

# Análise de Previsão de Geração Fotovoltaica na Região Metropolitana de Fortaleza Usando Técnicas de Aprendizado de Máquina: Um Estudo de Caso

Leonardo A. Vasconcelos de Oliveira  
*Departamento de Indústria*  
*IFCE - Campus Maracanaú*  
Maracanaú - CE, Brasil  
leonardoadriano.v@gmail.com

Rômulo Cesar Cunha Lima  
*Departamento de Indústria*  
*IFCE - Campus Pecém*  
São Gonçalo do Amarante - CE, Brasil  
romulo.cesar@ifce.edu.br

José Daniel de Alencar Santos  
*Departamento de Indústria*  
*IFCE - Campus Maracanaú*  
Maracanaú - CE, Brasil  
jdaniel@ifce.edu.br

**Abstract**—This paper presents a case study for analysis and prediction of photovoltaic generation in the metropolitan region of Fortaleza-CE, applying machine learning techniques in the paradigms of time series prediction and system identification. Computational simulations were performed using linear (least squares) and non-linear (artificial neural networks and kernel methods) estimators that are part of the state-of-the-art in machine learning. In time series scenario, using only power measurements, the best results were obtained with the MLP network, with a prediction horizon of seven days ahead. In system identification, using power and solar radiation measurements, the least squares estimator achieved the best performance among all tested estimators, even in the free simulation scenario, i.e., infinite steps ahead.

**Index Terms**—Photovoltaic Generation, Electric Power, Machine Learning, Time Series Prediction, System Identification.

**Resumo**—Este artigo apresenta um estudo de caso para análise e previsão de geração fotovoltaica na região metropolitana de Fortaleza-CE, aplicando técnicas de aprendizado de máquina nos paradigmas de predição de séries temporais e identificação de sistemas. Foram realizadas simulações computacionais usando estimadores lineares (mínimos quadrados) e não lineares (redes neurais artificiais e métodos de *kernel*) que fazem parte do estado da arte em aprendizado de máquina. No cenário de séries temporais, usando apenas medições da potência gerada, os melhores resultados obtidos foram com a rede MLP, em um horizonte de predição de sete dias à frente. Já em identificação de sistemas, usando medições de potência e de radiação solar, o estimador de mínimos quadrados alcançou o melhor desempenho dentre todos os estimadores testados, inclusive no cenário de simulação livre, ou seja, de infinitos passos à frente.

**Palavras-chaves**—Geração Fotovoltaica, Potência Elétrica, Aprendizado de Máquina, Predição de Séries Temporais, Identificação de Sistemas.

## I. INTRODUÇÃO

Ao longo das últimas décadas, as fontes renováveis de energia, principalmente solar e eólica, tem se consolidado como importantes alternativas para mitigar os impactos negativos da geração de energia a partir de combustíveis fósseis [1]. Em especial, o crescimento mundial da geração solar é consequência do aumento da eficiência dos painéis fotovoltaicos aliado à redução dos custos de implantação [2]. De acordo

com o relatório global de energia da *Renewable Energy Policy Network for the 21st Century* (REN21) [3], a capacidade de geração fotovoltaica instalada globalmente aumentou em mais de 50 vezes no período de 2007 a 2017, passando de 8GW de potência instalada para 402GW.

Essa tendência de crescimento também tem se mostrado em nível nacional, tendo em vista que o Brasil foi o quarto país no mundo que mais aumentou sua capacidade instalada de geração solar no ano de 2021, segundo a Associação Brasileira de Energia Solar Fotovoltaica (ABSOLAR) [4]. Nesse mesmo ano, especificamente na região Nordeste, geograficamente privilegiada como polo de geração solar, os estados de Bahia, Piauí e Ceará estiveram entre os cinco maiores geradores de energia solar no país [5]. Isso ilustra o papel de destaque do Brasil, e especialmente da região Nordeste, na geração solar para diversificação da matriz energética, promovendo sustentabilidade e impulsionando o desenvolvimento econômico.

A despeito das inúmeras vantagens, a utilização da energia solar também pode apresentar impactos que afetam a estabilidade e a operação do sistema elétrico [6]. Isso ocorre principalmente em função de sua natureza intermitente, altamente sensível e dependente das condições climáticas e sazonalidades [7]. Em outras palavras, seu comportamento é influenciado por fatores como ciclos diários de dias e noites, quadras chuvosas e secas, velocidade do vento, fluxo de radiação solar, temperatura ambiente, dentre outros.

Nesse contexto, a utilização de procedimentos e técnicas para análise e predição da potência gerada em sistemas fotovoltaicos torna-se essencial, no intuito de propor soluções eficazes para as flutuações da potência, tanto em sistemas de larga escala como em unidades menores conectadas à rede de distribuição [8]. Os resultados dessas análises e predições podem ainda ser úteis em aplicações como previsão e detecção de falhas, planejamento de ações de manutenção e diagnósticos de eficiência energética [9].

O uso das ferramentas de Inteligência Artificial (IA) e de Aprendizado de Máquina (*Machine Learning*) tem se mostrado uma abordagem promissora para criação de modelos preditivos precisos que estimam a potência gerada em sistemas foto-

voltaicos, em diferentes paradigmas e condições ambientais. Aplicações diversas de IA podem ser encontradas na perspectiva de predição de séries temporais [8], [10], considerando apenas a potência como única variável analisada, ou na perspectiva de identificação de sistemas [11], [12] considerando, além da potência, variáveis como radiação solar, temperatura e velocidade do vento para as predições de potência gerada.

Diante do exposto, este artigo apresenta um estudo de caso para análise e predição de geração fotovoltaica na região metropolitana de Fortaleza, aplicando ferramentas de aprendizado de máquina baseadas em estimadores lineares, neurais e em métodos de *kernel*, nos paradigmas de predição de séries temporais e de identificação de sistemas. A ideia é que os melhores resultados alcançados venham a servir de suporte para ações de planejamento de manutenção, como também para diagnóstico de eficiência energética.

O restante do artigo está organizado a partir da sequência a seguir. Na Seção II, são abordados o sistema de geração fotovoltaico e o sensor de fluxo de radiação solar utilizados. A Seção III apresenta a metodologia adotada para o desenvolvimento do presente artigo. Já na Seção IV, são apresentados e discutidos os resultados das simulações computacionais. Por fim, o artigo é concluído na Seção V.

## II. GERAÇÃO SOLAR - IFCE MARACANAÚ

O IFCE Maracanaú é um *campus* do Instituto Federal de Educação, Ciências e Tecnologia, situado na cidade de Maracanaú, região metropolitana de Fortaleza. O município está localizado a 24km da capital, possui cerca de 209.057 habitantes e se destaca como o maior polo industrial do estado do Ceará.

### A. Sistema Fotovoltaico

O IFCE Maracanaú possui um sistema próprio de geração fotovoltaica, do tipo *on-grid*, instalado desde meados de 2020. As placas do sistema são do modelo JAM72S01-390/PR, com capacidade nominal para fornecer até 390W de potência ativa de pico.



Figura 1: Vista superior dos conjuntos de painéis fotovoltaicos do IFCE Maracanaú.

Na Figura 1 é ilustrado o arranjo das placas no *campus*, dividido em dois conjuntos. O primeiro, no lado esquerdo da figura, possui 47 placas. O segundo conjunto, do lado direito, possui 80 placas. Cada um contém uma capacidade máxima de geração de até 18kW e 31kW, respectivamente, considerando condições ideais de operação.

O sistema de geração fotovoltaica conta ainda com dois inversores de frequência, mostrados na Figura 2, para condicionar a energia gerada à rede de distribuição e ao campus. O primeiro inversor (à esquerda, na Figura 2) é do modelo GW15KN-DT e possui a capacidade de gerenciar 15kW. Já o segundo (à direita, na Figura 2), é do modelo GW30K-MT e tem 30kW de capacidade de gerenciamento. Dessa forma, o primeiro inversor é responsável por fazer a conversão de energia do sistema que tem o menor número de placas, enquanto o segundo gerencia o sistema com o maior número de placas.



Figura 2: Sistema de gerenciamento elétrico dos inversores responsáveis pela geração fotovoltaica no IFCE Maracanaú.

A parte de monitoramento das grandezas medidas e geração de histórico de dados é realizada a partir de uma plataforma disponibilizada pela empresa fornecedora dos inversores (*GoodWe Technologies*). Chamada de estação de potência, ilustrada na Figura 3, fornece acesso às medições de diversas variáveis através dos inversores, tais como correntes e tensões de fase, temperatura do inversor e potência ativa gerada, dentre outras.

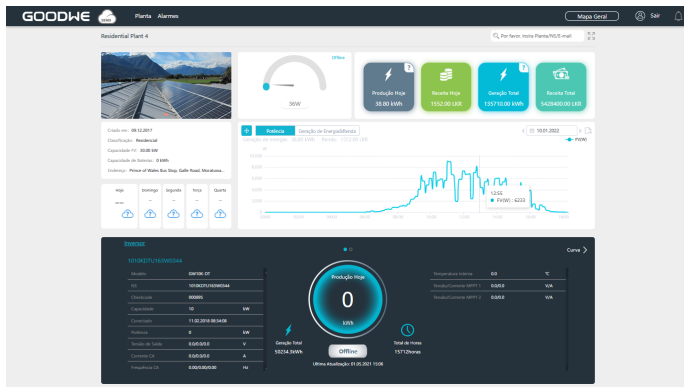


Figura 3: Tela do módulo de monitoramento e histórico das variáveis medidas pelos inversores.

Para os objetivos do presente trabalho, foram coletados durante cerca de quinze meses os dados da potência ativa processada pelos dois inversores. Além disso, a amostragem foi originalmente obtida na escala de minutos.

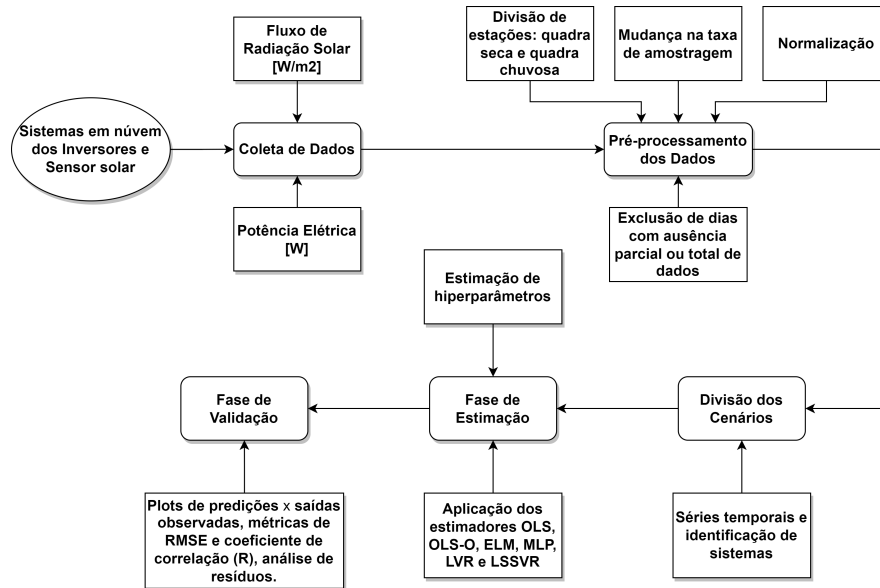


Figura 4: Metodologia da abordagem proposta.

### B. Sistema de Fluxo de Radiação Solar

De acordo com o estado da arte, existem outras grandezas que influenciam diretamente no desempenho da geração fotovoltaica como, por exemplo, velocidade do vento, pressão atmosférica, temperatura ambiente e radiação solar [13].

Com o intuito de aumentar a confiabilidade das predições de potência gerada, optou-se por utilizar adicionalmente um breve histórico de medições de radiação solar ( $W/m^2$ ), obtido a partir de um sensor piranômetro instalado nas dependências do IFCE Maracanaú.

Entretanto, por questões técnicas e por conta da pandemia de Covid-19, o sistema de aquisição dos dados de radiação solar ficou inoperante durante um longo período. Assim, para efeito de utilização nos experimentos do presente trabalho, foram consideradas as leituras de radiação no período de maio a agosto de 2022, totalizando cerca de 90 dias de coleta, concomitantes com o período de medições da potência gerada.

## III. METODOLOGIA PROPOSTA

Esta seção descreve a metodologia aplicada ao problema tratado, conforme ilustrada na Figura 4, desde a aquisição dos dados de potência elétrica e radiação solar, até a validação dos resultados obtidos a partir de simulações computacionais para a predição de potência ativa gerada, aplicando os paradigmas de previsão de séries temporais e identificação de sistemas.

### A. Coleta de Dados

A coleta de dados foi realizada em duas diferentes abordagens. Primeiramente, foram coletadas as medições de potência ativa a partir dos dois inversores instalados, no sistema em nuvem. Esses dados foram usados para as predições usando séries temporais. Já a segunda abordagem considerou adicionalmente a coleta dos dados de radiação solar, a partir do sensor piranômetro. Nesse caso, as medições de potência e

radiação foram usadas para obter predições via identificação de sistemas.

As medições da potência ativa gerada (em  $W$ ) foram inicialmente tratadas com uma taxa de amostragem de um minuto, no período de quinze meses, de 15 de maio de 2021 a 15 de agosto de 2022. Na Figura 5 são mostrados os dados coletados nesse período nos inversores 1 (Figura 5a) e 2 (Figura 5b), totalizando 283.953 e 281.383 amostras, respectivamente.

As medições realizadas com o fluxo de radiação solar (em  $W/m^2$ ) também foram obtidas com uma taxa de amostragem de um minuto em um período de três meses, de 18 de maio a 15 de agosto de 2022. Na Figura 5c são mostrados os dados de radiação coletados a partir do sensor piranômetro, totalizando 89.000 amostras.

### B. Pré-processamento dos dados

Uma vez coletadas as medições de potência ativa e radiação solar, foi necessário realizar um pré-processamento desses dados no intuito de torná-los mais representativos ao problema tratado.

1) *Mudança na taxa de amostragem*: Analisando os dados de potência ativa, percebeu-se que não possuíam uma taxa de amostragem fixa. Portanto, foi necessário realizar um pré-processamento para modificar a sua taxa de amostragem e mantê-la contínua. Já o fluxo de radiação possuía uma taxa de amostragem fixa em minutos, no entanto, observou-se ausências recorrentes nas amostras (valores nulos), aparentemente sem justificativa. Assim, para contornar tais problemas, ambas as base de dados foram modificadas para uma medição por hora, calculando-se a média dos dados coletados em cada intervalo.

Para critério de ilustração, na Figura 6a é apresentada uma visão dos primeiros dias de dados da potência ativa antes do pré-processamento, em que é possível observar um

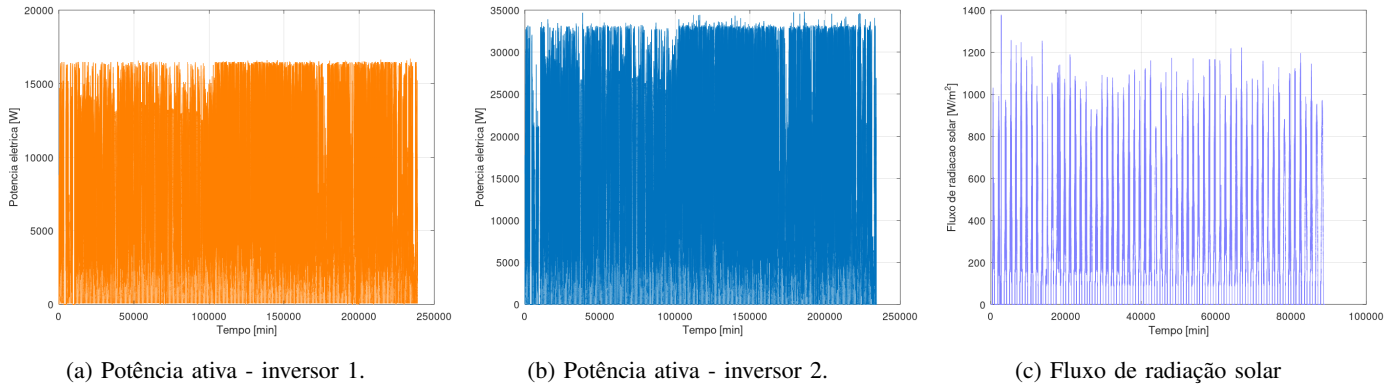


Figura 5: Amostras originais coletadas dos inversores e do sensor piranômetro.

pequeno deslocamento para a direita da potência gerada a partir do inversor 1 (curva vermelha). Esse comportamento é consequência da taxa de amostragem variável, o que acaba inviabilizando somar as duas potências para extrair a potência total do sistema.

2) *Exclusão de dias com ausência de dados*: Na Figura 6a também são apresentadas regiões em que as potências geradas nos dois inversores são nulas. Prováveis explicações para isso são: ocorrência de problemas no envio dos dados para a plataforma, manutenção nos inversores e interrupção de fornecimento elétrica no *campus*. Além disso, foi considerado que em um dia completo (24 horas) tem-se aproximadamente 15 horas de geração solar, visto que durante a noite não há geração. Portanto, todas essas regiões atípicas foram retiradas das amostras para não atrapalhar o procedimento de estimação. O resultado final do pré-processamento é mostrado na Figura 6b.

### C. Métricas de Validação

Para avaliação dos resultados obtidos, foram utilizadas como figuras de mérito os valores de RMSE (*Root Mean Squared Error*) e do coeficiente de correlação ( $R$ ) entre a saída real e a saída predita, usando as amostras do conjunto de validação. Matematicamente, os valores de RMSE e  $R$  podem ser calculados como:

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

$$R = \frac{1}{N-1} \sum_{i=1}^N \left( \frac{y_i - \mu_y}{\sigma_y} \right) \left( \frac{\hat{y}_i - \mu_{\hat{y}}}{\sigma_{\hat{y}}} \right), \quad (2)$$

em que  $N$  é o número de amostras da fase de validação,  $y_i$  é a  $i$ -ésima saída real,  $\hat{y}_i$  é a  $i$ -ésima saída estimada,  $\mu$  é a média amostral e  $\sigma$  é o desvio padrão.

De maneira a assegurar ainda mais a validade dos resultados obtidos, foi aplicado um teste estatístico para os resíduos ( $\xi = y_{\text{real}} - y_{\text{pred}}$ ) produzidos pelos modelos avaliados. O objetivo desse teste é verificar se os resíduos obtidos com os dados de estimação em predição de um passo à frente são

descorrelacionados, ou seja, se o modelo realmente conseguiu explicar tudo que for explicável nos dados [14]. Para um problema de séries temporais, como o modelo não apresenta entrada, é utilizada apenas a função de autocorrelação do vetor de resíduos ( $r_{\xi\xi}$ ). Para o caso de identificação de sistemas, além da autocorrelação dos resíduos, também são usadas correlações lineares ( $r_{u\xi}$ ) e não lineares ( $r_{u2'\xi}$  e  $r_{u2'\xi^2}$ ) entre os resíduos e a sequência de entrada do sistema [14]. Nesse contexto, um intervalo de confiança de 95% é usado para decidir se o teste é satisfeito e se os modelos são validados.

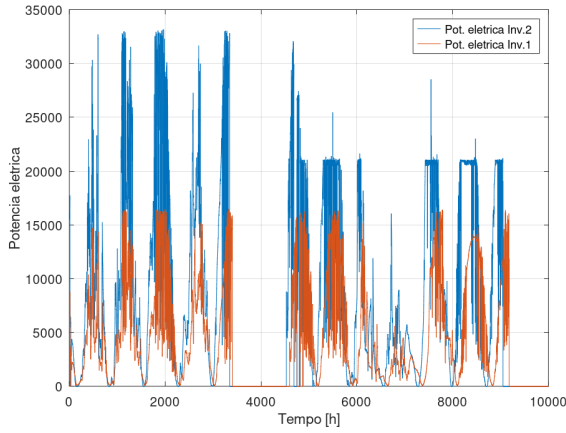
### D. Estimadores

Para os experimentos computacionais, foi aplicado o estimador linear *Ordinary Least Squares* (OLS) [14]. Já em termos de estimadores não lineares, foram usados dois baseados em métodos de *kernel* (com função de ativação gaussiana), *Support Vector Regression* (SVR) [15] e *Least Squares Support Vector Regression* (LSSVR) [16], e duas das mais usuais arquiteturas de redes neurais artificiais, *Extreme Learning Machine* (ELM) [17] e *MultiLayer Perceptron* (MLP) [18].

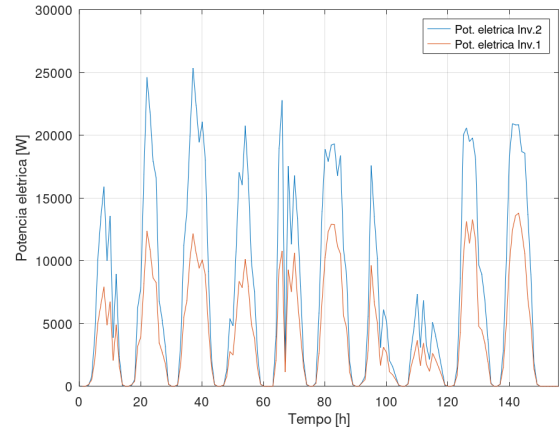
A seleção dos modelos de Aprendizado de Máquina supracitados foi fundamentada em sua ampla aplicabilidade em contextos relacionados à previsão de séries temporais e à identificação de sistemas. Especificamente na área de produção de energia fotovoltaica, diversos trabalhos recentes tem empregado esses estimadores, a exemplo de: ELM [19]–[21], MLP [22]–[24], SVR [25]–[27], LSSVR [10], [28], [29], para citar alguns.

## IV. RESULTADOS E DISCUSSÃO

Nesta seção, serão apresentados e discutidos os resultados obtidos a partir de simulações computacionais em diferentes tarefas de regressão dinâmica: predição de séries temporais com dados da quadra seca (potência ativa), predição de séries temporais com dados da quadra chuvosa (potência ativa) e identificação de sistemas (potência ativa e radiação solar). Para tanto, todos os algoritmos foram escritos em *GNU Octave 8.2.0*, e os experimentos rodaram em computadores *desktop IlhaWay*, com processador *Intel Core i5 3.20Ghz*, 4GB de memória RAM e com *Windows 10 Pro 64 bits*.



(a) Potências geradas a partir dos inversores 1 e 2 ANTES do pré-processamento.



(b) Potências geradas a partir dos inversores 1 e 2 DEPOIS do pré-processamento.

Figura 6: Exemplos de amostras das potências geradas a partir dos inversores 1 e 2 antes e depois da fase de pré-processamento.

De maneira a buscar os hiperparâmetros ótimos para cada estimador (OLS, ELM, MLP, SVR e LSSVR), utilizou-se a técnica *grid search* com *10-fold cross validation* (*10-fold CV*) para as tarefas de predição de séries temporais. Já em identificação de sistemas, foi aplicada *grid search* com *5-fold cross validation* (*5-fold CV*) em virtude da baixa cardinalidade da respectiva base de dados. Adicionalmente, decidiu-se por otimizar a ordem dos regressores para o estimador OLS,  $\hat{n}_y$  para predição de séries temporais e  $\hat{n}_u, \hat{n}_y$  para identificação de sistemas, fixando-se esses valores para os demais estimadores. Por fim, para mitigar as disparidades de escala entre as amostras de dados, optou-se por normalizá-las entre 0 e 1.

Uma vez definidas as ordens dos regressores para os estimadores em predição de séries temporais ( $\hat{n}_y = 15$  e 14 para quadra seca e chuvosa, respectivamente) e identificação de sistemas ( $\hat{n}_y = 14$  e  $\hat{n}_u = 13$ ), foram aplicadas *10-fold CV* e *5-fold CV*, respectivamente, para otimizar o número de neurônios ocultos ( $q$ ) nas redes ELM e MLP, a largura de kernel gaussiano ( $\sigma$ ), o *trade off* entre a solução de norma mínima e os erros cometidos ( $C$ ) para SVR e LSSVR. Já o hiperparâmetro que controla a tolerância de erros ( $\epsilon$ ) do estimador SVR foi setado no valor 0,3 a partir de experimentos preliminares. Os valores setados dos hiperparâmetros para cada estimador são mostrados na Tabela I.

Nos experimentos computacionais, os *hold-outs* escolhidos para as fases de estimação e validação em cada cenário foram: 69% – 31% e 66.5% – 33.5% para predição de séries temporais quadra seca e quadra chuvosa, respectivamente, e 79% – 21% para identificação de sistemas. Para avaliar os resultados, foram realizadas 50 rodadas independentes de cada estimador para calcular valores médios (e dispersão) de RMSE e R com os dados de validação. Conforme mencionado anteriormente, um dia de medição tem cerca de 15 horas de geração fotovoltaica. Assim, os horizontes de predição analisados a seguir foram baseados em predições de 1 passo

 Tabela I: Valores dos hiperparâmetros setados para cada estimador via *grid search*, nos diferentes cenários de simulação.

Cenário	Estimadores			
	ELM	MLP	SVR	LSSVR
<b>Séries temporais</b> <b>Quadra Seca</b>	$q = 47$	$q = 23$	$C = 64$ $\sigma^{-2} = 3, 13$ $\epsilon = 0, 3$	$C = 64$ $\sigma^{-2} = 50, 00$ $\epsilon = 0, 3$
<b>Séries temporais</b> <b>Quadra Chuvosa</b>	$q = 41$	$q = 22$	$C = 16$ $\sigma^{-2} = 3, 13$ $\epsilon = 0, 3$	$C = 16$ $\sigma^{-2} = 50, 00$ $\epsilon = 0, 3$
<b>Identificação de Sistemas</b>	$q = 41$	$q = 28$	$C = 64$ $\sigma^{-2} = 0, 20$ $\epsilon = 0, 3$	$C = 64$ $\sigma^{-2} = 3, 13$ $\epsilon = 0, 3$

à frente (1 hora), 45 passos à frente (45 horas - correspondem a três dias de predição), 105 passos à frente (105 horas - correspondem a sete dias de predição) e simulação livre (infinitos passos à frente).

#### A. Predição de Séries Temporais

Os resultados obtidos a partir dos experimentos computacionais para o cenário de predição de séries temporais são mostrados nas Tabelas II e III para as quadras seca e chuvosa, respectivamente. Observando os resultados na Tabela II, é possível afirmar que a rede MLP obteve os menores valores médios de RMSE nos três horizontes de predição simulados, uma hora à frente, 45 horas à frente e 105 horas à frente. Da mesma forma, a rede MLP também alcançou os maiores valores do coeficiente de correlação (R), entre todos os estimadores analisados. Vale ressaltar que quanto maior o valor de R, maior a correlação entre a saída predita e a saída real do sistema e, portanto, melhor tende a ser a qualidade das predições.

Para os resultados na quadra chuvosa, conforme mostra a Tabela III, pode-se observar que novamente a rede MLP

Tabela II: Resultados obtidos de RMSE e R com amostras de validação, ao longo de 50 rodadas independentes de cada estimador, em previsão de séries temporais - quadra SECA.

Estimador	1 Hora		45 Horas		105 horas	
	RMSE	R	RMSE	R	RMSE	R
<i>OLS</i>	7, 87E-2 ±2, 24E-16	9, 68E-1 ±9, 32E-16	9, 13E-2 ±4, 02E-15	9, 57E-1 ±3, 92E-15	9, 37E-2 ±1, 93E-14	9, 58E-1 ±1, 26E-14
<i>ELM</i>	7, 56E-2 ±1, 15E-03	9, 71E-1 ±9, 14E-04	8, 36E-2 ±3, 58E-03	9, 65E-1 ±3, 05E-03	9, 34E-2 ±1, 68E-02	9, 58E-1 ±1, 72E-02
<i>MLP</i>	<b>7, 58E-2 ±6, 07E-04</b>	<b>9, 71E-1 ±4, 29E-04</b>	<b>8, 13E-2 ±2, 56E-03</b>	<b>9, 69E-1 ±1, 05E-03</b>	<b>8, 34E-2 ±4, 74E-03</b>	<b>9, 67E-1 ±3, 91E-03</b>
<i>SVR</i>	1, 47E-1 ±1, 01E-11	9, 42E-1 ±3, 46E-12	3, 05E-1 ±8, 30E-12	4, 70E-1 ±2, 38E-11	3, 12E-1 ±9, 03E-12	1, 87E-1 ±8, 41E-12
<i>LSSVR</i>	8, 14E-2 ±5, 98E-15	9, 66E-1 ±4, 89E-15	1, 05E-1 ±6, 02E-13	9, 61E-1 ±1, 68E-13	1, 25E-1 ±7, 37E-13	9, 62E-1 ±5, 38E-13

Tabela III: Resultados obtidos de RMSE e R com amostras de validação, ao longo de 50 rodadas independentes de cada estimador, em previsão de séries temporais - quadra CHUVOSA.

Estimador	1 Hora		45 Horas		105 horas	
	RMSE	R	RMSE	R	RMSE	R
<i>OLS</i>	9, 69E-2 ±9, 37E-17	9, 28E-1 ±7, 98E-16	1, 73E-1 ±1, 61E-15	7, 56E-1 ±1, 13E-14	2, 06E-1 ±7, 03E-15	6, 60E-1 ±3, 54E-14
<i>ELM</i>	9, 66E-2 ±1, 14E-03	9, 29E-1 ±1, 77E-03	1, 69E-1 ±1, 26E-02	7, 84E-1 ±3, 13E-02	2, 02E-1 ±4, 61E-02	6, 75E-1 ±1, 75E-01
<i>MLP</i>	<b>9, 59E-2 ±2, 56E-03</b>	<b>9, 32E-1 ±3, 45E-03</b>	<b>1, 63E-1 ±1, 06E-02</b>	<b>8, 16E-1 ±1, 17E-02</b>	<b>1, 75E-1 ±1, 42E-02</b>	<b>8, 15E-1 ±2, 60E-02</b>
<i>SVR</i>	1, 44E-1 ±5, 65E-11	9, 09E-1 ±5, 20E-13	2, 88E-1 ±1, 52E-10	2, 87E-1 ±8, 06E-11	2, 93E-1 ±1, 21E-10	1, 38E-1 ±3, 82E-11
<i>LSSVR</i>	1, 10E-1 ±3, 26E-15	9, 13E-1 ±4, 28E-15	2, 07E-1 ±8, 57E-14	7, 83E-1 ±1, 71E-13	2, 20E-1 ±1, 62E-12	7, 60E-1 ±1, 98E-12

Tabela IV: Resultados obtidos de RMSE e R com amostras de validação, ao longo de 50 rodadas independentes de cada estimador, em identificação de sistemas.

Estimador	1 Hora		105 Horas		Simulação livre	
	RMSE	R	RMSE	R	RMSE	R
<i>OLS</i>	8, 92E-2 ±1, 24E-15	9, 63E-1 ±9, 45E-16	<b>1, 46E-1 ±4, 93E-15</b>	<b>8, 99E-1 ±5, 88E-15</b>	<b>1, 46E-1 ±4, 95E-15</b>	<b>8, 99E-1 ±5, 85E-15</b>
<i>ELM</i>	8, 74E-2 ±2, 50E-03	9, 65E-1 ±1, 88E-03	1, 58E-1 ±2, 97E-02	8, 85E-1 ±3, 88E-02	1, 51E-1 ±1, 90E-02	8, 94E-1 ±2, 62E-02
<i>MLP</i>	9, 23E-2 ±7, 56E-04	9, 61E-1 ±1, 11E-03	1, 73E-1 ±2, 89E-02	8, 58E-1 ±4, 19E-02	1, 68E-1 ±2, 80E-02	8, 59E-1 ±4, 01E-02
<i>SVR</i>	1, 50E-1 ±2, 05E-13	9, 44E-1 ±7, 08E-14	1, 94E-1 ±2, 66E-13	9, 29E-1 ±1, 56E-13	1, 94E-1 ±2, 33E-13	9, 29E-1 ±1, 34E-13
<i>LSSVR</i>	<b>8, 61E-2 ±1, 75E-14</b>	<b>9, 67E-1 ±5, 86E-15</b>	1, 67E-1 ±3, 52E-14	8, 68E-1 ±4, 55E-14	1, 69E-1 ±4, 23E-14	8, 64E-1 ±5, 65E-14

alcançou os melhores valores de RMSE e R dentre todos os estimadores testados, nos três horizontes de previsão analisados. Entretanto, em geral, os desempenhos dos estimadores, inclusive da rede MLP, apresentaram redução em suas performances comparativamente às obtidas para a quadra seca, na Tabela II. Uma possível explicação para isso seja o fato de que durante a quadra seca nos estados da região do Nordeste do Brasil, a maior parte desse período costuma ser de dias realmente ensolarados, com pouquíssimas precipitações chuvosas. Por outro lado, na quadra chuvosa, mesmo apresentando dias nublados e/ou bastante chuvosos, costuma-se ter períodos ou dias ensolarados, o que pode causar uma variação significativa nos níveis de potência gerada e, consequentemente, reduzir a qualidade das previsões realizadas.

As previsões para o horizonte de 105 horas à frente (uma semana) no pior caso, ou seja as previsões que obtiveram os maiores valores de RMSE ao longo das 50 rodadas, para o cenário de séries temporais com a rede MLP nas quadras seca e chuvosa são ilustradas nas Figuras 7a e 7b, respectivamente. Em ambos os casos, mesmo se tratando da análise no pior caso, é possível verificar que as saídas da rede MLP conseguiram, em geral, acompanhar o comportamento dinâmico do histórico de potência com desempenho razoável. Isso é ainda mais evidente no caso das previsões da quadra seca (Figura 7a) uma vez que o comportamento da potência gerada é um tanto mais uniforme, quando comparada à quadra chuvosa, na Figura 7b.

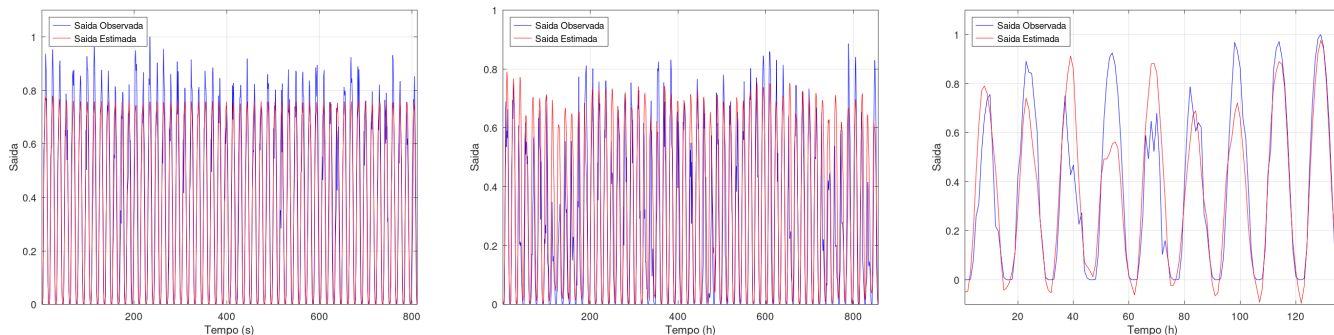
### B. Identificação de Sistemas

A segunda parte dos experimentos considerou também as medições de radiação solar para compor o cenário de identificação de sistemas. Os resultados foram obtidos com os estimadores nos horizontes de previsão de um passo à frente (uma hora), 105 passos à frente (uma semana) e simulação livre, os quais são mostrados na Tabela IV. Observando os valores médios e dispersões de RMSE e R, percebe-se que para o cenário de um passo à frente o LSSVR obteve o melhor desempenho dentre os modelos testados. Entretanto, para os demais horizontes de previsão, 105 passos à frente e simulação livre, o estimador linear dos mínimos quadrados alcançou os menores valores de RMSE, bem como os maiores valores de R, ou seja, os mais próximos da unidade.

Há algumas pontuações que merecem destaque nos valores mostrados na Tabela IV. Os resultados obtidos com o OLS em identificação de sistemas, mesmo no cenário de simulação livre e usando uma quantidade significativamente menor de dados, foram superiores aos obtidos em todos os cenários de previsão de séries temporais. Tal fato é um indicativo de que a inclusão das medições de radiação solar contribuiu significativamente para a melhoria na qualidade das previsões da potência elétrica gerada.

Percebe-se ainda que a utilização dos sinais de radiação solar (entrada) e potência ativa (saída) parece ter transformado a tarefa de identificação de sistemas para predominantemente





(a) Predição de uma semana a frente em séries temporais - quadra seca do estimador *MLP* no cenário de maior *RMSE* dentro das 50 rodadas independentes.

(b) Predição de uma semana à frente no cenário de séries temporais - quadra chuvosa do estimador *MLP* no cenário de maior *RMSE* dentro das 50 rodadas independentes.

(c) Predição infinitos passos a frente em identificação de sistemas do *OLS* no cenário de maior *RMSE* dentro das 50 rodadas independentes.

Figura 7: Resultados de simulação de máxima predição dos melhores estimadores em cada cenário abordado, no pior caso dentro das 50 rodadas independentes.

linear. O desempenho da predição em simulação livre do estimador *OLS* no pior caso, aquele que obteve o maior valor de *RMSE* nas 50 rodadas, é ilustrado na Figura 7c.

### C. Análise de Resíduos

Por fim, nesta seção discute-se os resultados obtidos com a aplicação de alguns testes de análise de resíduos aos modelos que alcançaram os melhores desempenhos nas Tabelas II, III e IV, correspondendo à rede *MLP* para séries temporais nas quadras seca e chuvosa e *OLS* para identificação de sistemas.

Para o cenário de predição de séries temporais, em que não há sinal de entrada, é realizado apenas o teste com a função de autocorrelação dos resíduos ( $r_{\xi\xi}$ ), que podem ser vistos na Figura 8a para a quadra seca e na Figura 8b para a quadra chuvosa. Pode-se dizer que os resíduos obtidos com a rede *MLP* em ambos os casos passaram no teste, dentro do intervalo de confiança de 95% (entre as linha azuis), o que é um indicativo de que os resíduos obtidos são ruído branco.

Já para o cenário de identificação de sistemas, uma vez que o estimador *OLS* é linear, optou-se em aplicar os testes com a função de autocorrelação do resíduo ( $r_{\xi\xi}$ ), mostrado na Figura 8c, e a função de correlação cruzada entre o resíduo e a entrada ( $r_{u\xi}$ ), conforme ilustrado na Figura 8d. No primeiro caso, com exceção de apenas uma amostra deslocada no tempo, o teste de autocorrelação do resíduo permaneceu entre os limites desejados. Entretanto, para o caso da correlação cruzada entre resíduo e entrada, percebe-se que para algumas amostras o teste extrapolou o intervalo de confiança. Talvez isso seja um indicativo da existência de algum grau de não linearidade no problema tratado, o que não impediu que o estimador *OLS* alcançasse desempenho satisfatório, mesmo no cenário de simulação livre.

## V. CONCLUSÕES

Neste artigo, foi realizado um estudo de caso sobre a predição de geração fotovoltaica na região metropolitana de

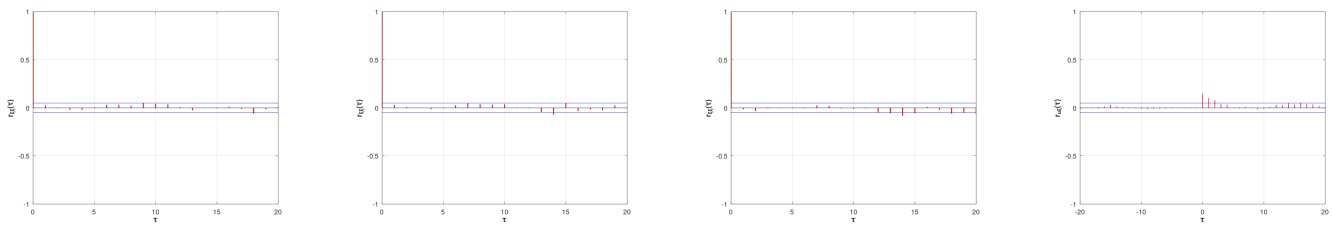
Fortaleza, mais especificamente no sistema instalado nas dependências do IFCE - *Campus* Maracanaú. Para tanto, a potência elétrica gerada foi analisada a partir da aplicação de estimadores lineares (*OLS*) e não lineares (*MLP*, *ELM*, *SVR* e *LSSVR*) em cenários de predição de séries temporais e de identificação de sistemas, usando radiação solar como entrada.

De acordo com os resultados obtidos a partir de simulações computacionais, percebeu-se que no cenário de predição de séries temporais, tanto para a quadra seca quanto para a quadra chuvosa, a rede neural *MLP* obteve o melhor desempenho em termos de *RMSE* e *R*, dentre todos os estimadores utilizados e em todos os horizontes de predição testados.

Já no cenário de identificação de sistemas, o estimador linear de mínimos quadrados (*OLS*) alcançou um desempenho ainda melhor, comparativamente aos obtidos pelos modelos lineares e não lineares em predição de séries temporais. A adição de medições da radiação solar como sinal de entrada acarretou uma melhoria considerável das predições do *OLS* para a potência elétrica gerada, mesmo no horizonte de predição de infinitos passos à frente (simulação livre).

Em virtude dos resultados satisfatórios obtidos com o estimador *OLS*, para trabalhos futuros pretende-se avaliar o desempenho de algumas de suas variantes, tais como mínimos quadrados com regularização, mínimos quadrados médios (*Least Mean Squares* - *LMS*) e mínimos quadrados recursivos (*Recursive Least Squares* - *RLS*).

Ainda como possíveis desdobramentos deste trabalho, pretende-se usar os modelos e predições realizadas para implementar sugestões de otimização da geração fotovoltaica e, a partir delas, reorganizar o calendário de manutenção preventiva. Adicionalmente, fazer um estudo análogo para predições do consumo de energia no campus para, juntamente com os resultados da análise de geração, produzir um diagnóstico técnico de eficiência energética para o IFCE Maracanaú.



(a)  $r_{\xi\xi}$  para a rede MLP - quadra seca. (b)  $r_{\xi\xi}$  para a rede MLP - quadra chuvosa. (c)  $r_{\xi\xi}$  para o OLS - identificação de sistemas. (d)  $r_{u\xi}$  para o OLS - identificação de sistemas.

Figura 8: Resultados da análise de resíduos para os estimadores que alcançaram os melhores desempenhos em previsão de séries temporais e identificação de sistemas.

#### AGRADECIMENTOS

Ao LAMSC (Laboratório de Aprendizado de Máquina e Simulações Computacionais), ao LPC (Laboratório de Potência e Controle) e ao LAESE (Laboratório de Eletrônica e Sistemas Embarcados) do IFCE - *Campus* Maracanaú, pelo apoio a esta pesquisa.

#### REFERÊNCIAS

- [1] B. Patnaik, M. Mishra, R. C. Bansal, and R. K. Jena, "Ac microgrid protection—a review: Current and future prospective," *Applied Energy*, vol. 271, p. 115210, 2020.
- [2] I. E. Agency, "Iea – international energy agency." 2023. [Online]. Available: <https://www.iea.org/>
- [3] A. K. Raturi, "Renewable 2018 global status," 2018.
- [4] A. A. B. de Energia Solar Fotovoltaica, "Energia solar fotovoltaica: Brasil é o 4º país que mais cresceu em 2021," 2023. [Online]. Available: <https://www.absolar.org.br/noticia/energia-solar-fotovoltaica-brasil-e-o-4o-pais-que-mais-cresceu-em-2021/>
- [5] O. O. N. do Sistema Elétrico, "Resultados da operação - histórico da geração eólica e fotovoltaica," 2023. [Online]. Available: <https://www.ons.org.br>
- [6] P. Asef, R. Bargallo, A. H. Karci, P. Niknejad, M. Barzegaran, and A. C. Laphorn, "Correlation of solar power prediction considering the nominal operating cell temperature under partial shading effect," *Measurement*, vol. 147, p. 106878, 2019.
- [7] K. Wang, X. Qi, and H. Liu, "A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network," *Applied Energy*, vol. 251, p. 113315, 2019.
- [8] H. Sharadga, S. Hajimirza, and R. S. Balog, "Time series forecasting of solar power generation for large-scale photovoltaic plants," *Renewable Energy*, vol. 150, pp. 797–807, 2020.
- [9] A. P. Gonzalo, A. P. Marugán, and F. P. G. Márquez, "Survey of maintenance management for photovoltaic power systems," *Renewable and Sustainable Energy Reviews*, vol. 134, p. 110347, 2020.
- [10] A. Fentis, M. Rafik, L. Bahatti, O. Bouattane, and M. Mestari, "Data driven approach to forecast the next day aggregate production of scattered small rooftop solar photovoltaic systems without meteorological parameters," *Energy Reports*, vol. 8, pp. 3221–3233, 2022.
- [11] P. M. Kumar, R. Saravanakumar, A. Karthick, and V. Mohanavel, "Artificial neural network-based output power prediction of grid-connected semitransparent photovoltaic system," *Environmental Science and Pollution Research*, pp. 1–10, 2022.
- [12] S. Al-Dahidi, O. Ayadi, M. Alrbai, and J. Adeeb, "Ensemble approach of optimized artificial neural networks for solar photovoltaic power prediction," *IEEE Access*, vol. 7, pp. 81 741–81 758, 2019.
- [13] R. Ahmed, V. Sreeram, Y. Mishra, and M. Arif, "A review and evaluation of the state-of-the-art in pv solar power forecasting: Techniques and optimization," *Renewable and Sustainable Energy Reviews*, vol. 124, p. 109792, 2020.
- [14] L. A. Aguirre, "Introdução à identificação de sistemas," *Editora UFMG, Belo Horizonte*, vol. 3, 2007.
- [15] A. Smola and V. Vapnik, "Support vector regression machines," *Advances in neural information processing systems*, vol. 9, pp. 155–161, 1997.
- [16] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.
- [17] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *2004 IEEE international joint conference on neural networks (IEEE Cat. No. 04CH37541)*, vol. 2. Ieee, 2004, pp. 985–990.
- [18] S. Haykin, *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1998.
- [19] H. Long, C. Zhang, R. Geng, Z. Wu, and W. Gu, "A combination interval prediction model based on biased convex cost function and auto-encoder in solar power prediction," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 3, pp. 1561–1570, 2021.
- [20] Y. Zhou, N. Zhou, L. Gong, and M. Jiang, "Prediction of photovoltaic power output based on similar day analysis, genetic algorithm and extreme learning machine," *Energy*, vol. 204, p. 117894, 2020.
- [21] R. K. Sahu, B. Shaw, J. R. Nayak *et al.*, "Short/medium term solar power forecasting of chhattisgarh state of india using modified tlbo optimized elm," *Engineering Science and Technology, an International Journal*, vol. 24, no. 5, pp. 1180–1200, 2021.
- [22] D. Markovics and M. J. Mayer, "Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction," *Renewable and Sustainable Energy Reviews*, vol. 161, p. 112364, 2022.
- [23] C. Keerthisinghe, E. Mickelson, D. S. Kirschen, N. Shih, and S. Gibson, "Improved pv forecasts for capacity firming," *IEEE Access*, vol. 8, pp. 152 173–152 182, 2020.
- [24] H. Zhou, Y. Zhang, L. Yang, Q. Liu, K. Yan, and Y. Du, "Short-term photovoltaic power forecasting based on long short term memory neural network and attention mechanism," *IEEE Access*, vol. 7, pp. 78 063–78 074, 2019.
- [25] P. Aji, K. Wakamori, and H. Mineno, "Short-term solar power forecasting using svr on hybrid pv power plant in indonesia," in *Advances in Intelligent Networking and Collaborative Systems: The 11th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2019)*. Springer, 2020, pp. 235–246.
- [26] P. Pawar, N. Mithulananthan, and M. Q. Raza, "Solar pv power forecasting using modified svr with gauss-newton method," in *2020 2nd Global Power, Energy and Communication Conference (GPECOM)*, 2020, pp. 226–231.
- [27] I. Jebli, F.-Z. Belouadha, M. I. Kabbaj, and A. Tilioua, "Prediction of solar energy guided by pearson correlation using machine learning," *Energy*, vol. 224, p. 120109, 2021.
- [28] A. Fentis, C. Lytridis, V. G. Kaburlasos, E. Vrochidou, T. Pachidis, E. Bahatti, and M. Mestari, "A machine learning based approach for next-day photovoltaic power forecasting," in *2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS)*. IEEE, 2020, pp. 1–8.
- [29] A. Fentis, L. Bahatti, M. Tabaa, and M. Mestari, "Short-term nonlinear autoregressive photovoltaic power forecasting using statistical learning approaches and in-situ observations," *International Journal of Energy and Environmental Engineering*, vol. 10, no. 2, pp. 189–206, 2019.