procedimento de treinamento e avaliação dos modelos.

4.3 Resultados e Discussão

No Apêndice A encontram-se o **nRMSE** e **MAPE** dos modelos de persistência dos índices de Céu Claro (*clk*) e Claridade (*cln*), RNN, RNN-sf, RNN-rfe e XGBoost (*xgb*), em séries temporais e horizontes de 12×5 minutos, 12×30 minutos, 3×1 hora, 6×1 hora e $3 \times$ média diária, separados de acordo com o conjunto de dados em Denver, Denver (NWP), Las Vegas e Las Vegas (NWP). Os resultados dos modelos RNN e RNN-sf são acompanhados de um rótulo subscrito indicando as variáveis contidas no conjunto que apresenta os melhores resultados, podendo significar: somente variáveis iniciais (*ird*), inclusão dos índices de Céu Claro (*clk*) ou de Claridade (*cln*) ou todos estes atributos (*all*). Embora também recebam sub-conjuntos de variáveis selecionadas, sua influência nos resultados dos modelos XGBoost é insignificante, motivo por não apresentarem rótulos.

Superficialmente, os resultados demonstram a hegemonia do modelo XGBoost, que apresenta o menor nRMSE para todos os experimentos. Contudo, também revelam que modelos de persistência, calculados em altas resoluções (minuto a minuto) superam os modelos RNN, aproximando-se dos melhores resultados em estimativas de curto prazo. Embora sejam superados no horizonte de média diária, em que estimativas de RNNs equiparam a acurácia do modelo XGBoost. A Figura 17 auxilia a interpretação dos resultados, indicando o nRMSE médio através das localizações e conjuntos de atributos, para cada modelo e horizonte.

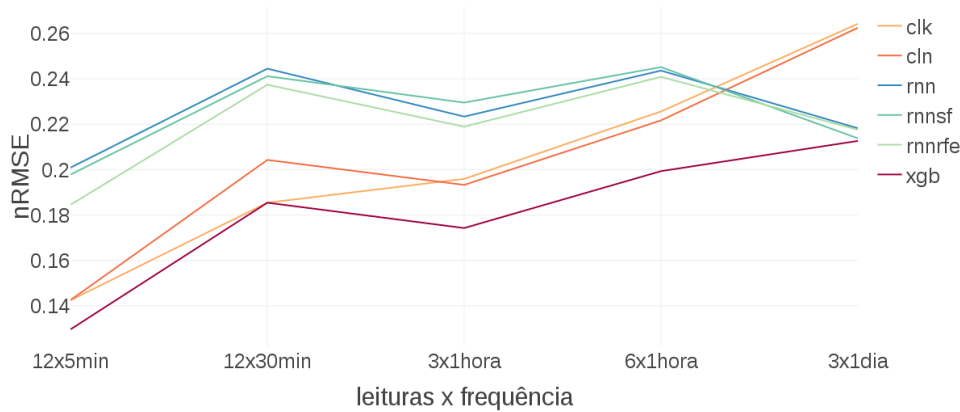


Figura 17 – $nRMSE$ médio através dos conjuntos de dados, para cada modelo e horizonte.

Projeções das estimativas em horizontes de 5 minutos, restritas aos dias 8 a 12 de junho de 2015, são comparadas à irradiância efetiva nas Figuras 18 a 21. A princípio, as estimativas apresentadas na Figura 18, demonstram a precisão do modelo **XGBoost**, assim como a ausência de variação entre resultados ao longo das épocas de treinamento e verificação realizadas.

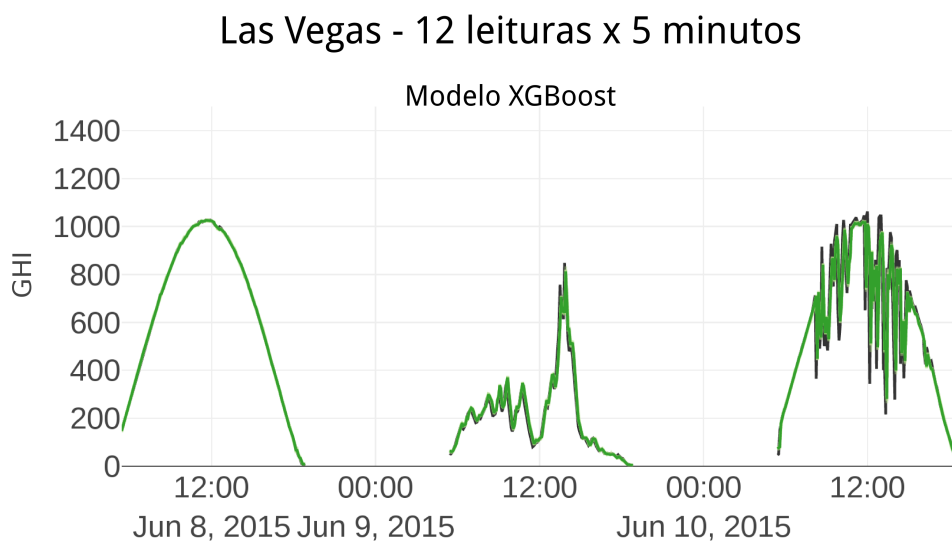


Figura 18 – Estimativas do modelo **XGBoost** entre os dias 8 e 12 de junho de 2015, comparadas à irradiância observada.

A projeção seguinte, apresentada na Figura 19, revela a superestimação da irradiância pelo modelo de persistência do índice de céu claro, sobretudo em dias de menor nebulosidade. Ainda assim, acompanham proximamente a variação causada pelas condições atmosféricas, em virtude da interpolação linear do modelo de Céu Claro horário em intervalos de 5 minutos, aliada à descrição da atenuação da irradiância estabelecida pelo índice respectivo, calculada na mesma frequência.

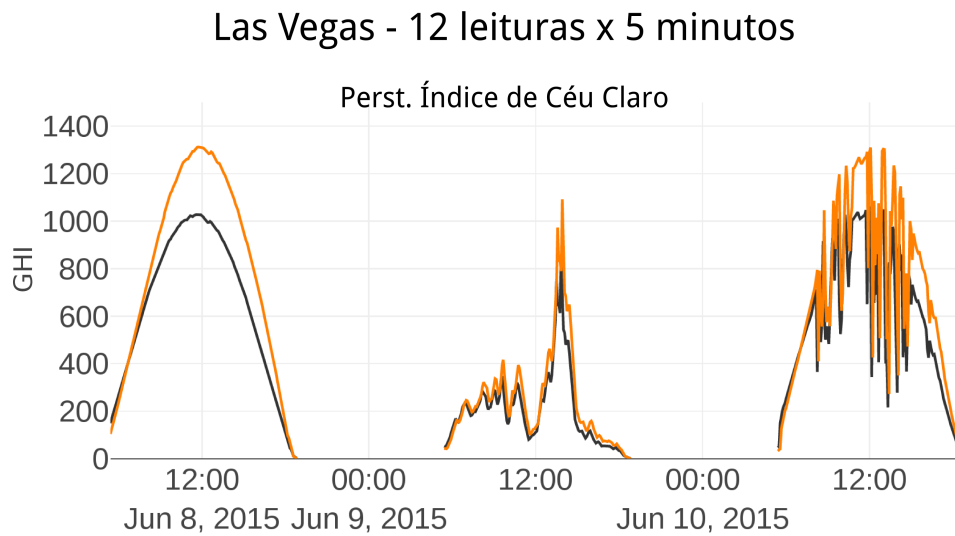


Figura 19 – Estimativas do modelo de persistência do índice de céu claro, entre os dias 8 e 12 de junho de 2015, comparadas à irradiância observada.

Estimativas de persistência do índice de claridade, apresentadas na Figura 20, demonstram os melhores resultados entre os modelos examinados neste horizonte, sendo possível observar a divergência entre estimativas e leituras somente ao reduzir a escala das projeções para algumas horas. Estas divergências resultam, principalmente, do atraso temporal intrínseco às estimativas de persistência. Todavia, ao estabelecer a irradiância extraterrestre e sua atenuação a cada minuto, estas são capazes de aproximadamente equiparar a variação do recurso solar.

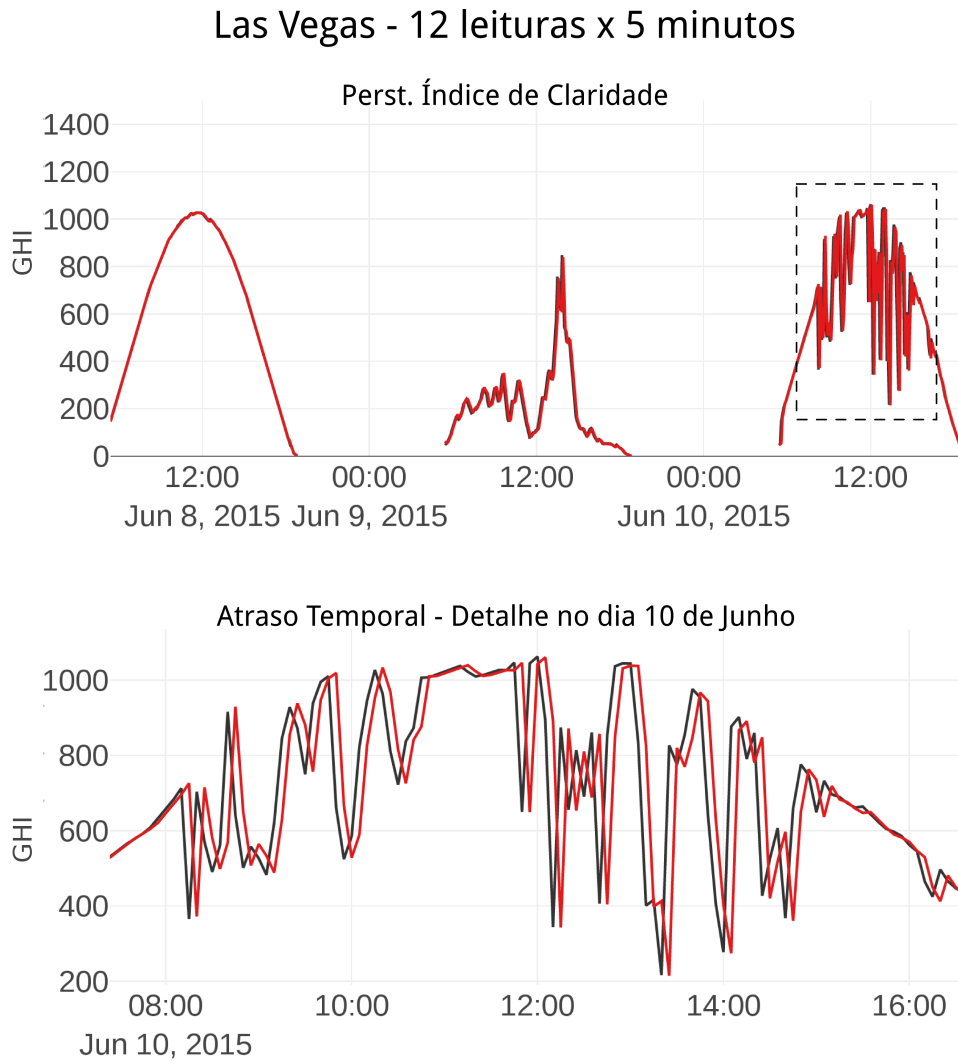


Figura 20 – Estimativas do modelo de persistência do índice de claridade, entre os dias 8 e 12 de junho de 2015, comparadas à irradiância observada. Seguidas da projeção do dia 10 de junho, entre as 8 e 16 horas.

Por fim, [RNNs](#) demonstram significativa divergência e variação ao longo das épocas de treinamento. Contudo, a restrição dos atributos aos responsáveis por 50% da importância estabelecida por [RFE](#), embora não aperfeiçoe as estimativas, sustenta precisão comparável e, aparentemente, diminui o ruído em dias ensolarados.

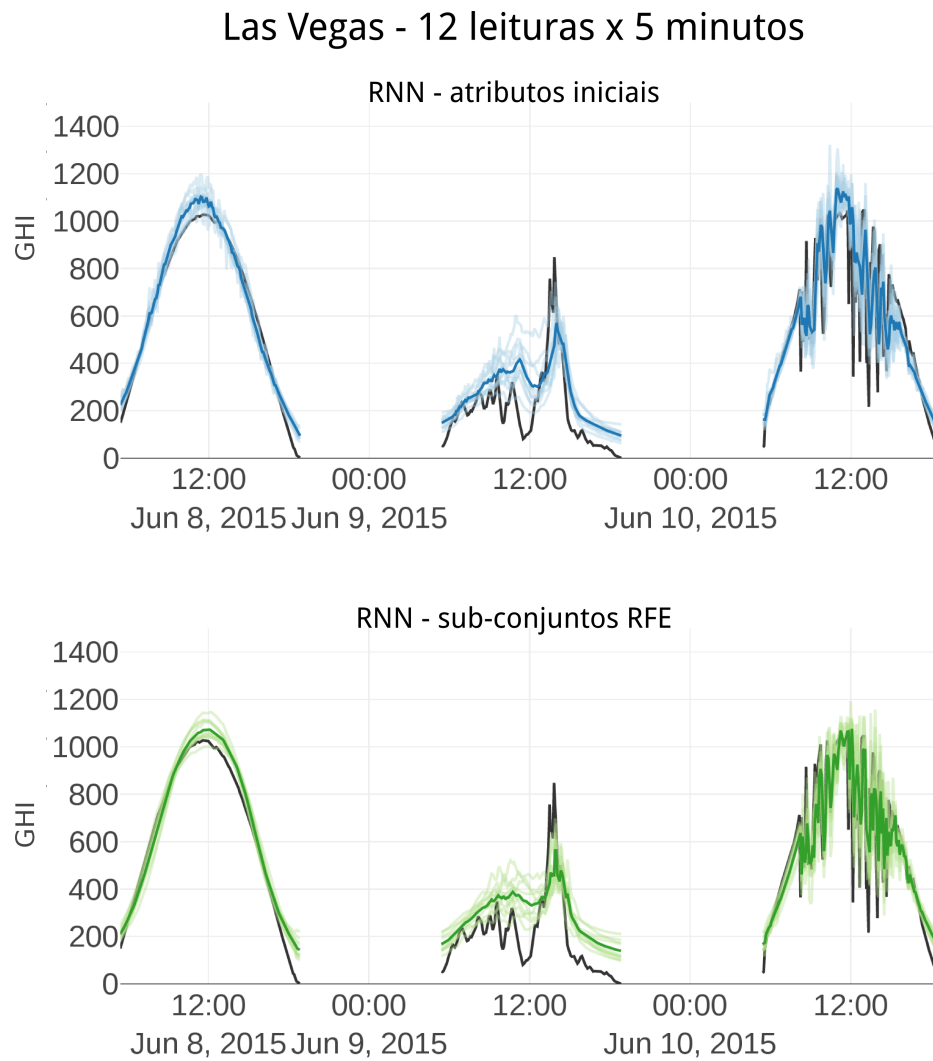


Figura 21 – Estimativas do modelo RNN, entre os dias 8 e 12 de junho de 2015, comparadas à irradiância observada.

5 Considerações Finais

Esta dissertação revisou uma amostra da literatura que concerne a previsão de radiação solar e produção fotovoltaica por métodos de Aprendizado de Máquina e Análise Probabilística de Séries Temporais. Partindo desta, apresentou dois experimentos destinados a examinar a aplicação de [Redes Neurais Recorrentes \(RNNs\)](#) e modelos [Extreme Gradient Boosting \(XGBoost\)](#) em estimativas da irradiância global horizontal (GHI) em horizontes de curto prazo.

A análise da literatura selecionada, revelou a predominância de [Redes Neurais Artificiais \(RNAs\)](#) entre os modelos com melhores resultados em experimentos publicados, seguidas por métodos de Análise Probabilística de Séries Temporais, sendo pouco frequente a utilização de RNNs. Demonstrou a ausência de um procedimento de avaliação homogêneo, tanto no que diz respeito às medidas de erro quanto às localidades de origem dos dados das quais as estimativas são fortemente dependentes. E, também, contribuiu para a elaboração dos modelos de persistência dos índices de Céu Claro e Claridade que demonstraram precisão significativa, alcançando os melhores resultados em horizontes de 5 minutos. Contudo, sobre estes é importante esclarecer que determinar a atenuação imposta pelas condições atmosféricas minuto a minuto antes de re-amostrar as leituras, pode ter dado origem a uma correspondência artificial entre estes índices e a GHI média em cada intervalo.

Os experimentos, por sua vez, foram realizados com o intuito de verificar duas hipóteses principais: **(1)** a precisão de RNNs é comparável a de outros modelos e experimentos na literatura, em virtude de sua capacidade de identificar relacionamentos temporais entre os dados; **(2)** limitar os exemplos a subconjuntos de atributos representativos favorece o aprendizado dos modelos e, conseqüentemente, a precisão das estimativas. Sobre a primeira hipótese, em horizontes distantes, RNNs demonstram resultados equivalentes aos melhores obtidos e, mesmo em outros horizontes, sua precisão é comparável a de experimentos similares na literatura. A segunda hipótese, por sua vez, é confirmada pelos resultados obtidos nos conjuntos limitados às variáveis responsáveis por 50% da importância determinada por [Eliminação Recursiva de Atributos \(RFE\)](#) e corroborada por resultados em estudos que empregam métodos de seleção e pelo desempenho do modelo *XGBoost*, que intrinsecamente seleciona subconjuntos informativos de atributos.

Estudos subsequentes são necessários para averiguar a precisão de estimativas produzidas por índices de Céu Claro e Claridade calculados em altas resoluções, estabelecer quais atributos carregam informação relevante sobre a irradiância solar, por meio da contribuição de cada um destes, e examinar a ocorrência de *vanishing* ou *exploding gradients*,

bem como sua relação com o número de exemplos e atributos utilizados durante o treinamento das RNNs.

Referências

- AHMAD, A.; ANDERSON, T.; LIE, T. Hourly global solar irradiation forecasting for new zealand. *Solar Energy*, Elsevier, v. 122, p. 1398–1408, 2015. Citado na página 41.
- ANTONANZAS, J. et al. Review of photovoltaic power forecasting. *Solar Energy*, Elsevier, v. 136, p. 78–111, 2016. Citado 3 vezes nas páginas 23, 26 e 32.
- ARIA, M.; CUCCURULLO, C. bibliometrix: An r-tool for comprehensive science mapping analysis. *Journal of Informetrics*, Elsevier, v. 11, n. 4, p. 959–975, 2017. Citado na página 30.
- BACHER, P.; MADSEN, H.; NIELSEN, H. A. Online short-term solar power forecasting. *Solar Energy*, Elsevier, v. 83, n. 10, p. 1772–1783, 2009. Citado 5 vezes nas páginas 29, 31, 32, 34 e 49.
- BIRD, R. E.; HULSTROM, R. L. *Simplified clear sky model for direct and diffuse insolation on horizontal surfaces*. [S.l.], 1981. Citado na página 26.
- BISGAARD, S.; KULAHCI, M. *Time series analysis and forecasting by example*. [S.l.]: John Wiley & Sons, 2011. Citado 3 vezes nas páginas 34, 35 e 36.
- BOLAND, J. Time series modelling of solar radiation. *Modeling Solar Radiation at the Earth's Surface*, Springer Verlag, Springer, p. 283–312, 2008. Citado na página 34.
- BRASIL, M. Terceira comunicação nacional do brasil à convenção-quadro das nações unidas sobre mudança do clima. *Mct, Brasília, dF, Brasil*, v. 3, 2016. Citado na página 22.
- CAO, S.; CAO, J. Forecast of solar irradiance using recurrent neural networks combined with wavelet analysis. *Applied Thermal Engineering*, Elsevier, v. 25, n. 2-3, p. 161–172, 2005. Citado 3 vezes nas páginas 11, 41 e 42.
- CHEN, C. et al. Online 24-h solar power forecasting based on weather type classification using artificial neural network. *Solar Energy*, Elsevier, v. 85, n. 11, p. 2856–2870, 2011. Citado na página 31.
- CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: ACM. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. [S.l.], 2016. p. 785–794. Citado na página 43.
- CLEVERT, D.-A.; UNTERTHINER, T.; HOCHREITER, S. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*, 2015. Citado na página 54.
- DAVID, M. et al. Probabilistic forecasting of the solar irradiance with recursive arma and garch models. *Solar Energy*, Elsevier, v. 133, p. 55–72, 2016. Citado na página 35.
- DIAGNE, M. et al. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renewable and Sustainable Energy Reviews*, Elsevier, v. 27, p. 65–76, 2013. Citado 3 vezes nas páginas 23, 28 e 31.

- ECK, N. J. van; WALTMAN, L. Visualizing bibliometric networks. In: *Measuring scholarly impact*. [S.l.]: Springer, 2014. p. 285–320. Citado na página 29.
- EDENHOFER, O. et al. *Renewable energy sources and climate change mitigation: Special report of the intergovernmental panel on climate change*. [S.l.]: Cambridge University Press, 2011. Citado na página 23.
- ELMAN, J. L. Finding structure in time. *Cognitive science*, Wiley Online Library, v. 14, n. 2, p. 179–211, 1990. Citado 2 vezes nas páginas 40 e 42.
- EPA, A. Inventory of us greenhouse gas emissions and sinks: 1990-2015. *Environmental Protection Agency 2017*, 2017. Citado na página 22.
- FRIEDMAN, J. H. Stochastic gradient boosting. *Computational Statistics & Data Analysis*, Elsevier, v. 38, n. 4, p. 367–378, 2002. Citado na página 43.
- GUYON, I. et al. Gene selection for cancer classification using support vector machines. *Machine learning*, Springer, v. 46, n. 1-3, p. 389–422, 2002. Citado na página 51.
- HALL, M. et al. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, ACM, v. 11, n. 1, p. 10–18, 2009. Citado na página 46.
- HAYKIN, S. S. *Neural networks and learning machines*. [S.l.]: Pearson Upper Saddle River, NJ, USA:, 2009. Citado 5 vezes nas páginas 11, 37, 38, 39 e 41.
- HÖÖK, M.; TANG, X. Depletion of fossil fuels and anthropogenic climate change—a review. *Energy Policy*, Elsevier, v. 52, p. 797–809, 2013. Citado na página 22.
- IEA. Co2 emissions from fuel combustion 2016. Organisation for Economic Co-operation and Development, OECD, 2016. Citado na página 22.
- IEA. World energy outlook 2016. Organisation for Economic Co-operation and Development, OECD, 2016. Citado na página 23.
- IEA. Tracking clean energy progress 2017. Organisation for Economic Co-operation and Development, OECD, 2017. Citado na página 23.
- INEICHEN, P.; PEREZ, R. A new air mass independent formulation for the link turbidity coefficient. *Solar Energy*, Elsevier, v. 73, n. 3, p. 151–157, 2002. Citado na página 33.
- INMAN, R. H.; PEDRO, H. T.; COIMBRA, C. F. Solar forecasting methods for renewable energy integration. *Progress in energy and combustion science*, Elsevier, v. 39, n. 6, p. 535–576, 2013. Citado 9 vezes nas páginas 23, 25, 26, 28, 31, 32, 34, 35 e 36.
- IQBAL, M. *An introduction to solar radiation*. [S.l.]: Elsevier, 1983. Citado na página 27.
- IRENA. Global energy transformation: A roadmap to 2050. IRENA, 2018. Citado 2 vezes nas páginas 22 e 23.
- KOPP, G.; LEAN, J. L. A new, lower value of total solar irradiance: Evidence and climate significance. *Geophysical Research Letters*, Wiley Online Library, v. 38, n. 1, 2011. Citado na página 25.

- LIPPERHEIDE, M.; BOSCH, J.; KLEISSL, J. Embedded nowcasting method using cloud speed persistence for a photovoltaic power plant. *Solar Energy*, Elsevier, v. 112, p. 232–238, 2015. Citado 2 vezes nas páginas [32](#) e [33](#).
- MARQUEZ, R.; COIMBRA, C. F. Proposed metric for evaluation of solar forecasting models. *Journal of solar energy engineering*, American Society of Mechanical Engineers, v. 135, n. 1, p. 011016, 2013. Citado na página [26](#).
- MELLIT, A.; PAVAN, A. M. A 24-h forecast of solar irradiance using artificial neural network: application for performance prediction of a grid-connected pv plant at trieste, italy. *Solar Energy*, Elsevier, v. 84, n. 5, p. 807–821, 2010. Citado 3 vezes nas páginas [29](#), [31](#) e [39](#).
- MUELLER, R. et al. Rethinking satellite-based solar irradiance modelling: The solis clear-sky module. *Remote sensing of Environment*, Elsevier, v. 91, n. 2, p. 160–174, 2004. Citado na página [26](#).
- PACHAURI, R. K. et al. *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. [S.l.]: IPCC, 2014. Citado 2 vezes nas páginas [21](#) e [22](#).
- PAOLI, C. et al. Forecasting of preprocessed daily solar radiation time series using neural networks. *Solar Energy*, Elsevier, v. 84, n. 12, p. 2146–2160, 2010. Citado 2 vezes nas páginas [31](#) e [36](#).
- PASCANU, R.; MIKOLOV, T.; BENGIO, Y. On the difficulty of training recurrent neural networks. In: *International Conference on Machine Learning*. [S.l.: s.n.], 2013. p. 1310–1318. Citado 2 vezes nas páginas [24](#) e [41](#).
- PEDRO, H. T.; COIMBRA, C. F. Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy*, Elsevier, v. 86, n. 7, p. 2017–2028, 2012. Citado 7 vezes nas páginas [11](#), [26](#), [31](#), [32](#), [33](#), [40](#) e [49](#).
- REIKARD, G. Predicting solar radiation at high resolutions: A comparison of time series forecasts. *Solar Energy*, Elsevier, v. 83, n. 3, p. 342–349, 2009. Citado 2 vezes nas páginas [31](#) e [36](#).
- ROBNIK-ŠIKONJA, M.; KONONENKO, I. An adaptation of relief for attribute estimation in regression. In: *Machine Learning: Proceedings of the Fourteenth International Conference (ICML'97)*. [S.l.: s.n.], 1997. p. 296–304. Citado na página [46](#).
- ROGELJ, J. et al. Paris agreement climate proposals need a boost to keep warming well below 2 c. *Nature*, Nature Publishing Group, v. 534, n. 7609, p. 631–639, 2016. Citado na página [21](#).
- SFETSOS, A.; COONICK, A. Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques. *Solar Energy*, Elsevier, v. 68, n. 2, p. 169–178, 2000. Citado 3 vezes nas páginas [31](#), [41](#) e [42](#).
- TOVAR-PESCADOR, J. Modelling the statistical properties of solar radiation and proposal of a technique based on boltzmann statistics. In: *Modeling Solar Radiation at the Earth's Surface*. [S.l.]: Springer, 2008. p. 55–91. Citado na página [27](#).

- UNFCCC, V. Adoption of the paris agreement. *I: Proposal by the President (Draft Decision)*, United Nations Office, Geneva (Switzerland), n. s 32, 2015. Citado na página 21.
- URRACA, R. et al. Estimation of daily global horizontal irradiation using extreme gradient boosting machines. In: SPRINGER. *International Joint Conference SOCO'16-CISIS'16-ICEUTE'16*. [S.l.], 2016. p. 105–113. Citado na página 43.
- VOYANT, C. et al. Machine learning methods for solar radiation forecasting: A review. *Renewable Energy*, Elsevier, v. 105, p. 569–582, 2017. Citado na página 46.
- VOYANT, C. et al. Multi-horizon solar radiation forecasting for mediterranean locations using time series models. *Renewable and Sustainable Energy Reviews*, Elsevier, v. 28, p. 44–52, 2013. Citado 2 vezes nas páginas 29 e 39.
- YADAV, A. K.; CHANDEL, S. Solar radiation prediction using artificial neural network techniques: A review. *Renewable and sustainable energy reviews*, Elsevier, v. 33, p. 772–781, 2014. Citado na página 31.
- YULE, G. U. et al. Vii. on a method of investigating periodicities disturbed series, with special reference to wolfer's sunspot numbers. *Phil. Trans. R. Soc. Lond. A*, The Royal Society, v. 226, n. 636-646, p. 267–298, 1927. Citado na página 34.

APÊNDICE A – Medidas de Erro para todos os Modelos e Horizontes

Tabela 5 – $nRMSE$ e $MAPE$ de todos os modelos e horizontes examinados para a cidade de Las Vegas. Valores decimais representam porcentagens ($0.25 = 25\%$)

Leituras × Frequência		RNN	RNN-sf	RNN-rfe	XGboost	Persist. Céu Claro	Persist. Clareza
Las Vegas							
12×5 min.	$nRMSE$	0.1471	0.1420	0.1614	0.1031	0.1111	0.1112
	$MAPE$	∞	∞	∞	∞	0.0962	0.0749
12×30 min.	$nRMSE$	0.1725 _(cln)	0.1725 _(cln)	0.1849	0.1410	0.1539	0.1519
	$MAPE$	1.5915	1.5940	2.5137	0.2675	0.4737	0.2067
3×1 hora	$nRMSE$	0.1706 _(clk)	0.1715 _(cln)	0.1940	0.1360	0.1546	0.1483
	$MAPE$	1.5924	1.4574	2.5451	0.2099	0.4456	0.2108
6×1 hora	$nRMSE$	0.1849 _(clk)	0.1815 _(cln)	0.1927	0.1518	0.1721	0.1639
	$MAPE$	1.6328	1.7265	2.4043	0.2524	0.5907	0.2591
3×1 dia	$nRMSE$	0.1715 _(all)	0.1683 _(cln)	0.1689	0.1654	0.2012	0.2088
	$MAPE$	0.2015	0.2163	0.2193	0.2104	0.2223	0.1944
Las Vegas NWP							
12×5 min.	$nRMSE$	0.1559 _(cln)	0.1477 _(cln)	0.1569	0.1085	0.1196	0.1196
	$MAPE$	1.2492	1.1072	1.4673	0.1387	0.0769	0.0975
12×30 min.	$nRMSE$	0.1877 _(clk)	0.1895 _(clk)	0.2041	0.1418	0.1577	0.1555
	$MAPE$	2.5590	2.0023	4.0857	0.7209	0.5006	0.2128
3×1 hora	$nRMSE$	0.1821 _(cln)	0.1817 _(clk)	0.1974	0.1406	0.1576	0.1508
	$MAPE$	2.8414	2.7270	3.6506	1.1840	0.5348	0.2322
6×1 hora	$nRMSE$	0.1997 _(cln)	0.1983 _(clk)	0.2020	0.1535	0.1766	0.1680
	$MAPE$	3.1044	3.5833	2.9408	1.6535	0.7152	0.2885
3×1 dia	$nRMSE$	0.1881 _(ird)	0.1907 _(ird)	0.1863	0.1838	0.2150	0.2195
	$MAPE$	0.1977	0.1982	0.1972	0.1904	0.1959	0.2044

Tabela 6 – Atributos melhor classificados por cada método de seleção.

Tabela 7 – nRMSE e MAPE de todos os modelos e horizontes examinados para a cidade de Denver. Valores decimais representam porcentagens (0.25 = 25%)

Leituras × Frequência		RNN	RNN-sf	RNN-rfe	XGboost	Persist. Céu Claro	Persist. Clareza
Denver							
12×5 min.	<i>nRMSE</i>	0.2093 _(cln)	0.2042 _(clk)	0.2181	0.1548	0.1692	0.1694
	<i>MAPE</i>	2.0559	1.8469	2.6412	0.2331	0.1179	0.1055
12×30 min.	<i>nRMSE</i>	0.2673 _(cln)	0.2692 _(cln)	0.2806	0.2338	0.2531	0.2552
	<i>MAPE</i>	4.8430	5.2348	5.6211	0.6761	0.4872	0.3203
3×1 hora	<i>nRMSE</i>	0.2480 _(clk)	0.2462 _(cln)	0.2592	0.2113	0.2354	0.2374
	<i>MAPE</i>	0.5049	2.9494	2.9868	3.3620	0.5695	0.3842
6×1 hora	<i>nRMSE</i>	0.2806 _(clk)	0.2752 _(cln)	0.2846	0.2487	0.2760	0.2776
	<i>MAPE</i>	2.7975	2.7747	3.1224	0.6655	0.7496	0.4846
3×1 dia	<i>nRMSE</i>	0.2546 _(all)	0.2539 _(all)	0.2554	0.2510	0.3239	0.3186
	<i>MAPE</i>	0.3020	0.2998	0.3002	0.2988	0.3319	0.3154
Denver NWP							
12×5 min.	<i>nRMSE</i>	0.2155 _(cln)	0.2145 _(cln)	0.2549	0.1498	0.1701	0.1699
	<i>MAPE</i>	1.2517	1.1934	1.9852	0.1896	0.1049	0.1167
12×30 min.	<i>nRMSE</i>	0.2937 _(cln)	0.2921 _(all)	0.2950	0.2230	0.2531	0.2545
	<i>MAPE</i>	1.7533	2.1643	2.2075	0.4806	0.4457	0.2949
3×1 hora	<i>nRMSE</i>	0.2547 _(cln)	0.2516 _(clk)	0.2674	0.2063	0.2361	0.2364
	<i>MAPE</i>	1.7666	1.6190	2.1642	0.5079	0.4770	0.3271
6×1 hora	<i>nRMSE</i>	0.2869 _(clk)	0.2819 _(all)	0.3012	0.2408	0.2777	0.2771
	<i>MAPE</i>	1.6331	1.6512	2.6735	0.7012	0.6181	0.4039
3×1 dia	<i>nRMSE</i>	0.2489 _(ird)	0.2501 _(all)	0.2445	0.2411	0.3167	0.3029
	<i>MAPE</i>	0.2871	0.2883	0.2772	0.2754	0.3074	0.2868



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Ciência da Computação

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado do candidato Eduarth Dapper Heinen, realizada em 28/08/2018:

Prof. Dr. Ricardo Augusto Souza Fernandes
UFSCar

Prof. Dr. Amilcar Flamarion Querubini Gonçalves
UFSCar

Prof. Dr. Vinicius Ponte Machado
UFPI

Certifico que a defesa realizou-se com a participação à distância do(s) membro(s) Vinicius Ponte Machado e, depois das arguições e deliberações realizadas, o(s) participante(s) à distância está(ao) de acordo com o conteúdo do parecer da banca examinadora redigido neste relatório de defesa.

Prof. Dr. Ricardo Augusto Souza Fernandes