



UNIVERSIDADE FEDERAL DE OURO PRETO
ESCOLA DE MINAS
COLEGIADO DO CURSO DE ENGENHARIA DE CONTROLE
E AUTOMAÇÃO - CECAU



LARISSA VIANA DA SILVA

ESTUDO DE MODELOS E PLATAFORMA WEB PARA
PREVISÃO DE IRRADIAÇÃO SOLAR

MONOGRAFIA DE GRADUAÇÃO EM ENGENHARIA DE CONTROLE E
AUTOMAÇÃO

Ouro Preto, 2022

LARISSA VIANA DA SILVA

**ESTUDO DE MODELOS E PLATAFORMA WEB PARA
PREVISÃO DE IRRADIAÇÃO SOLAR**

**Monografia apresentada ao Curso de
Engenharia de Controle e Automação
da Universidade Federal de Ouro Preto
como parte dos requisitos para a obten-
ção do Grau de Engenheiro de Controle
e Automação.**

Orientador: Prof. Rodrigo Cesar Pedrosa Silva, Ph.D.

Coorientador: Prof. Agnaldo J. R. Reis, Dr.Sc.

**Ouro Preto
Escola de Minas – UFOP
2022**

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

S586e Silva, Larissa Viana da.
Estudo de modelos e plataforma web para previsão de irradiação solar.
[manuscrito] / Larissa Viana da Silva. - 2022.
45 f.: il.: color., gráf., tab..

Orientador: Prof. Dr. Rodrigo Cesar Pedrosa Silva.

Coorientador: Prof. Dr. Agnaldo J. R. Reis.

Monografia (Bacharelado). Universidade Federal de Ouro Preto. Escola de Minas. Graduação em Engenharia de Controle e Automação .

Área de Concentração: Engenharia de Controle e Automação.

1. Fontes renovável. 2. Geração de energia fotovoltaica. 3. Tecnologia.
4. Estatística. 5. Previsão. I. Reis, Agnaldo J. R.. II. Silva, Rodrigo Cesar Pedrosa. III. Universidade Federal de Ouro Preto. IV. Título.

CDU 621

Bibliotecário(a) Responsável: Angela Maria Raimundo - SIAPE: 1.644.803



ATA DA SESSÃO DE DEFESA DE TRABALHO DE CONCLUSÃO DE CURSO

No sexto dia do mês de janeiro de dois mil e vinte dois, realizou-se às 08 horas, de forma não presencial, por meio do aplicativo Google Meet (<https://meet.google.com/fnh-jjaq-kys>), a sessão de defesa de Trabalho de Conclusão de Curso da candidata ao grau de Engenharia de Controle e Automação, Larissa Viana da Silva, intitulada “Estudo de Modelos e Plataforma Web para Previsão de Irradiação Solar”. A Banca Examinadora foi constituída por Rodrigo César Pedrosa Silva (Orientador, UFOP/ICEB/DECOM), Agnaldo José da Rocha Reis (Coorientador, UFOP/EM/DECAT), Sílvia Grasiella Moreira Almeida (IFMG/OP) e Marcos Antônio Alves (FITec). O Prof. Rodrigo abriu a sessão agradecendo a participação dos examinadores supracitados e passou a palavra à candidata, que fez a exposição do seu trabalho. Em seguida, foi realizada a arguição pelos examinadores, com a respectiva defesa da candidata. Finalizada a arguição, a Banca Examinadora, sem a presença da candidata, deliberou pela sua **Aprovação**. Nada mais havendo para constar, lavrou-se a presente ata que será assinada eletronicamente pelos orientadores do trabalho via SEI!UFOP em nome de todos os membros da banca.



Documento assinado eletronicamente por **Agnaldo Jose da Rocha Reis, PROFESSOR DE MAGISTERIO SUPERIOR**, em 06/01/2022, às 12:07, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Rodrigo Cesar Pedrosa Silva, PROFESSOR DE MAGISTERIO SUPERIOR**, em 07/01/2022, às 07:46, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0265174** e o código CRC **E3CDD73E**.

*Este trabalho é dedicado à todos que,
de alguma maneira, me deram forças para continuar.*

AGRADECIMENTOS

Primeiramente gostaria de agradecer à Deus, que me acompanhou durante toda esta jornada, me abençoando e me cobrindo com seu imenso amor. Quero agradecer à minha mãe, por ser exemplo de perseverança e por ter me passado valores que jamais esquecerei. Seu amor sempre esteve comigo, e me deu coragem para ir viver meus sonhos. Esta conquista é por nós duas! Agradecer ao meu pai, por me impulsionar a voar voos cada vez mais longos. Você me inspira a ser melhor todos os dias!

Agradecer a toda família Viana e Carvalho, pelo suporte e carinho nesses anos. Em especial tia Preta e tio Juarez, o amor que sinto por vocês é imensurável. Agradecer a família que conquistei em Ouro Preto: a inigualável República Toka. Moradoras, e ex-alunas, que me abraçaram, acolheram e me ensinaram tanto. Graças a vocês hoje eu sou uma mulher melhor. Obrigado por me darem um lugar para chamar de lar.

Agradecer as minhas amigas de Vila-Velha: Renata e Julia. Vocês são meu coração fora do peito. Obrigada por me encorajarem a ir nessa jornada, e me darem um porto seguro para voltar. O amor de vocês foi luz no meu caminho. Agradecer ao Alan, que sempre esteve comigo, independente da distância. Te levarei comigo pra onde for. Agradecer ao Igor, irmão que a vida me deu. Obrigada por ser tão companheiro, e sempre torcer por mim.

Quero agradecer também ao meu orientador, o Prof. Rodrigo, que de forma tão genuína abriu meus olhos para os milhares de caminhos que a vida pode tomar. Você é exemplo de profissional e de pessoa. Tenho muito orgulho de ter dividido essa pesquisa com você. Obrigada por tanto.

Por fim, agradecer à Universidade Federal de Ouro Preto, pelo ensino gratuito e de qualidade. Por me capacitar com tanta excelência e me por me dar todo o suporte para concluir minha graduação. À todos os professores e técnicos: muito obrigada! Vocês fizeram a diferença na minha vida.

*“Por vezes sentimos que aquilo que fazemos não é senão uma gota de água no mar.
Mas o mar seria menor se lhe faltasse uma gota.” (Madre Teresa de Calcutá)*

RESUMO

Com a crescente demanda por recursos energéticos e a urgência em se reduzir o aquecimento global, amplia-se a necessidade de meios de produção de energia renováveis, capazes de fornecer de maneira eficiente à população a energia necessária às suas atividades. Neste cenário, a energia fotovoltaica se mostra uma promissora alternativa, devido à seu grande potencial e disponibilidade. Entretanto, a irradiância recebida na Terra é variável, fazendo com que a produção de energia fotovoltaica seja intermitente. Esta característica ameaça o gerenciamento e a distribuição de energia, pois não há garantias de estabilidade e confiança, uma vez que os sistemas fotovoltaicos se mostram incapazes de fornecer a energia requerida pela rede a todo momento.

Assim, realizar previsões acuradas da produção de energia viabiliza sua integração à rede, permitindo que todo seu potencial seja explorado. Dessa forma, os algoritmos ARIMA, SARIMA e SARIMAX foram analisados utilizando a velocidade do vento como variável exógena na previsão de radiação solar, que é o fator preponderante na produção de energia fotovoltaica. Além disso, foi proposto um método para a identificação rápida dos parâmetros de controle dos algoritmos, e um sistema web, denominado WindSun, no qual os modelos são disponibilizados ao público. Os resultados obtidos apontaram que o algoritmo SARIMA obteve resultados superiores aos demais modelos candidatos devido ao componente sazonal marcante nas séries de radiação solar. A heurística proposta se mostrou eficaz e produziu resultados melhores tanto na previsão quanto no tempo de treinamento. A velocidade do vento não se mostrou um variável exógena tão interessante para este problema, uma vez que não foi capaz de atribuir uma melhora significativa à precisão das previsões. Em trabalhos futuros este é um aspecto que deve ser analisado.

Palavras-chaves: fontes renováveis; fotovoltaica, eficiência energética; novas tecnologias; estatística; previsão.

ABSTRACT

With the growing demand for energy resources and the urgency to reduce global warming, the need for renewable energy production means, capable of efficiently supplying the population with the energy necessary for their activities, increases. In this scenario, photovoltaic energy is a promising alternative, due to its great potential and availability. However, the irradiance received on Earth is variable, causing the production of photovoltaic energy to be intermittent. This feature threatens energy management and distribution, as there is no guarantee of stability and reliability, since photovoltaic systems are unable to supply the energy required by the grid at all times.

Thus, making accurate forecasts of energy production enables its integration into the grid, allowing its full potential to be explored. In this work, the ARIMA, SARIMA and SARIMAX algorithms are analyzed using wind speed as an exogenous variable in the prediction of solar radiation, which is the preponderant factor in the production of photovoltaic energy. In addition, a method for the rapid identification of the control parameters of the algorithms is proposed, and a web system, called WindSun, in which the models are made available to the public. The results obtained showed that the SARIMA algorithm obtained better results than the other candidate models due to the marked seasonal component in the solar radiation series. The proposed heuristic proved to be effective and produced better results in both prediction and training time. Wind speed was not such an interesting exogenous variable for this problem, as it was not able to attribute a significant improvement to forecast accuracy. In future work, this is an aspect that should be analyzed.

Key-words: renewable sources; photovoltaics, energy efficiency; new technologies; statistic; forecast.

LISTA DE ILUSTRAÇÕES

Figura 1 – Histograma da radiação solar	29
Figura 2 – Histograma da velocidade do vento	29
Figura 3 – Radiação solar durante uma semana	29
Figura 4 – Velocidade do vento durante uma semana	29
Figura 5 – Gráfico de Auto-Correlação	33
Figura 6 – Gráfico do PACF	33
Figura 7 – Página inicial do site que contém informações sobre o projeto. . .	37
Figura 8 – Página de mapa onde o usuário pode navegar entre as estações. . .	38
Figura 9 – Página de gráficos que são gerados através da escolha das variáveis, do horizonte de previsão e do modelo.	38
Figura 10 – Previsões de uma semana do método SARIMA (AC)	41
Figura 11 – Previsões de uma semana do método SARIMAX (AC)	41
Figura 12 – Previsões de uma semana do método SARIMAX (AUTO)	41

LISTA DE TABELAS

- Tabela 1 – 12 horas à frente - ARIMA(AUTO): (2,1,4), ARIMA(AC): (24,0,24),
SARIMA(AUTO):(3,0,1),(0,0,3,4), SARIMA(AC): (2,0,2),(1,0,1,24) 40
- Tabela 2 – 24 horas à frente - ARIMA(AUTO): (2,1,4), ARIMA(AC): (24,0,24),
SARIMA(AUTO):(3,0,1),(0,0,3,4), SARIMA(AC): (2,0,2),(1,0,1,24) 40
- Tabela 3 – 168 horas à frente - ARIMA(AUTO): (2,1,4), ARIMA(AC): (24,0,24),
SARIMA(AUTO):(3,0,1),(0,0,3,4), SARIMA(AC): (2,0,2),(1,0,1,24) 40

LISTA DE ABREVIATURAS E SIGLAS

AR	Autoregressive
MA	Moving Average
ARIMA	Autoregressive Integrated Moving Average
SARIMA	Seasonal Autoregressive Integrated Moving Average
SARIMAX	Seasonal ARIMA with EXogenous variables
AC	Auto-Correlação
AUTO	Auto-ARIMA
AC	Auto-Correlação
AIC	Akaike Information Criterion
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MDA	Mean Directional Accuracy
INMET	Instituto Nacional de Meteorologia
ADF	Augmented Dickey Fuller Test
KPSS	Kwiatkowski – Phillips – Schmidt – Shin Test
API	Interface de Programação de Aplicações
ACF	Função de Autocorrelação
PACF	Função de Autocorrelação Parcial
UFOP	Universidade Federal de Ouro Preto

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Revisão da Literatura	16
1.1.1	<i>Métodos regressivos</i>	16
1.1.2	<i>Técnicas de inteligência artificial</i>	17
1.1.3	<i>Métodos híbridos</i>	18
1.1.4	<i>Modelos estatísticos</i>	18
2	MÉTODOS DE PREVISÃO EM SÉRIES TEMPORAIS E MÉTODOS DE SELEÇÃO DE PARÂMETROS	20
2.1	Métodos de previsão	20
2.1.1	<i>ARIMA</i>	20
2.1.2	<i>SARIMA</i>	21
2.1.3	<i>SARIMAX</i>	22
2.2	Métodos para Seleção de Parâmetros	22
2.2.1	<i>Auto-ARIMA</i>	23
2.2.2	<i>Identificação de Parâmetros baseada na Autocorrelação</i>	23
3	TRATAMENTO E ANÁLISE DOS DADOS	27
3.1	A base de dados	27
3.2	Análise gráfica dos dados	28
4	PROJETO EXPERIMENTAL	30
4.1	Sistema Computacional	30
4.2	Definição dos parâmetros	30
4.3	Treinamento dos algoritmos	33
4.4	Métricas de comparação de séries temporais	34
4.5	Implementação dos métodos	35
5	SISTEMA WEB PARA APRESENTAÇÃO DOS DADOS E PREVISÕES	36
5.1	WindSun: Previsão meteorológica através dos algoritmos	36
6	RESULTADOS	39

7	CONCLUSÃO	42
7.1	Trabalhos Futuros	42
	REFERÊNCIAS	44

1 INTRODUÇÃO

Ao longo dos anos tem-se notado um aumento no consumo de energia elétrica proveniente principalmente da industrialização, modernização e crescimento populacional, além do aumento do padrão de vida das pessoas (DAS et al., 2018). Em paralelo, o aumento de emissão de poluentes também pode ser observado, resultante da queima de combustíveis fósseis. Esta crescente problemática tem causado alarde entre os dirigentes de todo o mundo e ganhado atenção especial nas discussões acerca de pautas climáticas (INATOMI; UDAETA, 2005). Durante a Conferência das Nações Unidas sobre Mudança Climática (COP21), em 2015, a maioria dos países concordaram em fazer um esforço para limitar o aquecimento global (BARBIERI; RAJAKARUNA; GHOSH, 2017), a fim de garantir um planeta mais sustentável para as futuras gerações.

Neste sentido, estima-se que em 2030, 30% de toda a energia produzida será a partir de fontes renováveis e que atingirá 100% em 2050 (DAS et al., 2018). Este cenário se apresenta favorável à energia solar, pois entre os recursos renováveis de energia, a solar está entre as fontes mais aceitáveis e promissoras devido ao seu potencial e disponibilidade.

Nos últimos 10 anos, a capacidade global de energia fotovoltaica instalada aumentou de 5,1 para 227,0 GW (DAS et al., 2018). Espera-se que esse crescimento continue a uma taxa semelhante ou superior no futuro devido ao seu benefício ecológico (DAS et al., 2018).

Em paralelo, a participação solar vem aumentando anualmente no mercado de eletricidade, sendo 7,8% na Itália e perto de 6,5% na Alemanha e na Grécia. Além disso, 22 países têm uma porcentagem superior a 1% (AHMED; KHALID, 2019). De acordo com a projeção mais alta da Agência Internacional de Energia (AIE), a instalação da capacidade global de energia fotovoltaica solar poderá exceder 1700 GW até 2030 (DAS et al., 2018). Assim, os sistemas fotovoltaicos vêm conquistando aceitação mundial e estão desempenhando um papel significativo no fornecimento de energia limpa e sustentável, registrando um enorme crescimento nos últimos anos.

Embora apresente diversos benefícios econômicos, e socioambientais, a produção de energia solar possui a intermitência como característica marcante, decorrente de fatores como: (i) sazonalidade, (ii) variações diurnas decorrentes do movimento de rotação da terra e (iii) a vários processos meteorológicos que ocorrem na atmosfera, em

particular o movimento estocástico das nuvens (MEER; WIDÉN; MUNKHAMMAR, 2018). Ainda há outros fatores que também contribuem com esta variação, como pressão e velocidade do vento, temperatura do ambiente e umidade relativa do ar (DAS et al., 2018).

Desta maneira, a característica intermitente dos sistemas fotovoltaicos, provocada pelos fatores ambientais e meteorológicos supracitados afeta diretamente o comportamento da energia em sua saída. O que a torna uma fonte insegura de produção e distribuição de energia, dificultando sua integração à rede (BARBIERI; RAJAKARUNA; GHOSH, 2017). Estudos apontam que o aumento da inserção da energia fotovoltaica podem levar a problemas de flutuação de tensão, perturbação no carregamento dos alimentadores (*feederloading*), perdas de potência da rede e aumento da corrente de curto-circuito (MEER; WIDÉN; MUNKHAMMAR, 2018).

Esta integração se torna uma questão importante quando se deseja que os sistemas fotovoltaicos convertam-se em um dos principais meios de produção energética, e que alcancem uma alta penetração nas redes de distribuição (BARBIERI; RAJAKARUNA; GHOSH, 2017). Contudo, para que isso ocorra, é necessário que hajam garantias de estabilidade e confiança, onde os sistemas fotovoltaicos devem ser capazes de fornecer a energia requerida pela rede.

Há ainda o aumento do custo final da energia, uma vez que os recursos nem sempre estão disponíveis ou não são garantidos, sendo necessários geradores de fornecimento de energia alternativos capazes de acomodar a subprodução (NOTTON et al., 2018).

Assim, com o objetivo de mitigar este problema, tem-se realizado previsões capazes de determinar com precisão a produção de energia naquele instante (NOTTON et al., 2018), atribuindo maior segurança a rede elétrica, sendo capaz de viabilizar a continuidade do serviço, e garantindo que chegue ao consumidor a energia requerida, sem quaisquer falhas.

Com a previsão, o operador da rede é capaz de antecipar suas ações, sabendo previamente qual será a condição de operação e, portanto, ter maior capacidade para tomar decisões acertadas, obtendo menores custos operacionais.

Além disso, há a melhora no gerenciamento do balanceamento elétrico, realizando-se o controle ideal dos fluxos de energia que ocorrem dentro do sistema fotovoltaico, sendo capaz de estimar reservas, programar o sistema de energia, gerenciar da melhor maneira eventuais congestionamentos e ser capaz de negociar preços no mercado de

energia (NOTTON et al., 2018).

Por fim, o sistema passa a ser mais confiável, uma vez que com a previsão, é possível lidar melhor com situações esperadas, inesperadas (tanto em produção quanto em demanda) e distúrbios, enquanto mantém-se o serviço com qualidade e continuidade aos consumidores (NOTTON, 2015).

Para que tudo isso ocorra, é necessário que a previsão seja precisa, pois um pequeno erro induz ao menos dois efeitos negativos: operador de rede pode receber altas penalidades porque a previsão imprecisa não permitiu alcançar o perfil de produção previsto, o que gera desconfiança. E o uso de geradores de energia alternativos se tornam essenciais, sendo requisitados para compensar a diferença entre a produção prevista e a real, o que eleva os custos de produção (NOTTON et al., 2018).

Novas técnicas e abordagens surgem em todo o mundo a cada ano para melhorar a precisão das previsões e reduzir os custos computacionais envolvidos no processo. A previsão de produção de energia solar é um campo ativo de pesquisa, especialmente para curto prazo (AHMED; KHALID, 2019). Entretanto, ainda há a necessidade de aprofundamento nas pesquisas afim de aumentar a acessibilidade dos modelos ao público geral, melhorar a precisão das previsões, principalmente se atreladas a um baixo custo computacional e transparência nos modelos.

1.1 Revisão da Literatura

Os estudos relacionados com a previsão da produção de plantas fotovoltaicas estão concentrados principalmente em dois campos: na previsão indireta e direta. Nas previsões indiretas, primeiro é prevista a irradiação solar e, em seguida, usado um modelo analítico de desempenho fotovoltaico da planta, responsável por obter a energia produzida. Já na previsão direta, calcula-se diretamente a produção de energia da planta e os modelos são construídos sobre esta variável (ANTONANZAS et al., 2016).

A seguir são descritas as principais técnicas de previsão encontradas na literatura analisada.

1.1.1 Métodos regressivos

Os métodos regressivos estimam a relação entre uma variável dependente (no caso deste estudo a irradiação solar) e outras variáveis, denominadas preditoras (ANTONANZAS et al., 2016). E podem ser subdivididos conforme descrito a seguir:

- Modelos lineares estacionários: englobam os modelos Auto-Regressivos (*Auto-Regressive (AR)*) (BACHER; MADSEN; NIELSEN, 2009), onde a sua saída é uma combinação linear dos valores passados dos preditores. O modelo de Médias Móveis (*Moving Average (MA)*) (LI; SU; SHU, 2014), usado quando os dados apresentam uma variação constante em torno da média, onde a média dos dados anteriores é usada como resultado da previsão. E o modelo Auto-Regressivo de Médias Móveis (*Auto-Regressive Moving Average (ARMA)*), que leva em consideração os valores passados defasados e os erros (BACHER; MADSEN; NIELSEN, 2009).
- Modelos lineares não-estacionários: aqui são encontrados o modelo Auto-Regressivo Integrado de Médias Móveis (*Auto-Regressive Integrated Moving Average (ARIMA)*), que é a união do modelo AR e MA (PEDRO; COIMBRA, 2012), descritos acima. E do modelo ARIMA sazonal, o SARIMA, que introduz um componente sazonal ao modelo (BOUZERDOUM; MELLIT; PAVAN, 2013).

1.1.2 Técnicas de inteligência artificial

- Redes neurais artificiais (Artificial Neural Networks): redes neurais artificiais é a técnica mais utilizada na previsão de energia solar (ANTONANZAS et al., 2016). Sua construção é baseada no comportamento biológico humano, onde um grupo de neurônios está interconectado para formar uma rede neural.
- K-vizinhos mais próximos (k-Nearest Neighbors (k-NN)): este é um dos métodos mais simples de aprendizado de máquina. Consiste em um algoritmo para reconhecimento de padrão, que compara o estado atual com conjuntos de treinamento em um espaço de tempo futuro (PEDRO; COIMBRA, 2012). A ideia é usar vizinhanças ou casos vizinhos como preditores para classificar um caso particular. Os k vizinhos mais próximos funcionam como um exemplo ao redor de um caso específico (FERRERO, 2009).
- Máquina de vetores de suporte (Support Vector Machine (SVM)): destacam-se pela forte capacidade de generalização e pela habilidade de lidar com problemas não lineares, funcionando como uma regressão linear múltipla usando preditores transformados, mas mantendo baixa complexidade e um bom ajuste de dados (ANTONANZAS et al., 2016).
- Floresta aleatória (Random Forest): este método consiste em um conjunto de

árvores de decisão, cujo resultado mostram a previsão média das árvores individuais (ANTONANZAS et al., 2016).

1.1.3 Métodos híbridos

Como cada modelo tem sua forma particular de lidar com os dados, pode ser que durante o processo de análise alguma informação se perca. Então, é interessante que se combine diferentes técnicas de previsão para fomentar os pontos fortes de cada método, visando melhorar a precisão das previsões. A combinação de dois ou mais modelos dá origem a um novo modelo, denominado modelo híbrido (ANTONANZAS et al., 2016).

Duas abordagens podem ser seguidas, seja combinando duas ou mais técnicas estatísticas (híbrido-estatístico), ou juntando uma técnica estatística a um modelo de desempenho fotovoltaico (híbrido-físico). Diversos estudos são encontrados na literatura utilizando o modelo híbrido-físico (ANTONANZAS et al., 2016), que combinam o método SARIMA com o SVM nas previsões horas à frente (BOUZERDOUM; MELLIT; PAVAN, 2013). Ou que utilizaram ARIMA com redes neurais, aplicados a sistemas fotovoltaicos (RAMSAMI; OREE, 2015).

1.1.4 Modelos estatísticos

Os modelos estatísticos utilizam dos dados fornecidos para criar relações entre eles, e serem capazes de prever o futuro, com base nas informações do passado. Dessa forma, não precisam de nenhuma informação interna do sistema para modelá-los (ANTONANZAS et al., 2016). Por outro lado, é necessário ter um volume grande de dados históricos, e que estes possuam qualidade, para que se possa assegurar a precisão das previsões.

A irradiação solar pode ser decomposta como uma série temporal, onde os dados são distribuídos ao longo do tempo. Assim, modelos estatísticos são amplamente utilizados para sua análise e previsão, em particular os métodos da família ARIMA: ARIMA, SARIMA e SARIMAX (ANTONANZAS et al., 2016). Estes são derivações dos conceitos de auto-regressão, média móvel e integração. A sigla ARIMA corresponde a um modelo auto-regressivo integrado de médias móveis, enquanto o SARIMA inclui um componente sazonal para especificar a periodicidade da sazonalidade da série. E o SARIMAX, que se estende ao SARIMA, incluindo uma variável exógena ao modelo (ANTONANZAS et al., 2016).

No processo de produção de energia solar diversos fatores meteorológicos estão

envolvidos além da irradiação, como pressão e velocidade do vento, o movimento das nuvens, a temperatura do ambiente e umidade relativa do ar (ANTONANZAS et al., 2016). Então, para ser capaz de realizar previsões precisas da produção da energia solar, é imprescindível levar ao menos uma dessas variáveis em consideração.

Assim, neste trabalho estudam-se três modelos estatísticos clássicos para o problema de previsão de curto e médio prazo de irradiação solar que é o principal componente relacionado à produção de energia fotovoltaica. Os modelos selecionados foram os clássicos *Autoregressive Integrated Moving Average* (ARIMA), *Seasonal ARIMA* (SARIMA) e SARIMA com variáveis exógenas (SARIMAX).

O uso dos métodos da família ARIMA na previsão se torna interessante, pois são métodos simples, que requerem um custo computacional mais baixo (HYNDMAN; ATHANASOPOULOS, 2018). Seu modelo matemático é pouco complexo, o que torna mais fácil realizar a sua implementação, e há ainda a possibilidade de incluir variáveis exógenas ao modelo, como no caso do SARIMAX.

Entretanto, parte significativa do custo computacional relacionado à construção destes modelos está ligada à determinação de seus parâmetros. Então, para diminuir este custo também foi proposta uma heurística que encontra tais parâmetros rapidamente.

Finalmente, para dar acessibilidade e visibilidade aos resultados da pesquisa, um sistema web denominado WindSun foi construído. Este utiliza a base de dados fornecida pelo INMET (Instituto Nacional de Meteorologia) e permite aos usuários analisar os dados em diferentes estações meteorológicas e visualizar os gráficos de previsões com horizontes de previsão de 1 hora, 1 dia e 1 semana.

Na seção 2 são descritos os métodos de previsão ARIMA, SARIMA e SARIMAX, bem como os métodos para seleção de parâmetros: o Auto-Arima e o método de identificação baseado na autocorrelação. Na seção 3 é descrito o treinamento e análise dos dados. Na sessão 4 tem-se o projeto experimental, que contém as etapas utilizadas no desenvolvimento do estudo. Na seção 5 é apresentado o sistema web construído, o Windsun. Na seção 6 estão descritos os resultados obtidos. E por fim, na seção 7 é feita a conclusão do trabalho.

2 MÉTODOS DE PREVISÃO EM SÉRIES TEMPORAIS E MÉTODOS DE SELEÇÃO DE PARÂMETROS

2.1 Métodos de previsão

2.1.1 ARIMA

Uma série temporal pode ser definida pelo conjunto $\mathcal{Y} = \{Y_0, Y_1, \dots, Y_T\}$ no qual Y_i é o valor da variável de interesse no instante de tempo i e \mathcal{Y} o conjunto de observações que é sempre ordenado pelo tempo das observações.

Neste contexto, os modelos *Autoregressive Integrated Moving Average* (ARIMA) são compostos por três processos:

- Processo auto-regressivo (*AR*): Este processo modela uma relação de dependência entre uma observação e algum número de observações passadas. Formalmente, este processo modela a observação Y_t como:

$$Y_t = \sum_{i=1}^p w_i Y_{t-i} + w_0 \quad (2.1)$$

onde p é um parâmetro do modelo que define quantas observações passadas, ou atrasos (em inglês *lags*) serão utilizados.

- Processo de diferenciação (*I*): Este processo tenta eliminar padrões de tendência e sazonalidade da série de forma a deixá-la estacionária. Para isso, o modelo faz uma transformação da série subtraindo o valor observado no instante de tempo passado da observação corrente. Formalmente, este processo é definido como:

$$Y_t^* = Y_t - Y_{t-1} \quad (2.2)$$

Este processo de diferenciação pode ser repetido sucessivas vezes transformando um série já transformada. O número de execuções sucessivas deste processo é controlado pelo parâmetro d .

- Processo de média móvel (*MA*): Este processo modela uma relação de dependência entre uma observação e o erro residual de uma média móvel aplicada à observações passadas. Formalmente, este processo é definido como:

$$Y_t = \mu + \sum_{i=1}^q \varphi_i \epsilon_{t-i} + \epsilon_0 \quad (2.3)$$

onde μ é a média móvel, ϵ_i é o resíduo no instante tempo i e q controla o número de observações passadas que serão utilizadas neste processo.

Uma vez que os parâmetros p , d e q são definidos, a série é transformada pelo processo de diferenciação e uma observação no tempo t é descrita pela equação do modelo ARIMA abaixo:

$$ARIMA(p, d, q) : Y_t^* = \sum_{i=1}^p w_i Y_{t-i}^* + w_0 + \mu + \sum_{i=1}^q \varphi_i \epsilon_{t-i} + \epsilon_0 \quad (2.4)$$

Para mais detalhes sobre o ARIMA, ver (NIELSEN, 2019a).

2.1.2 SARIMA

O modelo ARIMA não possui suporte para séries temporais com ciclos de repetição, também conhecidas como séries sazonais ou periódicas.

Seasonal Autoregressive Integrated Moving Average (SARIMA), ou ARIMA sazonal, é uma extensão do ARIMA para lidar com o componente sazonal da série. Além dos processos do ARIMA definidos na seção anterior, o SARIMA adiciona três novos parâmetros, P , D e Q , para especificar os processos AR, I e MA para o componente sazonal e um parâmetro adicional S para especificar a periodicidade da sazonalidade. Assim, dado S , podemos especificar os processos sazonais da seguinte forma:

- Processo auto-regressivo sazonal ($AR(S)$):

$$Y_t = \sum_{i=1}^P \varpi_i Y_{t-iS} + w_0 \quad (2.5)$$

- Processo de diferenciação sazonal ($I(S)$):

$$Y_t^* = Y_t - Y_{t-S} \quad (2.6)$$

- Processo de média móvel sazonal ($MA(S)$):

$$Y_t = \mu + \sum_{i=1}^Q \phi_i \epsilon_{t-iS} + \epsilon_0 \quad (2.7)$$

Depois de feitas as diferenciações especificadas pelos parâmetros d e D , definidos os parâmetros p , q , P , Q e S , uma observação no SARIMA é modelada como:

$$SARIMA(p, d, q)(P, D, Q)_S : Y_t^* = \sum_{i=1}^p w_i Y_{t-i}^* + w_0 + \sum_{i=1}^q \varphi_i \epsilon_{t-i} + \epsilon_0 + \sum_{i=1}^P \varpi_i Y_{t-iS}^* + \varpi_0 + \sum_{i=1}^Q \phi_i \epsilon_{t-iS} + \epsilon_0 \quad (2.8)$$

Para mais detalhes sobre o SARIMA, ver (NIELSEN, 2019a).

2.1.3 SARIMAX

O *Seasonal ARIMA with EXogenous variables* (SARIMAX), permite que outras variáveis independentes diferentes da variável de interesse possam ser incorporadas no modelo.

O modelo SARIMAX pode ser definido como:

$$SARIMAX(p, d, q)(P, D, Q)_S : Y_t^* = \lambda X_t \sum_{i=1}^p w_i Y_{t-i}^* + w_0 + \sum_{i=1}^q \varphi_i \epsilon_{t-i} + \epsilon_0 + \sum_{i=1}^P \varpi_i Y_{t-iS}^* + \varpi_0 + \sum_{i=1}^Q \phi_i \epsilon_{t-iS} + \epsilon_0 \quad (2.9)$$

onde, X_t é uma variável exógena, que pode ser estimada, por exemplo, pelo seu próprio ARIMA.

Na aplicação de previsão de irradiação solar, uma variável exógena interessante é a velocidade do vento. Como este fator influencia na movimentação das nuvens, ele pode ser um indicativo de mudanças na irradiação solar recebida pela placa de produção de energia.

2.2 Métodos para Seleção de Parâmetros

Um dos problemas na utilização dos modelos da família ARIMA é a determinação dos parâmetros p , d e q . Este problema se torna ainda mais complexo quando utilizamos modelos SARIMA devido à adição dos parâmetros P , Q , D e S . Uma

solução comum para este problema é utilização do método Auto-ARIMA (SMITH, 2013) que faz uma busca pelo espaço de parâmetros, tanto do ARIMA quando do SARIMA, encontrando a melhor combinação.

Um problema do Auto-ARIMA é o seu alto custo computacional uma vez que ele precisa treinar e avaliar vários modelos para encontrar o melhor conjunto de parâmetros. Por este motivo, os intervalos de busca dos parâmetros p, d, q, P, Q e D não podem ser muito largos para evitar estouro da memória do processo.

Na próxima seção é introduzido o Auto-ARIMA. Posteriormente, na Seção 2.2.2, para eliminar o gargalo computacional causado pelo auto-ARIMA, apresenta-se um método para a identificação de parâmetros baseado no gráfico de auto-correlação.

2.2.1 Auto-ARIMA

Os modelos da família ARIMA contam com diversas bibliotecas disponíveis pela comunidade que possuem recursos que auxiliam em sua implementação. São métodos para realizar testes estatísticos, realizar decomposição sazonal, buscar parâmetros automaticamente, entre outros.

No caso da busca automática por parâmetros, o método em questão é o auto-arima (SMITH, 2013). O auto-arima funciona como uma busca em grade que experimenta diversas combinações de p e q (também P e Q para modelos sazonais), selecionando a combinação que minimiza o AIC (Akaike Information Criterion) (CRITERIO..., 2021), cujo modelo matemático é definido por:

$$AIC = 2k - 2Ln(\hat{L}) \quad (2.10)$$

onde k corresponde ao número de parâmetros do modelo, e \hat{L} é o valor máximo da função de verossimilhança.

Para selecionar o termo d o auto-arima usa um teste de estacionariedade (como o teste Dickey-Fuller) e sazonalidade (como o teste Canova-Hansen) para modelos sazonais (SMITH, 2013). Assim, é possível inferir que o custo computacional deste métodos é alto. Para aliviar este problema, é proposto o método descrito na seção a seguir.

2.2.2 Identificação de Parâmetros baseada na Autocorrelação

Medidas de autocorrelação e autocorrelação parcial são indicadores importantes de quais atrasos são os mais significativos para previsão em uma série temporal

(HYNDMAN; ATHANASOPOULOS, 2018).

Nesta seção, com o objetivo de diminuir o custo computacional da identificação de parâmetros, apresenta-se uma heurística baseada nas medidas de auto-correlação e auto-correlação parcial.

A autocorrelação, ou correlação serial, é a correlação de um sinal com uma cópia atrasada de si mesmo como uma função do atraso. Em sua forma mais fundamental, autocorrelação de uma série temporal é a ideia de que um valor da série em determinado ponto no tempo pode ter correlação com um valor em outro ponto no tempo. Ou seja, é como se os dados em diferentes pontos da série fossem linearmente relacionados uns com os outros em função de sua diferença de tempo (NIELSEN, 2019b). É exatamente esta ideia que o processo autoregressivo (AR) do ARIMA tenta implementar. A autocorrelação, entre o espaço de tempo t_1 e t_2 , pode ser definida como:

$$R_{xx}(t_1, t_2) = E[X_{t_1} * \overline{X_{t_2}}] \quad (2.11)$$

onde E é o operador de valor esperado e a barra representa a conjugação complexa (AUTOCORRELATION, 2019).

Já a autocorrelação parcial de uma série temporal para determinado atraso é a correlação parcial da série com ela mesma em um atraso que contém toda a informação entre dois pontos no tempo. Isso é feito calculando-se um número de correlações condicionais e subtraindo-as da correlação total (NIELSEN, 2019b). Ou seja, a autocorrelação parcial é a correlação de um sinal com uma cópia atrasada de si mesmo como uma função do atraso controlando para outros atrasos. É exatamente esta a ideia implementada pelo processo de média móvel (MA) do ARIMA. A autocorrelação parcial em uma série temporal Z_t , em um atraso k , denotado $\alpha(k)$, é a autocorrelação entre Z_t e Z_{t+k} em que são contabilizados as defasagens de 1 a $k - 1$. Ou seja,

$$\alpha(1) = corr(z_{t+1}, z_t) \quad (2.12)$$

Para $k = 1$.

$$\alpha(k) = corr(z_{t+k} - P_{t,k}, z_t - P_{t,k}(z_t)) \quad (2.13)$$

Para $k \geq 2$.

onde $P_{t,k}$ é o operador sobrejetivo da projeção ortogonal de x no subespaço linear do espaço de Hilbert medido por $z_{t+1}, \dots, z_{t+k-1}$ (PARTIAL... , 2021).

Assim, para o ARIMA, o método pode ser definido da seguinte forma:

1. Para cada $d \in \{0, 1, 2\}$
 - a) Gerar o gráfico de auto-correlação;
 - b) Gerar o gráfico de auto-correlação parcial;
 - c) p é o valor do atraso correspondente ao segundo maior pico da autocorrelação para a série diferenciada, em valores absolutos;
 - d) q é o valor do atraso correspondente ao segundo maior pico da autocorrelação parcial para a série diferenciada, em valores absolutos.
2. Dos modelos gerados para cada valor de d , escolher o melhor em termos de Root Mean Square Error (*RMSE*) no conjunto de dados de treinamento.

Vale ressaltar que embora o maior pico da autocorrelação para a série diferenciada seja o mais representativo, a escolha do segundo maior se deu para que fosse possível obter padrões mais complexos da série, atribuindo ao modelo características particulares dos dados. Este mesmo procedimento vale para a seleção dos parâmetros do SARIMA.

No caso do SARIMA, o modelo ARIMA é estendido com componentes de sazonalidade. Assim, o procedimento para a escolha de seus parâmetros é definido da seguinte forma:

1. Para cada $d \in \{0, 1, 2\}$
 - a) Gerar o gráfico de auto-correlação;
 - b) Gerar o gráfico de auto-correlação parcial;
 - c) p é o valor do atraso correspondente ao segundo maior valor absoluto de autocorrelação;
 - d) q é o valor do atraso correspondente ao segundo maior valor absoluto da autocorrelação parcial;
 - e) S é a periodicidade do componente sazonal extraído da série;
 - f) D é igual a 0 pois uma vez que já exploramos três possibilidades para d , fazer um diferenciação sazonal adicional poderia transformar a série em um passeio aleatório, removendo toda a característica particular dos dados;

- g) P é igual a 1 simplesmente para que o modelo possa explorar o componente sazonal enquanto mantém-se compacto;
 - h) Q é igual a 1 simplesmente para que o modelo possa explorar o componente sazonal enquanto mantém-se compacto.
2. Dos modelos gerados para cada valor de d , escolher o melhor em termos de $RMSE$ no conjunto de dados de treinamento.

Note que o procedimento de escolha dos parâmetros dos modelos ARIMA e SARIMA (consequentemente do SARIMAX) agora tem o custo de criação e treinamento de 3 modelos apenas.

3 TRATAMENTO E ANÁLISE DOS DADOS

3.1 A base de dados

A base de dados utilizada foi construída através dos dados fornecidos pelo INMET (Instituto Nacional de Meteorologia). São disponibilizados dados referentes à precipitação total (mm), pressão atmosférica (mB), radiação global (W/m^2), temperatura do ar ($^{\circ}C$), temperatura máxima e mínima ($^{\circ}C$), temperatura de orvalho máxima e mínima ($^{\circ}C$), umidade relativa máxima e mínima (%), umidade relativa do ar (%), velocidade do vento (m/s) e vento, rajada máxima (m/s).

Os dados são coletados de hora em hora por meio de estações automáticas. Através do site <http://www.inmet.gov.br> é possível fazer o download do pacote de dados anuais de todas as estações automáticas separadas por ano, desde 2000.

Nas etapas de treino e teste dos algoritmos foram utilizados dados de velocidade do vento e radiação da estação da Pampulha - Belo Horizonte. Nossa base de dados contempla 1720 dados coletados entre os anos de 2018 e 2019. Foram necessários alguns ajustes nos dados antes de inseri-los nos algoritmos. A seguir está relatado o passo a passo realizado:

1. Formato da hora refeito: a sigla "UTC" e um acréscimo de zeros desnecessários foram removidos dos dados.
2. Ajuste nos dados de velocidade do vento: substituição do separador "," por "." para que os dados pudessem ser reconhecidos como número. Acréscimo de um zero a esquerda do separador para indicar números com casas decimais. E a inclusão de valores faltantes por zero, para que a série pudesse ser consistente, já que não há essa informação disponível.
3. Ajuste nos dados de radiação: substituição do separador "," por "." para que os dados pudessem ser reconhecidos como número. Substituição dos valores negativos e faltantes por zero.
4. Agrupamento da data e hora: as colunas de data e hora foram agrupadas para facilitar a indexação dos dados.

5. Transformação em *datetime*: a biblioteca *datetime* do Pandas (PANDAS, 2021) permite transformar a string data+hora que foi agrupada no passo anterior em uma variável interpretável pelo compilador como sendo de fato "data+hora".
6. Criação de todas as datas existentes no banco de dados: ao longo do banco de dados haviam dias incompletos, com dados de horas faltantes. Assim, criamos todas as datas e horários existentes entre 2018 e 2019 para verificar quais estavam faltando e inclui-las.
7. Inclusão de datas e horários faltantes e substituição do valor de radiação e velocidade do vento por NaN: o banco de dados foi vasculhado e comparado com o conjunto de datas e horários criado. Os dados faltantes foram incluídos e nos valores foi colocado um NaN. NaN é um valor ou símbolo usado nas linguagens de programação para representar um valor numérico indefinido ou irrepresentável.
8. Dados transformados em série e os valores NaN interpolados: foi criada uma série temporal para os dados de radiação e de velocidade do vento, e os valores NaN foram interpolados para garantir a continuidade da série.
9. Séries exportadas para um arquivo .csv: as séries foram exportadas para um arquivo .csv para que pudessem posteriormente serem utilizadas pelos algoritmos.

Para código do tratamento dos dados, acesse <https://bit.ly/32xslp3>

3.2 Análise gráfica dos dados

Nas figuras 1 e 2 podemos ver a distribuição dos dados de radiação solar e de velocidade do vento, respectivamente. A radiação solar é o fator preponderante na produção de energia fotovoltaica e a velocidade do vento será utilizada como variável exógena com o objetivo de melhorar a precisão das previsões. Note que a radiação solar tem uma mediana de aproximadamente $2000W/m^2$, mas o valor mais comum é, obviamente, 0 uma vez que durante toda a noite não temos radiação solar.

Nas figuras 3 e 4 apresentamos as séries de radiação solar e velocidade do vento no período de uma semana. Este período será o horizonte máximo de previsão nos experimentos.

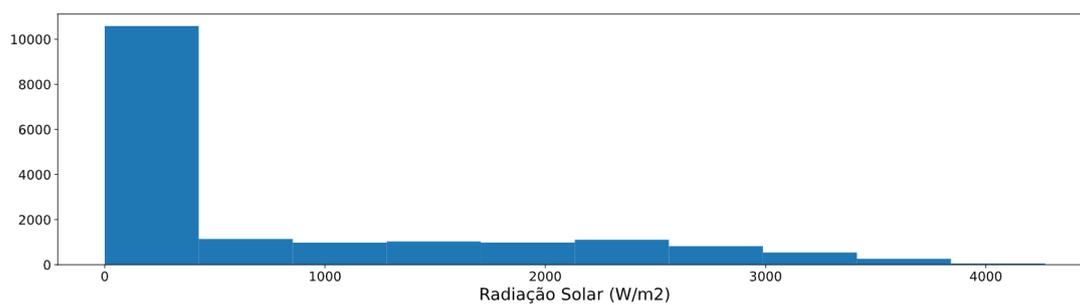


Figura 1 – Histograma da radiação solar

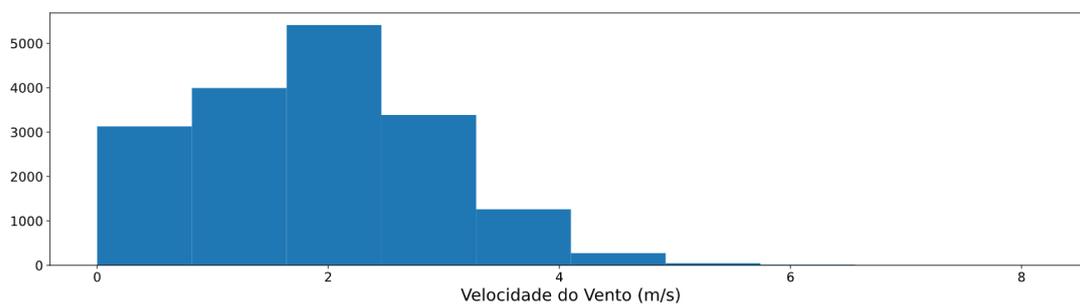


Figura 2 – Histograma da velocidade do vento

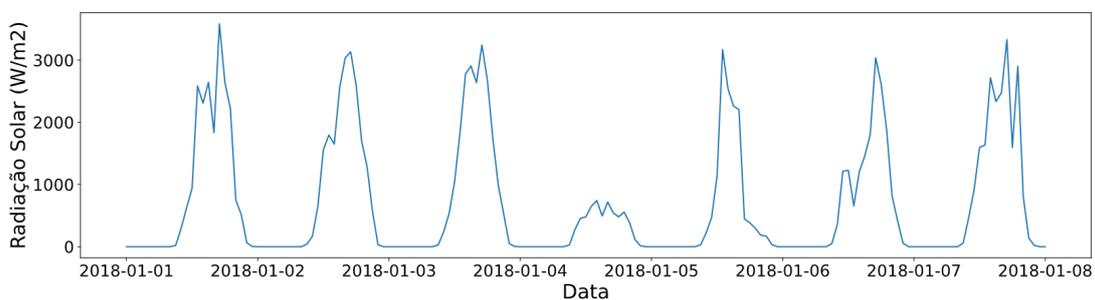


Figura 3 – Radiação solar durante uma semana

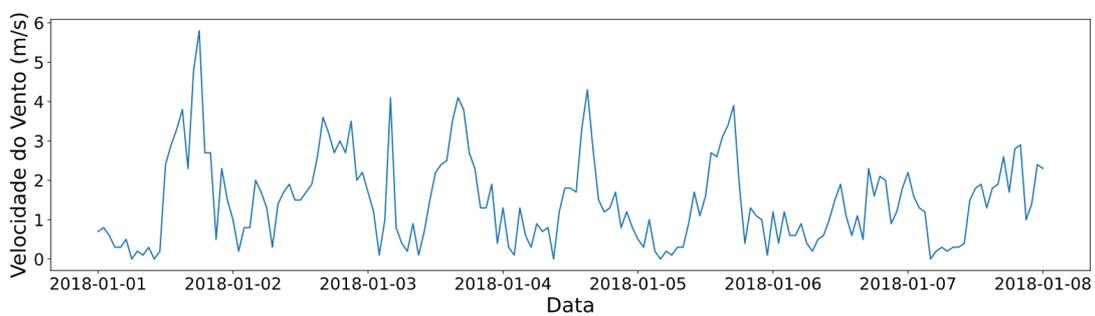


Figura 4 – Velocidade do vento durante uma semana

4 PROJETO EXPERIMENTAL

Como mencionado anteriormente, após o tratamento realizado nos dados, pudemos inserir as séries temporais nos algoritmos. Nas próximas seções estão descritas as etapas realizadas.

4.1 Sistema Computacional

Os experimentos foram executados em um Intel(R) Core(TM) i7-10510U com 16GB de RAM rodando o sistema operacional Windows 10.

4.2 Definição dos parâmetros

Como descrito na seções anteriores, a escolha dos parâmetros é crucial para o desempenho do algoritmo pois eles influenciam diretamente na precisão da previsão e no custo computacional do modelo. Na próxima seção, definimos como o auto-ARIMA foi parametrizado. Em seguida apresentamos a aplicação do método baseado na Autocorrelação.

1. Auto-ARIMA

Para a implementação do auto-ARIMA foi utilizada a API *pmdarima*. A API disponibiliza os testes para estacionariedade e o método auto-ARIMA. Através da documentação é possível analisar todos os parâmetros disponíveis e selecionar os que se adequam ao nosso problema. Para mais informações, acesse: <https://alkaline-ml.com/pmdarima/index.html>.

O primeiro passo foi verificar a estacionariedade para determinar o parametro d . Foram implementados os testes ADF e KPSS disponíveis pela API, e elegemos o resultado do teste com melhor desempenho. É passado ao método *ndiffs*:

- train: dados do conjunto de treino;
- alpha: grau de significância do teste;
- max_d: intervalo de d , que vai de 0 ao valor informado, em que teste deverá ser executado.

Segue código exemplificando o uso:

```

1      kpss_diffs = ndiffs(train, alpha=0.05, test='kpss',
      max_d=6)
      adf_diffs = ndiffs(train, alpha=0.05, test='adf', max_d
      =6)
3      d = max(adf_diffs, kpss_diffs)

```

Em seguida, utilizou-se o método auto-ARIMA definindo os parâmetros que entendemos como necessários para a busca dos parâmetros do ARIMA respeitando os limites de memória do sistema de teste:

- train: dados do conjunto de treino;
- d: o número de diferenciações definido através dos testes ADF e KPSS;
- seasonal: é passado o valor False porque o ARIMA não admite sazonalidade;
- stepwise: quando passado o valor True é feita uma otimização na busca dos parâmetros que diminui o tempo de processamento e é menos propenso ao sobreajuste do modelo;
- max_p: intervalo de p , que vai de 0 ao valor informado, em que o teste deverá ser executado;
- max_q: intervalo de q , que vai de 0 ao valor informado, em que o teste deverá ser executado.

Segue código exemplificando o uso:

```

1      auto = pm.auto_arima(train, d=d, seasonal=False,
      stepwise=True, max_p=12, max_q=12)
      resultado = auto.order

```

É importante ressaltar que não foi possível utilizar o auto-ARIMA com valores maiores de max_p e max_q por conta dos limites de memória impostos pelo sistema operacional.

Para a busca dos parâmetros do SARIMA o método é o mesmo, só que utilizando outros parâmetros do auto-ARIMA:

- train: dados do conjunto de treino;
- d: o número de diferenciações definido através dos testes ADF e KPSS;

- m : período para diferenciação sazonal, referindo-se ao número de períodos em cada estação. Para definir o valor de m é preciso entender o comportamento da série, como a sazonalidade age nos dados;
- `seasonal`: é passado o valor `True`, pois o SARIMA admite sazonalidade;
- `stepwise`: quando passado o valor `True` é feita uma otimização na busca dos parâmetros que diminui o tempo de processamento e é menos propenso ao sobreajuste do modelo.
- `max_p`: serão testados todas as combinações de p no intervalo de 0 ao valor informado;
- `max_q`: serão testados todas as combinações de q no intervalo de 0 ao valor informado;
- `max_P`: serão testados todas as combinações de P no intervalo de 0 ao valor informado;
- `max_Q`: serão testados todas as combinações de Q no intervalo de 0 ao valor informado;

```

2 auto = pm.auto_arima(train, d=d, m=4, seasonal=True,
    stepwise=True, max_p=12, max_q=12, max_P=12, max_Q
    =12)
    resultado = auto. + auto.seasonal_order

```

Como o `seasonal = True` e não foi informado nenhum valor de D , por pré-definição, o parâmetro `seasonal_test` é definido automaticamente como o teste 'OCSB' (Osborn, Chui, Smith, and Birchenhall). E informa ao `auto_arima` qual é o valor estimado de D .

2. Análise dos gráficos de autocorrelação (ACF) e autocorrelação parcial (PACF)

O gráfico do ACF determina valor do termo p . O ACF (ver figura 5) informa quais os atrasos mais fortemente correlacionados com um valor observado.

O gráfico de PACF (ver figura 6) determina o valor do termo q , onde a autocorrelação parcial pode ser vista como a correlação entre a série e sua defasagem, após excluir as contribuições das defasagens intermediárias. Então, podemos dizer que o PACF transmite a correlação pura entre um atraso e a série.

Além destas informações, para determinar os parâmetros dos métodos SARIMA e SARIMAX, como descrito na seção 2.2.2, é necessário conhecer a periodicidade da série que neste estudo foi obtida por inspeção visual do gráfico apresentado na figura 1 que mostra ciclos de 24h em 24h.

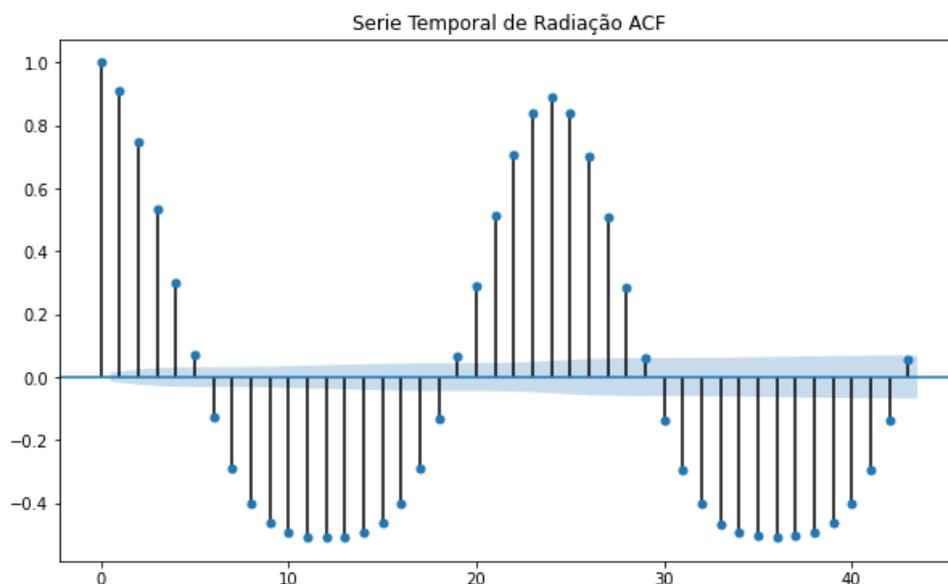


Figura 5 – Gráfico de Auto-Correlação

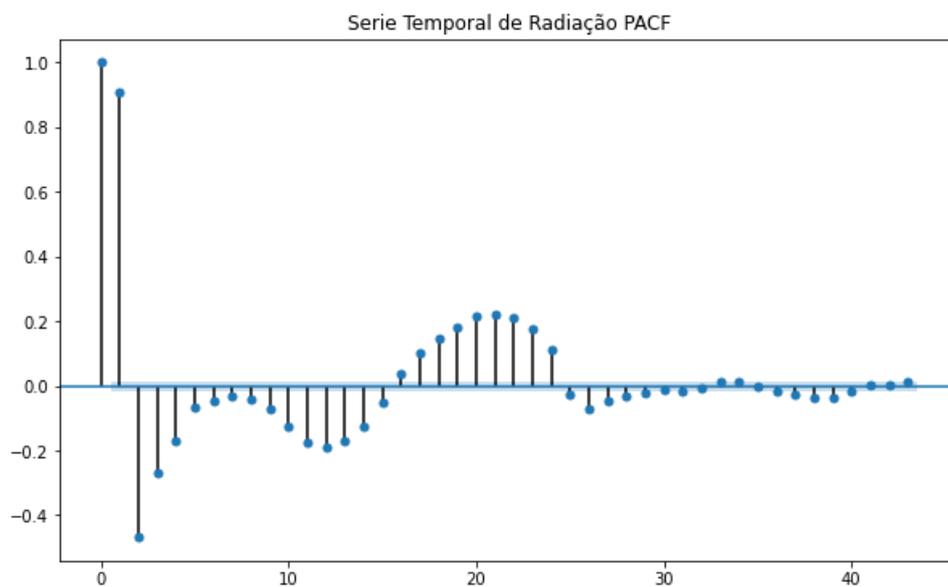


Figura 6 – Gráfico do PACF

4.3 Treinamento dos algoritmos

Para análise do desempenho dos métodos separou-se os dados em conjuntos de treino e teste; onde o algoritmo é modelado com os dados de treino e a previsão validada com os dados do conjunto de teste.

Começando da data 2018-01-01 foram criados 10 conjuntos de treinamento de

3 meses cada, pois estimou-se ser um volume adequado de dados para o problema, sem haver estouro da capacidade computacional. Os conjuntos de dados de teste consistiram dos sete dias que seguem do fim do conjunto de treinamento.

Os algoritmos foram treinados com cada conjunto treinamento e as previsões foram realizadas nos horizontes de 12 horas, 24 horas (1 dia) e 168 horas (1 semana). Como variável exógena do SARIMAX escolhemos a velocidade do vento. E assim, analisou-se os parâmetros estabelecidos pelo auto-arima e pela análise dos gráficos PACF e ACF, avaliando os resultados através das métricas de erro absoluto médio (MAE), erro absoluto percentual médio (MAPE) e acurácia direcional média (MDA).

4.4 Métricas de comparação de séries temporais

- Erro Absoluto Médio (Mean Absolute Error - MAE)

O erro absoluto médio (MAE) é a medida de erros entre observações que expresam o mesmo fenômeno, como por exemplo, o valor observado versus o previsto. É uma métrica comum para mensurar o erro de previsão em séries temporais. O MAE é calculado como:

$$MAE = \frac{\sum_{t=1}^n |y_t - x_t|}{n} = \frac{\sum_{t=1}^n |\epsilon_t|}{n} \quad (4.1)$$

em que y_t é o valor previsto e x_t é o valor observado. Portanto, temos que $|\epsilon_t| = |y_t - x_t|$ é uma média aritmética dos valores absolutos.

- Erro Absoluto Percentual Médio (Mean absolute percentage error - MAPE)

O erro absoluto percentual médio (MAPE) é a média dos erros percentuais absolutos das previsões. Também é comumente utilizado para mensurar o erro de previsão em séries temporais, mas neste caso resultando em uma porcentagem. O MAPE é calculado como:

$$MAPE = \left(\frac{1}{n} \sum_{t=1}^n \frac{|y_t - x_t|}{|y_t|} \right) * 100 \quad (4.2)$$

em que y_t é o valor observado e x_t é o valor previsto.

- Acurácia Direcional Média (*Mean Directional Accuracy* - MDA)

A acurácia direcional média é uma medida de acurácia da previsão de um método que compara a direção da previsão com a direção realizada pelo valor observado.

Isto é, o MDA fornece a probabilidade que o método de previsão consegue detectar a direção correta da série temporal. O MDA é calculado como:

$$MDA = \frac{1}{n} \sum_t 1_{\text{sgn}(y_t - y_{t-1}) = \text{sgn}(x_t - y_{t-1})} \quad (4.3)$$

Em que y_t é o valor observado, x_t é o valor previsto, 1 é o indicador da função e sgn é a função de sinal.

- Tempo para treinamento: É o tempo necessário para a identificação de parâmetros e ajuste dos pesos das equações lineares dos modelos.
- Tempo para inferência: É o tempo utilizado pelo modelo para produzir as previsões.

4.5 Implementação dos métodos

Após o treino e análise dos resultados, foram definidos os parâmetros que melhor se ajustaram aos dados. Com esses parâmetros estabelecidos, os algoritmos ARIMA, SARIMA e SARIMAX foram implementados no projeto WindSun e disponibilizados para que o público em geral pudesse ter acesso e realizar as previsões.

5 SISTEMA WEB PARA APRESENTAÇÃO DOS DADOS E PREVISÕES

5.1 WindSun: Previsão meteorológica através dos algoritmos

Como forma de divulgação dos métodos que foram implementados, e para ampliar o alcance do conhecimento científico, foi criado um site denominado *WindSun*. O sistema tem como objetivo mostrar ao usuário de maneira simples algumas informações de interesse sobre energias renováveis. São exibidos dados sobre a velocidade do vento e radiação solar em determinadas localidades, gerando previsões destas variáveis.

O site foi desenvolvido em parceria com o TerraLab, que é um laboratório de pesquisa e desenvolvimento do departamento de ciência da computação da UFOP, atuando como um programa para capacitação de recursos humanos destinados à implementação de softwares como serviço e datascience.

Para a construção do sistema, foi desenvolvido um documento de especificação de requisitos, responsável por determinar toda as regras de negócio relacionadas ao WindSun. Segundo (MACHADO, 2018), fazer um bom levantamento e uma especificação de requisitos é primordial para quem trabalha com o desenvolvimento de sistemas. Este levantamento pode não cobrir todas necessidades do sistema inicialmente, mas tende a antecipar eventuais falhas e erros de produtos de software.

Além disso, para guiar o desenvolvimento, é importante que haja um layout representando a interface do sistema, que irá determinar a sua aparência final, a tipologia, a paleta de cores, e todos os demais elementos visuais que o compõe. Para tal, utilizou-se a ferramenta Figma, um editor gráfico de vetor e prototipagem de projetos de design online.

No primeiro contato com o site, o usuário tem acesso a uma página com todas as informações sobre o projeto. Lá estão disponibilizadas informações sobre a equipe que atuou no desenvolvimento do WindSun e mais sobre o TerraLab, conforme mostra a figura 7. Em seguida, é exibido um mapa com diversas estações meteorológicas brasileiras onde o usuário pode clicar em um marcador e ter acesso informações sobre velocidade do vento e radiação naquele local, conforme figura 8. Há ainda a possibilidade de visualizar um gráfico com dados históricos e de previsão. O usuário pode escolher a variável, o horizonte de previsão e o método desejados (ver figura 9). Para acesso ao código fonte dos modelos, acesse: <https://github.com/lrsvg/>

TimeSeriesForecastingWindSun. E para acesso ao site, visite: <https://windsun-fronteste.web.app/>.

WindSun

Início Sobre Quem somos Contato

Previsão meteorológica através da inteligência artificial

Visualize e gere gráficos com dados de radiação solar, velocidade do vento e evapotranspiração.

Entrar

Sobre

O que é o WindSun?

O conhecimento mais perto de você
Gráficos com dados do INMET e a possibilidade de realizar previsões

Nossa missão é aprimorar a interação do público com as atividades desenvolvidas dentro da universidade. Dessa maneira, utilizamos um assunto de relevância global - geração de energia por fontes renováveis e limpas - para demonstrar, na prática, o funcionamento de algoritmos de aprendizado de máquina para a previsão de dados meteorológicos.

Dados Meteorológicos Geolocalizados

Como funciona?

Escolha a estação meteorológica Nagueve pelo mapa e selecione a estação do seu interesse

No WindSun, você tem acesso a um mapa em que estão disponíveis estações meteorológicas brasileiras. Selecione a estação de seu interesse e visualize as informações mais recentes sobre radiação, velocidade do vento e evapotranspiração naquele local.

Figura 7 – Página inicial do site que contém informações sobre o projeto.

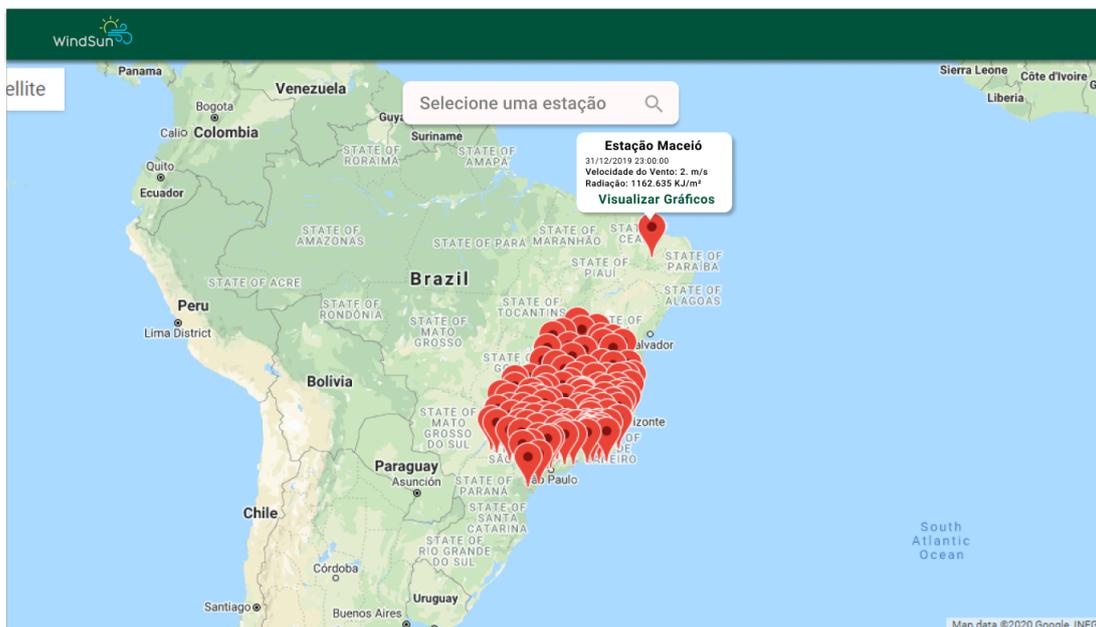


Figura 8 – Página de mapa onde o usuário pode navegar entre as estações.

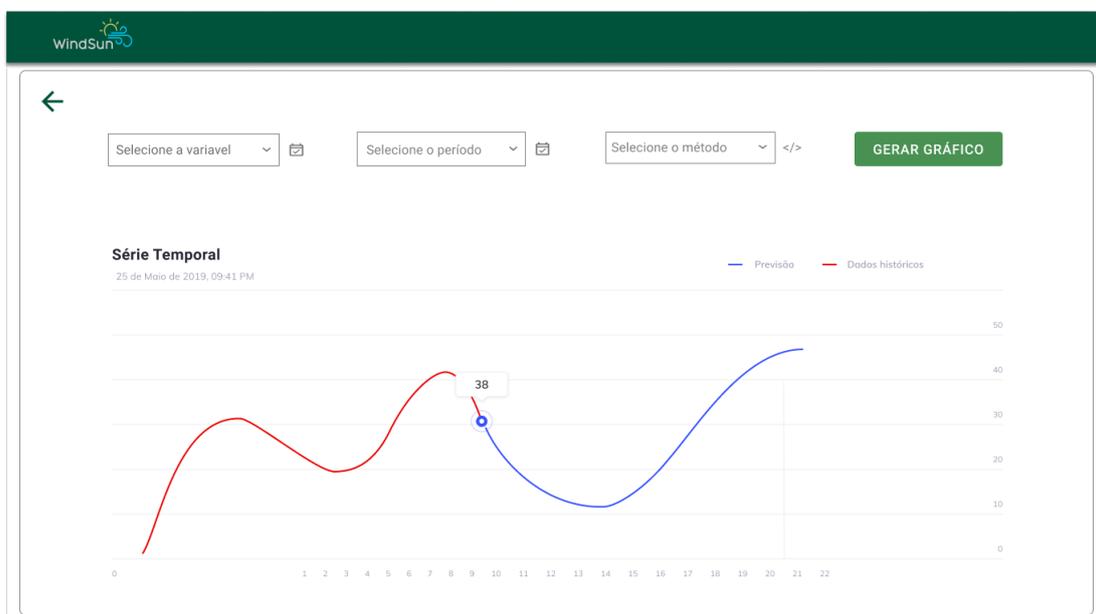


Figura 9 – Página de gráficos que são gerados através da escolha das variáveis, do horizonte de previsão e do modelo.

6 RESULTADOS

As tabelas 1, 2 e 3 apresentam os resultados obtidos pelos métodos, ARIMA, SARIMA e SARIMAX, utilizando o auto-ARIMA (AUTO) e o método baseado nos valores de autocorrelação (AC). Elas apresentam os tempos médios de treinamento e inferência além das medidas de qualidade descritas na seção 4.4 para os 10 pares de conjuntos de treinamento e teste com horizontes de previsão de 12h, 1 dia (24h) e 1 semana (168h).

Pode-se observar que, como esperado, em termos de tempo de treinamento, o auto-ARIMA é mais custoso. Por outro lado, como o tamanho dos modelos gerados por ele é limitado, os métodos que o utilizam possuem tempos de inferência ligeiramente menores. Em relação aos erros e à acurácia direcional, o método AC teve desempenho muito superior. Como este método tem um tempo fixo de treinamento (avalia 3 modelos para determinar os parâmetros) é possível construir modelos muito maiores que, neste problema, se mostraram muito mais adequados.

No caso do método AUTO, como há o gargalo computacional, não é possível permitir que o modelo explore muitas combinações de parâmetros, e com isso, limita-se o tamanho do espaço de busca pela melhor combinação. Portanto, a técnica de definição dos parâmetros que resultou em melhor desempenho, levando-se em conta todas as métricas de qualidade, foi a AC. Se mostrando superior em termos de tempo de treinamento e capacidade de gerar previsões com baixo índice de erro e maior acurácia direcional para todos os horizontes de previsão testados.

Em comparação ao desempenho dos algoritmos ARIMA, SARIMA e SARIMAX na realização da previsão da série temporal, o modelo que se mostrou mais preciso foi o SARIMA, devido ao componente sazonal marcante nas séries de radiação solar. Seus resultados apresentaram os menores erros, e tiveram uma boa acurácia em relação aos demais. É importante ressaltar, contudo, que embora o SARIMAX apresente erros maiores, a adição da velocidade do vento melhorou a acurácia direcional.

As figuras 10, 11, 12, mostram as previsões de uma semana dos métodos SARIMA (AC), SARIMAX (AC) e SARIMAX(AUTO), respectivamente. Pode-se observar claramente na Figura 12 que o método AUTO subestimou os parâmetros e produziu modelos incapazes de reproduzir os padrões da série.

Na figura 10 é possível ver que o SARIMA (AC) captura a estrutura global da

série mas não consegue representar algum detalhes e variações locais. Por outro lado, na Figura 11, podemos notar que a adição da velocidade do vento permite ao modelo representar padrões mais complexos. Apesar na piora do MAE e do RMSE a acurácia direcional melhora ligeiramente. Os tempos de treinamento e inferência foram medidos em segundos.

Método	T_{train}	T_{infer}	MAE	$RMSE$	MDA
ARIMA (AUTO)	5399.66	0.0004	432.96	591.00	43.51
ARIMA (AC)	466,62	0.0006	101.23	213.10	55.73
SARIMA (AUTO)	22506.66	0.0005	187.41	442.35	43.51
SARIMA (AC)	125.79	0.0006	81.85	199.68	64.89
SARIMAX (AUTO)	22523.63	0.0006	184.92	396.26	61.83
SARIMAX (AC)	127.92	0.0007	157.22	287.05	67.94

Tabela 1 – 12 horas à frente - ARIMA(AUTO): (2,1,4), ARIMA(AC): (24,0,24), SARIMA(AUTO):(3,0,1),(0,0,3,4), SARIMA(AC): (2,0,2),(1,0,1,24)

Método	T_{train}	T_{infer}	MAE	$RMSE$	MDA
ARIMA (AUTO)	5399.66	0.0007	678.42	941.38	44.87
ARIMA (AC)	466,62	0.0011	254.59	463.33	61.60
SARIMA (AUTO)	22506.66	0.0009	718.97	1248.17	47.91
SARIMA (AC)	125.79	0.0011	231.53	464.65	67.68
SARIMAX (AUTO)	22523.63	0.0013	629.61	1050.86	55.13
SARIMAX (AC)	127.92	0.0014	277.72	474.16	70.34

Tabela 2 – 24 horas à frente - ARIMA(AUTO): (2,1,4), ARIMA(AC): (24,0,24), SARIMA(AUTO):(3,0,1),(0,0,3,4), SARIMA(AC): (2,0,2),(1,0,1,24)

Método	T_{train}	T_{infer}	MAE	$RMSE$	MDA
ARIMA (AUTO)	5399.66	0.0051	829.24	1062.56	36.87
ARIMA (AC)	466,62	0.0079	264.71	455.65	58.53
SARIMA (AUTO)	22506.66	0.0066	750.78	1286.21	50.08
SARIMA (AC)	125.79	0.0078	214.15	419.69	65.46
SARIMAX (AUTO)	22523.63	0.0089	723.40	1137.08	39.52
SARIMAX (AC)	127.92	0.0099	281.83	462.13	65.51

Tabela 3 – 168 horas à frente - ARIMA(AUTO): (2,1,4), ARIMA(AC): (24,0,24), SARIMA(AUTO):(3,0,1),(0,0,3,4), SARIMA(AC): (2,0,2),(1,0,1,24)

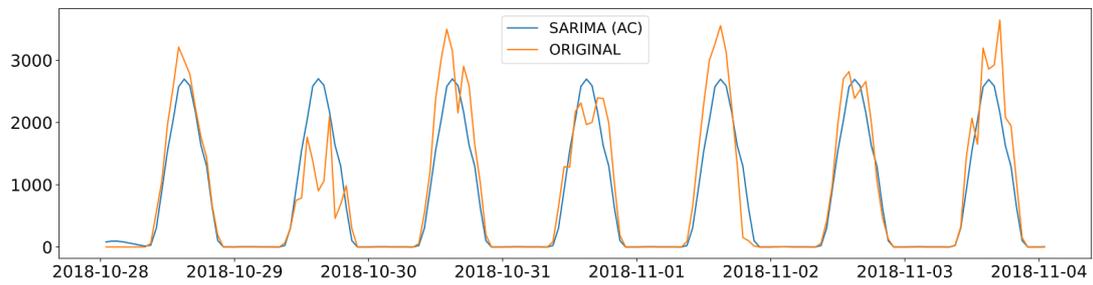


Figura 10 – Previsões de uma semana do método SARIMA (AC)

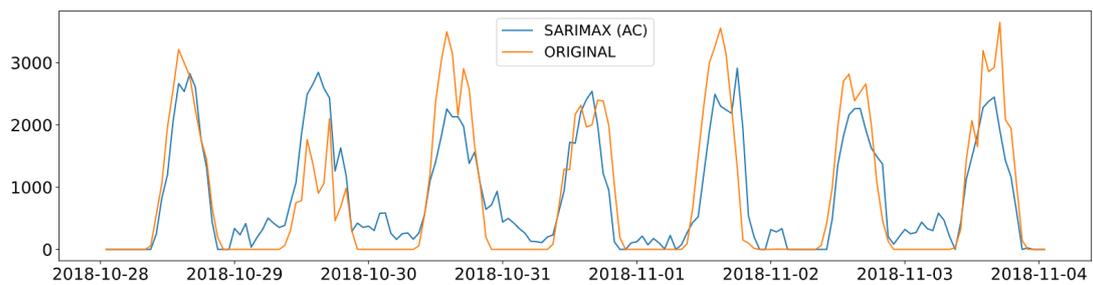


Figura 11 – Previsões de uma semana do método SARIMAX (AC)

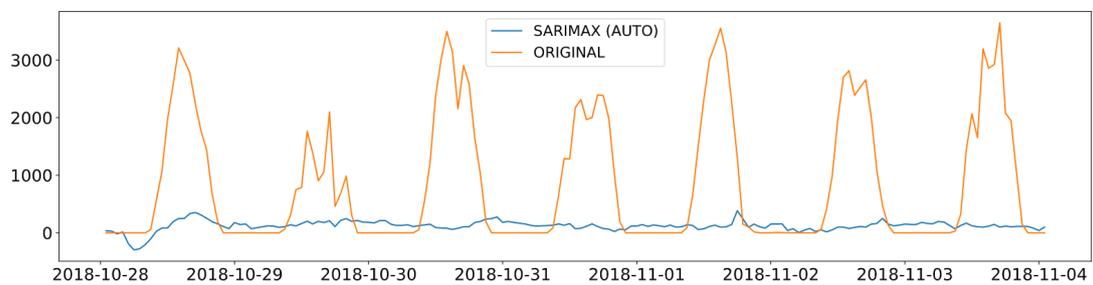


Figura 12 – Previsões de uma semana do método SARIMAX (AUTO)

7 CONCLUSÃO

A crescente busca por energias renováveis se dá graças à urgente necessidade de redução das emissões de gases poluentes, e devido ao crescente aumento no consumo de energia. Dentre as opções, a energia solar se destaca, principalmente por seu potencial e disponibilidade. Entretanto, sua característica intermitente dificulta o gerenciamento da rede de distribuição de energia. Uma solução proposta é realizar a previsão da produção de energia solar, principalmente a curto prazo.

A previsão de energia é um campo de pesquisa ativo, entretanto ainda há a necessidade de aprofundamento. Desta maneira, buscou-se com este trabalho desenvolver os modelos de previsão ARIMA, SARIMA e SARIMAX, de forma a melhorar a precisão das previsões, atrelado à baixos custos computacionais e transparência nos modelos. Foram desenvolvidas duas técnicas diferentes de definição dos parâmetros p , q , d e P , Q e D : o auto-ARIMA e a análise dos gráficos PACF e ACF.

Os resultados mostraram que o algoritmo SARIMA tem resultados superiores devido ao componente sazonal marcante nas séries de radiação solar. O método de identificação de parâmetros baseado na AC se mostrou eficaz e produziu resultados melhores tanto na previsão quanto no tempo de treinamento. O uso da velocidade do vento no modelo SARIMAX, embora tenha atribuído uma acurácia direcional maior, apresentou maiores erros se comparado aos demais. Assim, a velocidade do vento não se mostrou uma variável exógena tão interessante para este problema.

Além disso, foi desenvolvido o sistema *WindSun* para viabilizar a interação do público com os algoritmos desenvolvidos, permitindo maior difusão do conhecimento ao público em geral acerca dos modelos de previsão.

7.1 Trabalhos Futuros

A utilização dos métodos ARIMA e SARIMA para o problemas nos moldes testados aqui esbarraram de forma significativa em limites computacionais. A adição de muito atrasos no modelo se mostrou inviável. Assim, uma possibilidade de trabalhos futuros é a implementação de outros métodos para a comparação do custo computacional versus precisão das previsões.

Além deste aspecto, foi interessante observar que a série de radiação solar tem um padrão global bem definido mas localmente apresenta nuances que são difíceis

de aproximar. Neste contexto, a utilização de mais variáveis exógenas pode melhorar os modelos e a junção de mais variáveis deste tipo devem ser estudadas com mais profundidade.

REFERÊNCIAS

- AHMED, A.; KHALID, M. A review on the selected applications of forecasting models in renewable power systems. *Renewable and Sustainable Energy Reviews*, 2019. Elsevier, v. 100, p. 9–21, 2019. Citado 2 vezes nas páginas 14 e 16.
- ANTONANZAS, J. et al. Review of photovoltaic power forecasting. *Solar Energy*, 2016. Elsevier, v. 136, p. 78–111, 2016. Citado 4 vezes nas páginas 16, 17, 18 e 19.
- AUTOCORRELATION. 2019. <https://perma.cc/U8JY-QD7U>. Acessado em: 09/12/2021. Citado na página 24.
- BACHER, P.; MADSEN, H.; NIELSEN, H. A. Online short-term solar power forecasting. *Solar energy*, 2009. Elsevier, v. 83, n. 10, p. 1772–1783, 2009. Citado na página 17.
- BARBIERI, F.; RAJAKARUNA, S.; GHOSH, A. Very short-term photovoltaic power forecasting with cloud modeling: A review. *Renewable and Sustainable Energy Reviews*, 2017. Elsevier, v. 75, p. 242–263, 2017. Citado 2 vezes nas páginas 14 e 15.
- BOUZERDOUM, M.; MELLIT, A.; PAVAN, A. M. A hybrid model (sarima–svm) for short-term power forecasting of a small-scale grid-connected photovoltaic plant. *Solar Energy*, 2013. Elsevier, v. 98, p. 226–235, 2013. Citado 2 vezes nas páginas 17 e 18.
- CRITERIO de informação de Akaike. 2021. https://pt.wikipedia.org/wiki/CritC3A9rio_de_informaC3A7C3A3o_de_Akaike. Acessado em: 10/12/2021. Citado na página 23.
- DAS, U. K. et al. Forecasting of photovoltaic power generation and model optimization: A review. *Renewable and Sustainable Energy Reviews*, 2018. Elsevier, v. 81, p. 912–928, 2018. Citado 2 vezes nas páginas 14 e 15.
- FERRERO, C. A. *Algoritmo kNN para previsão de dados temporais: funções de previsão e critérios de seleção de vizinhos próximos aplicados a variáveis ambientais em limnologia*. Tese (Doutorado) — Universidade de São Paulo, 2009. Citado na página 17.
- HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting: principles and practice*. [S.l.]: OTexts, 2018. Citado 2 vezes nas páginas 19 e 24.
- INATOMI, T. A. H.; UDAETA, M. E. M. Análise dos impactos ambientais na produção de energia dentro do planejamento integrado de recursos. In: *III Workshop Internacional Brasil-Japão: Implicações Regionais e Globais em Energia, Meio Ambiente e Desenvolvimento Sustentável*. [S.l.: s.n.], 2005. Citado na página 14.
- LI, Y.; SU, Y.; SHU, L. An armax model for forecasting the power output of a grid connected photovoltaic system. *Renewable Energy*, 2014. Elsevier, v. 66, p. 78–89, 2014. Citado na página 17.

MACHADO, F. N. R. *Análise e Gestão de Requisitos de Software—Onde nascem os sistemas*. [S.l.]: Saraiva Educação SA, 2018. Citado na página 36.

MEER, D. W. Van der; WIDÉN, J.; MUNKHAMMAR, J. Review on probabilistic forecasting of photovoltaic power production and electricity consumption. *Renewable and Sustainable Energy Reviews*, 2018. Elsevier, v. 81, p. 1484–1512, 2018. Citado na página 15.

NIELSEN, A. *Practical Time Series Analysis: Prediction with Statistics and Machine Learning*. [S.l.]: O’Reilly Media, Sebastopol, 2019. Citado 2 vezes nas páginas 21 e 22.

NIELSEN, A. *Practical Time Series Analysis: Prediction with Statistics and Machine Learning*. 1a edição. ed. [S.l.]: O’Reilly Media, 2019. Citado na página 24.

NOTTON, G. Importance of islands in renewable energy production and storage: The situation of the french islands. *Renewable and Sustainable Energy Reviews*, 2015. Elsevier, v. 47, p. 260–269, 2015. Citado na página 16.

NOTTON, G. et al. Intermittent and stochastic character of renewable energy sources: Consequences, cost of intermittence and benefit of forecasting. *Renewable and Sustainable Energy Reviews*, 2018. Elsevier, v. 87, p. 96–105, 2018. Citado 2 vezes nas páginas 15 e 16.

PANDAS. 2021. <https://pandas.pydata.org/>. Acessado em: 03/01/2022. Citado na página 28.

PARTIAL autocorrelation function. 2021. https://en.wikipedia.org/wiki/Partial_autocorrelation_function. Acessado em: 09/12/2021. Citado na página 25.

PEDRO, H. T.; COIMBRA, C. F. Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy*, 2012. Elsevier, v. 86, n. 7, p. 2017–2028, 2012. Citado na página 17.

RAMSAMI, P.; OREE, V. A hybrid method for forecasting the energy output of photovoltaic systems. *Energy Conversion and Management*, 2015. Elsevier, v. 95, p. 406–413, 2015. Citado na página 18.

SMITH, T. G. *pmdarima: ARIMA estimators for Python*. 2013. https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.auto_arima.html,. Citado na página 23.