



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CENTRO DE TECNOLOGIA**  
**DEPARTAMENTO DE ENGENHARIA MECÂNICA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA**  
**MESTRADO ACADÊMICO EM ENGENHARIA MECÂNICA**

**ANTONIO WELLINGTON DANTAS DA COSTA**

**MODELOS DE APRENDIZADO DE MÁQUINA PARA ESTIMATIVA DE  
TEMPERATURA DE OPERAÇÃO DE MÓDULOS FOTOVOLTAICOS EM CLIMA  
SEMIÁRIDO**

**FORTALEZA**

**2026**

ANTONIO WELLINGTON DANTAS DA COSTA

MODELOS DE APRENDIZADO DE MÁQUINA PARA ESTIMATIVA DE TEMPERATURA  
DE OPERAÇÃO DE MÓDULOS FOTOVOLTAICOS EM CLIMA SEMIÁRIDO

Dissertação apresentada ao Curso de Mestrado Acadêmico em Engenharia Mecânica do Programa de Pós-Graduação em Engenharia Mecânica do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia Mecânica. Área de Concentração: Processos, equipamentos e sistemas para energias renováveis.

Orientador: Prof. Dr. Paulo Cesar Marques de Carvalho.

Coorientadora: Dra. Leticia Oliveira Santos.

FORTALEZA

2026

ANTONIO WELLINGTON DANTAS DA COSTA

MODELOS DE APRENDIZADO DE MÁQUINA PARA ESTIMATIVA DE TEMPERATURA  
DE OPERAÇÃO DE MÓDULOS FOTOVOLTAICOS EM CLIMA SEMIÁRIDO

Dissertação apresentada ao Curso de Mestrado Acadêmico em Engenharia Mecânica do Programa de Pós-Graduação em Engenharia Mecânica do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia Mecânica. Área de Concentração: Processos, equipamentos e sistemas para energias renováveis.

Aprovada em: 10/03/2026.

BANCA EXAMINADORA

---

Prof. Dr. Paulo Cesar Marques de  
Carvalho (Orientador)  
Universidade Federal do Ceará (UFC)

---

Dra. Leticia Oliveira Santos (Coorientadora)  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Paulo Alexandre Costa Rocha  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Francisco Nilvaldo Aguiar Freire  
Universidade Federal do Ceará (UFC)

À minha esposa, companheira de todas as horas,  
que com seu amor incondicional e apoio cons-  
tante tornou possível cada passo desta jornada.  
Ao meu filho, grande amor, que trouxe luz e  
alegria aos momentos mais desafiadores.  
Aos meus pais, por me proporcionarem os ali-  
cerces para chegar até aqui.



## **AGRADECIMENTOS**

Ao Prof. Dr. Paulo Cesar Marques de Carvalho, pela excelente orientação e pelo apoio a mim fornecido.

À Dra. Leticia de Oliveira Santos, pela coorientação durante todo o processo.

Aos membros da banca examinadora, Prof. Dr. Paulo Alexandre Costa Rocha e Prof. Dr. Francisco Nivaldo Aguiar Freire, pela disponibilidade e pelas valiosas colaborações e sugestões.

À minha esposa, por me apoiar durante todo o período do mestrado, e ao meu filho, por me motivar a sempre lutar pelos meus sonhos.

Aos colegas do Laboratório de Energias Alternativas, pelas contribuições e pelo suporte concedidos durante a minha pesquisa.

Aos colegas e docentes do curso de Mestrado em Engenharia Mecânica, pelas reflexões, críticas e sugestões recebidas.

À FUNCAP, vinculada à SECITECE, pelo apoio financeiro com a concessão da bolsa de mestrado.

Porque o Senhor dá a sabedoria, e da sua boca  
vem o conhecimento e o entendimento.

(Provérbios 2:6)

## RESUMO

A crescente demanda por fontes de energia limpa e renovável impulsiona a expansão de sistemas fotovoltaicos (FV) em regiões de clima semiárido, onde a alta irradiância solar se combina com temperaturas elevadas que reduzem a eficiência de conversão energética. A temperatura de operação dos módulos FV ( $T_{mod}$ ) é variável crítica para dimensionamento, monitoramento e otimização de desempenho. Este trabalho implementa e avalia modelos de *Machine Learning* (Aprendizado de Máquina – ML) para estimativa de  $T_{mod}$  em condições de clima semiárido tropical. Utilizando dados experimentais coletados no Laboratório de Energias Alternativas da Universidade Federal do Ceará (LEA-UFC), em Fortaleza, durante dois anos de monitoramento (maio de 2023 a maio de 2025), foram implementados e comparados modelos organizados em três categorias: algoritmos de ML (Regressão Linear - RL, *K-Nearest Neighbors* - KNN, *Random Forest* - RF e *eXtreme Gradient Boosting* - XGBoost), correlações semiempíricas híbridas (linear, não-linear e racional) e modelos físico-empíricos consolidados (*Ross*, *Faiman* e *PVsys*). O conjunto de dados final compreendeu 273.310 observações ao longo de dois anos, utilizando cinco variáveis preditoras meteorológicas: irradiância solar ( $G$ ), temperatura ambiente ( $T_a$ ), velocidade do vento ( $V_w$ ) e ângulos solares. Os resultados demonstraram superioridade dos modelos de ML, com destaque para o KNN, que alcançou *Mean Absolute Error* (MAE) de  $0,83^\circ\text{C}$ , *Root Mean Square Error* (RMSE) de  $1,44^\circ\text{C}$  e coeficiente de determinação  $R^2 = 0,97$ , representando redução de aproximadamente 77% no MAE comparado aos modelos físicos. As correlações semiempíricas híbridas apresentaram desempenho intermediário, com redução de 22% no RMSE e 68% no viés sistemático comparado aos modelos físicos convencionais. A análise de importância de variáveis revelou que  $G$  e  $T_a$  respondem por 70–80% da capacidade preditiva, enquanto  $V_w$  apresentou contribuição inferior a 2%, consequência do microclima urbano com rugosidade elevada e baixas velocidades médias. O desempenho superior do ML explica-se pela estabilidade do clima semiárido local, que facilitou o aprendizado dos padrões, e pela capacidade dos algoritmos de se ajustarem automaticamente aos dados locais, capturando variações térmicas complexas sob elevada irradiação que modelos físicos genéricos não preveem. O trabalho aprofunda a compreensão sobre o aquecimento dos módulos Fv no semiárido brasileiro, fundamentando estratégias de otimização, resfriamento e controle inteligente.

**Palavras-chave:** aprendizado de máquina; geração fotovoltaica; temperatura de módulos; clima semiárido.

## ABSTRACT

The growing demand for clean and renewable energy sources drives the expansion of photovoltaic (PV) systems in semi-arid regions, where high solar irradiance is coupled with high temperatures that reduce energy conversion efficiency. The operating temperature of PV modules ( $T_{mod}$ ) is a critical variable for sizing, monitoring, and performance optimization. This work implements and evaluates *Machine Learning* (ML) models for estimating  $T_{mod}$  under tropical semi-arid climate conditions. Using experimental data collected at the Alternative Energy Laboratory of the Federal University of Ceará (LEA-UFC), in Fortaleza, during two years of monitoring (May 2023 to May 2025), models organized into three categories were implemented and compared: ML algorithms (Linear Regression - LR, *K-Nearest Neighbors* - KNN, *Random Forest* - RF, and *eXtreme Gradient Boosting* - XGBoost), hybrid semi-empirical correlations (linear, non-linear, and rational), and consolidated physical-empirical models (*Ross*, *Faiman*, and *PVsyst*). The final dataset comprised 273,310 observations over two years, using five meteorological predictor variables: solar irradiance ( $G$ ), ambient temperature ( $T_a$ ), wind speed ( $V_w$ ), and solar angles. The results demonstrated the superiority of ML models, particularly KNN, which achieved a *Mean Absolute Error* (MAE) of  $0.83^\circ\text{C}$ , *Root Mean Square Error* (RMSE) of  $1.44^\circ\text{C}$ , and a coefficient of determination  $R^2 = 0.97$ , representing a reduction of approximately 77% in MAE compared to physical models. The hybrid semi-empirical correlations showed intermediate performance, with a 22% reduction in RMSE and a 68% reduction in systematic bias compared to conventional physical models. Feature importance analysis revealed that  $G$  and  $T_a$  account for 70–80% of the predictive capacity, while  $V_w$  contributed less than 2%, a consequence of the urban microclimate with high roughness and low average speeds. The superior performance of ML is explained by the stability of the local semi-arid climate, which facilitated pattern learning, and by the algorithms' ability to automatically adjust to local data, capturing complex thermal variations under high irradiation that generic physical models fail to predict. This work deepens the understanding of PV module heating in the Brazilian semi-arid region, providing a foundation for optimization, cooling, and intelligent control strategies.

**Keywords:** machine learning; photovoltaic generation; module temperature; semi-arid climate.

## LISTA DE FIGURAS

Figura 1 – Sazonalidade dos recursos climáticos no Ceará . . . . .	36
Figura 2 – Instalação experimental no LEA-UFC . . . . .	41
Figura 3 – Variação sazonal das variáveis meteorológicas e $T_{mod}$ ao longo do período de medição (22/mai/2023–30/mai/2025) no clima semiárido de Fortaleza/CE. . . . .	52
Figura 4 – Histogramas de frequência e correlações entre os pares de variáveis do banco de dados do LEA-UFC (22 de maio de 2023 a 30 de maio de 2025). . . . .	54
Figura 5 – Gráficos de caracterização de $V_w$ , $T_a$ , $T_{mod}$ e $G$ : (a) ao longo do dia (b) ao longo do ano (c) estatísticas descritivas. . . . .	55
Figura 6 – Boxplot de distribuição das variáveis por hora do dia. . . . .	56
Figura 7 – Comparação entre valores preditos ( $\hat{T}_{mod}$ ) e reais ( $T_{mod}$ ) para os modelos ML . . . . .	61
Figura 8 – Comparação entre $T_{mod}$ média medida e $\hat{T}_{mod}$ estimada para os modelos ao longo dos primeiros 5 dias do conjunto de teste . . . . .	62
Figura 9 – Estimativa de $\hat{T}_{mod}$ (amarelo) sobreposta aos dados reais (azul) vs. dia do ano para os modelos: (a) RF, (b) RL, (c) KNN, (d) <i>XGBoost</i> . . . . .	64
Figura 10 – Comparação entre $T_{mod}$ e $\hat{T}_{mod}$ nos modelos de referência. . . . .	67
Figura 11 – Comparação do desempenho de todos os modelos avaliados considerando MAE e RMSE. . . . .	72

## LISTA DE TABELAS

Tabela 1 – Equações para $T_{mod}$ a partir do ano 2000 . . . . .	28
Tabela 2 – Características gerais dos modelos aplicados . . . . .	49
Tabela 3 – Configurações e hiperparâmetros explorados . . . . .	49
Tabela 4 – Principais resultados dos modelos de ML . . . . .	62
Tabela 5 – Desempenho dos modelos físico-semiempíricos de referência no conjunto de teste . . . . .	65
Tabela 6 – Desempenho dos modelos semiempíricos com correlações híbridas no conjunto de teste . . . . .	68
Tabela 7 – Importância relativa das variáveis meteorológicas nos modelos de ML. . . . .	71
Tabela 8 – Comparação qualitativa entre os modelos de ML avaliados. . . . .	73

## LISTA DE ABREVIATURAS E SIGLAS

BIPV	<i>Building-Integrated Photovoltaics</i>
FF	Fator de Forma
FV	Fotovoltaica
KNN	<i>K-Nearest Neighbors</i>
LEA-UFC	Laboratório de Energias Alternativas da Universidade Federal do Ceará
LSTM	<i>Long Short-Term Memory</i>
MAE	<i>Mean Absolute Error</i>
MBE	<i>Mean Bias Error</i>
ML	<i>Machine Learning</i>
NMOT	<i>Nominal Module Operating Temperature</i>
RF	<i>Random Forest</i>
RL	Regressão Linear
RMSE	<i>Root Mean Square Error</i>
STC	<i>Standard Test Conditions</i>
XGBoost	<i>Extreme Gradient Boosting</i>

## LISTA DE SÍMBOLOS

$a_i, b_i, C_i, \dots$	Coefficientes empíricos das correlações semiempíricas híbridas
$C_{th}$	Capacitância térmica (J/K)
$G$	Irradiância solar global (W/m <sup>2</sup> )
$I_{SC}$	Corrente de curto-circuito (A)
$k$	Número de vizinhos mais próximos no algoritmo KNN
$n$	Número de amostras no conjunto de dados
$P_{95}$	Percentil 95 do erro absoluto (°C)
$P_{max}$	Potência máxima (W)
$R^2$	Coefficiente de determinação
$r$	Coefficiente de perda térmica do modelo de Ross (°C·m <sup>2</sup> /W)
$T_a$	Temperatura ambiente (°C)
$T_b$	Temperatura da face posterior (fundo) do módulo (°C)
$T_{mod}$	Temperatura de operação do módulo fotovoltaico (°C)
$\hat{T}_{mod}$	Temperatura estimada do módulo fotovoltaico (°C)
$\bar{T}_{mod}$	Temperatura média dos módulos fotovoltaicos (°C)
$T_{sky}$	Temperatura radiativa efetiva do céu (K ou °C)
$U_0$	Coefficiente de perda térmica de referência do modelo de Faiman (W/(m <sup>2</sup> ·K))
$U_1$	Coefficiente linear convectivo do modelo de Faiman (W·s/(m <sup>3</sup> ·K))
$U_L$	Coefficiente global de perdas sensíveis (W/(m <sup>2</sup> ·K))
$u_c$	Coefficiente de perda térmica constante do modelo PVsyst (W/(m <sup>2</sup> ·K))
$u_v$	Coefficiente de perda térmica convectiva do modelo PVsyst (W·s/(m <sup>3</sup> ·K))
$V_{OC}$	Tensão de circuito aberto (V)
$V_w$	Velocidade do vento (m/s)
$y_i$	Valor observado da i-ésima amostra
$\hat{y}_i$	Valor predito da i-ésima amostra
$\bar{y}$	Valor médio das observações

$\alpha$	Absortância solar do módulo (adimensional)
$\beta$	Ângulo de inclinação do painel (°)
$\gamma$	Azimute do painel (°)
$\Delta T$	Diferença característica de temperatura (°C)
$\varepsilon$	Emissividade no infravermelho (adimensional)
$\eta$	Eficiência de conversão energética (adimensional)
$\theta_z$	Ângulo zenital solar (°)
$\lambda$	Parâmetro de regularização (adimensional)
$\rho$	Refletância (adimensional)
$\sigma$	Constante de Stefan-Boltzmann ( $\text{W}/(\text{m}^2 \cdot \text{K}^4)$ )
$\tau$	Transmissividade do vidro (adimensional)
$\phi_s$	Azimute solar (°)

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>16</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>20</b>
<b>2.1</b>	<b>Balanco de energia e mecanismos de transferência de calor</b>	<b>21</b>
<b>2.2</b>	<b>Fatores determinantes da temperatura de operação</b>	<b>22</b>
<b>2.3</b>	<b>Modelos de estimativa de temperatura</b>	<b>23</b>
<b>2.3.1</b>	<i>Classificação geral</i>	<b>24</b>
<b>2.3.1.1</b>	<i>Modelos físicos estacionários</i>	<b>25</b>
<b>2.3.1.2</b>	<i>Modelos físicos dinâmicos</i>	<b>25</b>
<b>2.3.1.3</b>	<i>Modelos baseados em parâmetros elétricos</i>	<b>26</b>
<b>2.3.1.4</b>	<i>Modelos baseados em dados</i>	<b>26</b>
<b>2.3.2</b>	<i>Síntese das correlações da literatura</i>	<b>27</b>
<b>2.3.3</b>	<i>Modelos físicos de referência</i>	<b>29</b>
<b>2.3.3.1</b>	<i>Modelo de Ross</i>	<b>29</b>
<b>2.3.3.2</b>	<i>Modelo de Faiman</i>	<b>30</b>
<b>2.3.3.3</b>	<i>Modelo PVsyst</i>	<b>30</b>
<b>2.3.3.4</b>	<i>Limitações estruturais das correlações convencionais</i>	<b>31</b>
<b>2.4</b>	<b>Modelos baseados em ML</b>	<b>32</b>
<b>2.4.1</b>	<i>Algoritmos comumente utilizados</i>	<b>32</b>
<b>2.4.2</b>	<i>Vantagens e limitações dos modelos ML</i>	<b>34</b>
<b>2.4.3</b>	<i>Interpretabilidade e coerência físico-estatística</i>	<b>35</b>
<b>2.5</b>	<b>Medição de <math>T_{mod}</math>: técnicas e limitações</b>	<b>35</b>
<b>2.6</b>	<b>Considerações para clima semiárido</b>	<b>36</b>
<b>2.7</b>	<b>Lacunas na literatura e justificativa do estudo</b>	<b>37</b>
<b>3</b>	<b>METODOLOGIA</b>	<b>39</b>
<b>3.1</b>	<b>Caracterização do sistema experimental</b>	<b>40</b>
<b>3.2</b>	<b>Tratamento dos dados</b>	<b>40</b>
<b>3.3</b>	<b>Seleção das variáveis preditivas</b>	<b>42</b>
<b>3.4</b>	<b>Métricas de avaliação e interpretabilidade</b>	<b>43</b>
<b>3.5</b>	<b>Modelagem por ML e justificativas</b>	<b>44</b>
<b>3.6</b>	<b>Modelagem híbrida</b>	<b>45</b>

3.6.1	<i>Otimização de modelos físicos</i>	45
3.6.2	<i>Calibração de correlações semi-empíricas</i>	47
3.6.3	<i>Validação e comparação dos modelos</i>	48
3.7	<b>Estratégias de validação e otimização</b>	48
4	<b>RESULTADOS</b>	51
4.1	<b>Análise dos dados coletados</b>	51
4.1.1	<i>Irradiância solar</i>	53
4.1.2	<i>Temperatura ambiente</i>	56
4.1.3	<i>Temperatura do módulo</i>	57
4.1.4	<i>Velocidade do vento</i>	58
4.1.5	<i>Ângulos solares</i>	58
4.1.6	<i>Variabilidade diária e padrões sazonais</i>	59
4.2	<b>Análise de desempenho dos modelos de ML</b>	60
4.3	<b>Modelos físicos de referência</b>	64
4.4	<b>Modelagem híbrida</b>	66
4.4.1	<i>Desempenho dos modelos semiempíricos com correlações híbridas</i>	68
4.4.2	<i>Análise da importância das variáveis</i>	70
4.5	<b>Análise comparativa de desempenho</b>	71
4.5.1	<i>Fatores determinantes do desempenho relativo dos modelos</i>	75
4.5.1.1	<i>Previsibilidade climática e cobertura do espaço de estados</i>	75
4.5.1.2	<i>Auto-calibração matemática versus parametrização generalista</i>	76
4.5.1.3	<i>Captura de não-linearidades e interações locais</i>	77
4.5.1.4	<i>Modelos híbridos: equilíbrio entre interpretabilidade física e desempenho</i>	78
4.5.1.5	<i>Implicações para generalização e transferibilidade</i>	79
5	<b>CONCLUSÕES</b>	81
6	<b>PRODUÇÃO CIENTÍFICA DERIVADA DESTA DISSERTAÇÃO</b>	83
	<b>REFERÊNCIAS</b>	84

## 1 INTRODUÇÃO

A geração fotovoltaica (FV) tem apresentado crescimento expressivo no cenário energético global, destacando-se como uma alternativa sustentável e economicamente viável para a geração de eletricidade, impulsionada por reduções de custo, avanços tecnológicos e políticas de descarbonização (INTERNATIONAL ENERGY AGENCY, 2024; REN21, 2024). Apesar desses avanços, a fração da irradiância solar global ( $G$ ) efetivamente convertida em eletricidade ainda é limitada pelas características termodinâmicas das células, permanecendo a eficiência comercial média de módulos de silício cristalino em faixas próximas de 20%–23% (GREEN *et al.*, 2024). A parcela não convertida é predominantemente transformada em calor, elevando a temperatura de operação dos módulos fotovoltaicos ( $T_{mod}$ ) e impactando negativamente seu desempenho elétrico por mecanismos associados à redução do *bandgap* efetivo e ao aumento das correntes de recombinação (RAUSCHENBACH, 1980; SKOPLAKI; PALYVOS, 2009).

O aumento da temperatura das células ocasiona principalmente a redução da tensão de circuito aberto ( $V_{oc}$ ), fator de forma (FF), da potência máxima ( $P_{max}$ ) e da eficiência ( $\eta$ ), enquanto a corrente de curto-circuito ( $I_{sc}$ ) apresenta incremento modesto (KING *et al.*, 2004; SKOPLAKI; PALYVOS, 2009). Coeficientes típicos de temperatura para módulos de silício cristalino situam-se em torno de  $-0,4$  a  $-0,5\%/^{\circ}\text{C}$  na potência ou eficiência (FAIMAN, 2008; GREEN *et al.*, 2024). Em condições de campo, especialmente em regiões de elevada  $G$  e baixa convecção, a temperatura de operação pode exceder a ambiente em  $20^{\circ}\text{C}$  a  $40^{\circ}\text{C}$  (ROSS, 1976; MATTEI *et al.*, 2006), configurando desafio adicional em climas tropicais.

A modelagem da temperatura de módulos FV é fundamental para: (i) estimativa de geração energética; (ii) análise de confiabilidade e degradação; (iii) otimização termoelétrica de arranjos; e (iv) suporte a estratégias de controle e manutenção preditiva. Abordagens presentes na literatura podem ser agrupadas em três grandes classes: (a) modelos físicos (ou termoenergéticos); (b) modelos baseados em dados; e (c) modelos híbridos que combinam conhecimento físico com técnicas *data-driven* (baseadas em dados), abrangendo desde métodos de calibração e otimização até algoritmos de *Machine Learning* (Aprendizado de Máquina – ML) (DUFFIE; BECKMAN, 2013; VOYANT *et al.*, 2017).

Modelos físicos estacionários, frequentemente derivados de correlações empíricas associadas a parâmetros, como a *Nominal Module Operating Temperature* (Temperatura Nominal de Operação do Módulo – NMOT), apresentam baixo custo computacional e são amplamente utilizados em simulações de longo prazo (ROSS, 1976; SKOPLAKI; PALYVOS, 2009). Contudo,

sua simplificação pode introduzir desvios significativos sob condições transientes ou microclimas heterogêneos (BARYKINA; HAMMER, 2017b). Extensões semiempíricas, como o modelo de *Faiman*, incorporam efeitos de  $G$  e velocidade do vento ( $V_w$ , do inglês *wind velocity*) em forma mais explícita (FAIMAN, 2008), melhorando a aderência em cenários variados.

Modelos dinâmicos de balanço energético descrevem a evolução temporal da temperatura por meio de termos de radiação, convecção e capacidade térmica, integrando variáveis como temperatura ambiente ( $T_a$ ),  $G$ ,  $V_w$ , propriedades óticas (transmissividade  $\tau$ , absorptância  $\alpha$ ) e parâmetros construtivos, como a espessura das camadas, massa específica e calor específico dos materiais do módulo (MATTEI *et al.*, 2006; FUENTES, 1987). Esses modelos oferecem maior fidelidade, mas demandam parâmetros nem sempre disponíveis ou facilmente mensuráveis, exigindo etapas de calibração e caracterização local (KING *et al.*, 2004; DUFFIE; BECKMAN, 2013). Estudos comparativos em diferentes regimes climáticos demonstram que a acurácia de correlações generalistas, isto é, modelos empíricos cujos parâmetros foram originalmente calibrados para climas temperados ou condições de laboratório, pode degradar sensivelmente, reforçando a necessidade de metodologias adaptativas e integradas a dados recentes de operação em campo (BARYKINA; HAMMER, 2017b; VOYANT *et al.*, 2017).

Nesse contexto, a incorporação de técnicas de ML e abordagens híbridas surge como alternativa promissora para reduzir incertezas, ajustar modelos físicos às condições reais e capturar não linearidades complexas presentes nas interações entre vento,  $G$  e dinâmica térmica. Tais métodos, quando apoiados em bases de dados de alta resolução e critérios de validação robustos, podem mitigar limitações de generalização dos modelos puramente físicos e ampliar a confiabilidade das estimativas termoelétricas em sistemas FV sob climas tropicais e semiáridos (VOYANT *et al.*, 2017; SKOPLAKI; PALYVOS, 2009). Adicionalmente, esta abordagem permite a calibração de correlações semiempíricas locais, ajustadas especificamente aos dados observados, expandindo a análise para além dos modelos tradicionais como *Ross*, *Faiman* e *PVsyst*.

Modelos baseados em dados dispensam a formulação física explícita e são capazes de capturar relações não lineares e interações complexas entre variáveis climáticas e operacionais (ZOUINE *et al.*, 2019). (CAMPOS *et al.*, 2020a) demonstraram a eficácia de modelos como o *K-Nearest Neighbors* (K-Vizinhos Mais Próximos – KNN) para estimativas de temperatura, utilizando variáveis como  $G$ ,  $T_a$  e  $V_w$ . Uma vantagem importante do ML é a possibilidade de adaptar o modelo aos dados específicos de cada local, melhorando a acurácia em comparação

com modelos genéricos.

Modelos como Regressão Linear (RL), *Random Forest* (Floresta Aleatória –RF) e *extreme Gradient Boosting* (Reforço de Gradiente Extremo – XGBoost) são amplamente utilizados devido à boa capacidade preditiva, flexibilidade e relativa interpretabilidade (RENKEMA *et al.*, 2024b). Embora técnicas mais complexas, como redes neurais profundas, sejam populares em outras áreas, há evidências de que modelos como RF podem igualar ou até superar o desempenho de *Long Short-Term Memory* (Memória Longa de Curto Prazo – LSTM) em aplicações voltadas à previsão de geração FV (JAIN *et al.*, 2022). Assim, algoritmos menos complexos continuam adequados para tarefas de previsão com dados meteorológicos.

O objetivo geral deste trabalho é implementar, validar e comparar uma estrutura integrada para a estimativa de  $T_{mod}$  em módulos FV instalados em clima semiárido. Para isso, são utilizados dados experimentais coletados entre maio de 2023 e maio de 2025 no Laboratório de Energias Alternativas da Universidade Federal do Ceará (LEA-UFC), em Fortaleza (CE). A pesquisa combina e compara modelos físico-semiempíricos consolidados na literatura (Ross, Faiman e PVsyst), correlações híbridas otimizadas por dados e modelos de ML (como RL, KNN, RF e XGBoost).

Para alcançar o objetivo geral, definem-se objetivos específicos estruturados de forma sequencial. Inicialmente, busca-se consolidar uma base de dados de qualidade mediante procedimentos de pré-processamento, que incluem a detecção de outliers físicos, imputação seletiva e verificação de consistência entre  $G$  e ângulos solares (*zenith* e *azimuth*). Em seguida, propõe-se treinar, calibrar e comparar algoritmos de ML com exploração sistemática de hiperparâmetros e validação temporal, respeitando a estrutura sequencial das séries de dados.

Além disso, o trabalho visa implementar e reavaliar os modelos físico-semiempíricos clássicos sob as condições do semiárido local. Isso estabelece uma referência para quantificar os ganhos de desempenho proporcionados pelos métodos de ML. O estudo também pretende formular, calibrar e otimizar correlações semiempíricas adicionais, empregando métodos numéricos de ajuste robusto. Por fim, objetiva-se realizar análises de importância de variáveis para interpretar os mecanismos físico-estatísticos subjacentes, derivando recomendações operacionais para mitigação térmica em sistemas FV urbanos.

Este trabalho visa contribuir para o avanço do conhecimento científico sobre a modelagem térmica de módulos FV em condições de clima semiárido tropical. Essa região é caracterizada por altos valores de  $G$  e temperaturas elevadas. Tais condições desafiam a

aplicabilidade de modelos físicos cujas parametrizações originais foram calibradas sob condições idealizadas, em áreas abertas e climas temperados. A aplicação direta desses modelos clássicos a ambientes urbanos obstruídos em regiões tropicais resulta em desvios substanciais (IEA PVPS Task 13, 2017; WHEELER *et al.*, 2021; BARRY *et al.*, 2020).

A caracterização de dois ciclos sazonais completos de dados experimentais, aliada à avaliação sistemática dos paradigmas de modelagem, oferece uma base científica adequada para a otimização de sistemas FV no semiárido. As metodologias desenvolvidas têm potencial de aplicação direta no dimensionamento de sistemas de resfriamento e na implementação de estratégias de controle preditivo. Espera-se, com isso, contribuir para a viabilização técnico-econômica de instalações FV em larga escala no cenário brasileiro.

## 2 FUNDAMENTAÇÃO TEÓRICA

A temperatura de operação do módulo ( $T_{mod}$ ) é uma variável de estado central para a avaliação e previsão de desempenho de sistemas FV, pois afeta diretamente a tensão de circuito aberto ( $V_{oc}$ ), o fator de forma ( $FF$ ), a potência máxima ( $P_{max}$ ) e a eficiência ( $\eta$ ); a corrente de curto-circuito ( $I_{sc}$ ) sofre apenas acréscimos marginais com o aquecimento (SKOPLAKI; PALYVOS, 2009; CUCE *et al.*, 2017; SANTOS *et al.*, 2021). O mecanismo físico subjacente a essa queda de desempenho termodinâmico decorre da redução do *bandgap* efetivo com a elevação de  $T_{mod}$ , o que intensifica as taxas de recombinação, reduz o potencial químico e desloca a curva I-V para tensões menores (RODRIGO *et al.*, 2014; FERNANDEZ *et al.*, 2013). Por ditar diretamente o balanço energético real da planta, desvios de poucos graus Celsius na estimativa de  $T_{mod}$  em condições de operação de campo propagam-se em erros significativos no cálculo de perdas energéticas acumuladas ao longo da vida útil do sistema.

Para módulos cristalinos, coeficientes de temperatura típicos situam-se na faixa de  $-0,03\%$  a  $-0,06\%$  de perda absoluta de  $\eta$  por grau Celsius (CUCE *et al.*, 2017; RAHMAN *et al.*, 2015). Considerando coeficiente de potência ( $\gamma_p$ ) de aproximadamente  $-0,35$  a  $-0,50\%/^{\circ}\text{C}$ , um erro de  $5^{\circ}\text{C}$  na estimativa de  $T_{mod}$  pode induzir incerteza de  $1,75$ – $2,5\%$  em  $P_{max}$ , magnitude relevante em análises de desempenho energético e retorno econômico, especialmente em climas semiáridos onde picos de  $T_{mod}$  superiores a  $70^{\circ}\text{C}$  são possíveis (SANTOS *et al.*, 2021; KING *et al.*, 2004).

Apesar dessa relevância,  $T_{mod}$  não é facilmente acessível de forma direta e contínua, tornando-se um dos parâmetros mais desafiadores na caracterização térmica de módulos FV (SEGADO *et al.*, 2015). A instrumentação intrusiva (sensores sob células) é pouco usual em larga escala, e medições no dorso introduzem incertezas de alguns graus. Assim, a literatura enfatiza reiteradamente os efeitos adversos do aquecimento sobre o desempenho elétrico (MORA *et al.*, 2017; CUCE *et al.*, 2017), reforçando a necessidade de modelos de estimação robustos.

A temperatura de operação resulta de um sistema multifatorial que combina propriedades ópticas e térmicas dos materiais (refletância, emitância, condutividade), arquitetura construtiva (camadas, encapsulantes, espaçamento traseiro), condições de instalação (inclinação, ventilação, integração a edificações) e variáveis meteorológicas primárias, destacando-se  $G$ ,  $T_a$  e  $V_w$  (GARCÍA; BALENZATEGUI, 2004; MARION *et al.*, 1999). Em climas com altos valores de  $G$ , mas com  $V_w$  irregular, a interação desses fatores pode amplificar gradientes térmicos e acelerar perdas de desempenho.

Dada a dificuldade de medição direta e a importância de  $T_{mod}$  como variável intermediária, diferentes abordagens de estimação foram desenvolvidas: (i) correlações semiempíricas que expressam  $T_{mod}$  em função de  $G$ ,  $T_a$ ,  $V_w$  e parâmetros característicos; (ii) formulações baseadas em balanço energético (explícitas ou implícitas), que integram conversão fotônica, convecção e radiação; e (iii) métodos baseados em parâmetros elétricos (uso de  $V_{oc}$  ou curvas I–V para retroestimar  $T_{mod}$ ) (JAKHRANI *et al.*, 2011; FERNANDEZ *et al.*, 2013; RODRIGO *et al.*, 2014). Essas classes incorporam, em diferentes níveis, propriedades materiais e características do arranjo, servindo de base para etapas subsequentes de modelagem de geração de potência.

Este capítulo está organizado da seguinte forma: a Seção 2.1 apresenta os fundamentos do balanço energético e dos mecanismos de transferência de calor em módulos FV; a Seção 2.2 discute os fatores determinantes da temperatura de operação; a Seção 2.3 aborda os modelos de estimativa de temperatura, consolidando a classificação geral, modelos físicos estacionários e dinâmicos, modelos baseados em dados, síntese das correlações da literatura e os modelos físicos de referência utilizados neste trabalho; a Seção 2.4 trata dos modelos baseados em ML, incluindo algoritmos, vantagens, limitações e aspectos de interpretabilidade; a Seção 2.5 descreve técnicas e limitações da medição de  $T_{mod}$ ; a Seção 2.3.3.4 analisa as limitações estruturais das correlações convencionais; a Seção 2.6 apresenta considerações específicas para clima semiárido; e, finalmente, a Seção 2.7 identifica lacunas na literatura e apresenta a justificativa do estudo.

## 2.1 Balanço de energia e mecanismos de transferência de calor

O estado térmico do módulo é governado pelo balanço de energia entre a parcela da irradiância absorvida que não é convertida em eletricidade (devido às limitações de eficiência quântica e elétrica) e os mecanismos de dissipação de calor para o meio externo por convecção, radiação e, de forma secundária, condução estrutural (SKOPLAKI; PALYVOS, 2009; DUFFIE; BECKMAN, 2013). Esse balanço de energia pode ser formulado matematicamente em regime estacionário pela seguinte equação:

$$(1 - \eta)\alpha G = U_L(T_{mod} - T_a) + \varepsilon\sigma(T_{mod}^4 - T_{sky}^4) \quad (2.1)$$

onde  $\eta$  é a eficiência de conversão energética do módulo,  $\alpha$  é a absorptância solar efetiva da superfície,  $U_L$  é o coeficiente global de perdas sensíveis (convecção + condução equivalente)

(DUFFIE; BECKMAN, 2013; DAVIS *et al.*, 2001),  $\varepsilon$  é a emissividade no infravermelho,  $\sigma$  é a constante de Stefan-Boltzmann e  $T_{sky}$  é a temperatura radiativa efetiva do céu. A magnitude de  $U_L$  é fortemente modulada por  $V_w$  e pela configuração de montagem do arranjo fotovoltaico (FAIMAN, 2008; TINA; ABATE, 2008)<sup>1</sup>.

Na maioria dos casos, esses mecanismos afetam tanto a face frontal quanto a traseira do módulo, uma vez que em instalações típicas mantém-se uma distância entre o módulo e o telhado para facilitar a remoção do calor rejeitado, permitindo que o módulo opere de forma eficiente. No caso de arranjos em estruturas abertas (*free-standing* ou *open rack*), pode-se considerar também a condução de calor através da estrutura de suporte: o calor é transportado para as superfícies da estrutura por condução, que, por sua vez, o dissipam para o ambiente por convecção e radiação (SANTOS *et al.*, 2021). Em instalações convencionais sobre estruturas abertas, a resistência térmica ao escoamento é reduzida, ao passo que montagens aderidas ou integradas à edificação elevam  $T_{mod}$  pela restrição convectiva (DAVIS *et al.*, 2001; ZHANG *et al.*, 2017).

De fato,  $T_{mod}$  depende de diversos fatores, incluindo: propriedades de absorção solar das células; materiais constituintes dos módulos (semicondutores, células, camadas, encapsulante, entre outros); dissipação térmica para o ambiente; condições de instalação e ambientais (localização,  $G$ ,  $T_a$ , e  $V_w$ ) (SANTOS *et al.*, 2021).

## 2.2 Fatores determinantes da temperatura de operação

A sensibilidade de  $T_{mod}$  aos preditores meteorológicos segue hierarquia consolidada:  $G$  explica a maior fração de variância, com incrementos de 1,8–4,9 °C em  $T_{mod}$  para cada 100 W/m<sup>2</sup> adicionais dependendo da tecnologia e montagem (RAHMAN *et al.*, 2015; SKOPLAKI; PALYVOS, 2009); a temperatura ambiente  $T_a$  apresenta coeficiente próximo da unidade parcial ( $\approx 1$  °C de acréscimo em  $T_{mod}$  por 1 °C em  $T_a$ ) (GARCÍA; BALENZATEGUI, 2004; SANTOS *et al.*, 2021);  $V_w$  induz reduções não lineares, com quedas relatadas de aproximadamente 3,3 °C ao aumentar de 1 para 2 m/s (TINA; ABATE, 2008; FAIMAN, 2008).

Propriedades ópticas e elétricas ( $\alpha$ ,  $\tau$ ,  $\eta$ ) entram ora explicitamente, ora via produto  $\tau\alpha$  ou termo de perdas  $[1 - (\tau\alpha\eta)]$ , refletindo o fracionamento espectral e a fração convertida em calor (CUCE *et al.*, 2017; FERNANDEZ *et al.*, 2013). O conceito de irradiância efetiva (*effective*

<sup>1</sup> Em formulações alternativas focadas nos fenômenos ópticos e convectivos isolados, o balanço também pode ser expresso assumindo a forma  $(1 - \rho)G - \eta G = h_c(T_{mod} - T_a) + \varepsilon\sigma(T_{mod}^4 - T_{sky}^4)$ , onde  $h_c$  é o coeficiente convectivo e  $\rho$  representa a refletância efetiva do módulo.

$G$ ) corrige diferenças angulares, espectrais e de sujeira entre o sensor e o módulo, mitigando erro sistemático em correlações e modelos de potência (KING *et al.*, 2004; PÉREZ-HIGUERAS *et al.*, 2012).

### 2.3 Modelos de estimativa de temperatura

As abordagens para modelar  $T_{mod}$  classificam-se tradicionalmente em modelos estacionários (*steady-state*), modelos dinâmicos e métodos baseados em dados ou ML (SKOPLAKI; PALYVOS, 2009; COSKUN *et al.*, 2017).

Em condições padrões de teste (STC, do inglês *Standard Test Conditions*), os módulos FV são caracterizados por parâmetros de desempenho definidos em ensaios laboratoriais ou disponibilizados pelo fabricante em suas fichas técnicas. Essas especificações constituem referência fundamental para o dimensionamento de usinas FV, pois permitem comparar o comportamento em diferentes condições de operação a partir de um ponto de referência comum (DOWS; GOUGH, 1995). Entre esses parâmetros, destaca-se a NMOT, que idealmente deveria ser reproduzível independentemente do laboratório, local ou período do ano em que o ensaio fosse realizado (KUITCHE *et al.*, 2011).

Sob a ótica matemática, os modelos de  $T_{mod}$  podem ser classificados em dois grupos: explícitos e implícitos (MOHAMMED *et al.*, 2019b). Nos modelos explícitos, a  $T_{mod}$  é obtida diretamente a partir de equações analíticas; já os implícitos envolvem variáveis que dependem da própria  $T_{mod}$ , exigindo procedimentos iterativos baseados no balanço de energia do módulo. Um exemplo de modelo explícito relaciona a  $T_{mod}$  à temperatura de fundo do módulo ( $T_b$ ), sendo a diferença entre ambas dependente do material do substrato e do nível de irradiância ( $G$ ). Esse tipo de formulação pode fornecer estimativas de  $T_{mod}$  com incertezas em torno de  $\pm 5$  °C, o que representa impacto inferior a 3% na potência de saída do módulo (KING *et al.*, 2004):

$$T_{mod} = T_b + \frac{G}{G_{ref}} \Delta T \quad (2.2)$$

em que  $G$  é a irradiância incidente no módulo FV ( $\text{W/m}^2$ ),  $G_{ref} = 1000 \text{ W/m}^2$  corresponde às condições STC, e  $\Delta T$  representa a diferença característica entre  $T_{mod}$  e  $T_b$  em  $G_{ref}$ . Em condições de montagem aberta (*open rack*),  $T_{mod} - T_b$  pode variar alguns graus Celsius, convergindo a zero em módulos com isolamento posterior ou quando a resistência térmica interna é pequena frente às perdas convectivas externas (KING *et al.*, 2004).

Uma formulação mais geral que incorpora a eficiência de conversão é dada por:

$$T_{mod} = T_a + \left( \frac{G}{G_{NMOT}} \right) \left( \frac{U_{L,NMOT}}{U_L} \right) (T_{NMOT} - T_{a,NMOT}) \left[ 1 - \frac{\eta}{\tau\alpha} \right] \quad (2.3)$$

na qual, em condições NMOT, o módulo opera em circuito aberto, de modo que  $\eta = 0$  (DUFFIE; BECKMAN, 2013). Como  $\eta$  depende de  $T_{mod}$ , essa expressão torna-se implícita, exigindo solução iterativa. Esta formulação evidencia o acoplamento entre desempenho elétrico e térmico: a fração da energia solar convertida em eletricidade ( $\eta$ ) reduz o aquecimento do módulo, enquanto a fração não convertida ( $1 - \eta/\tau\alpha$ ) contribui para a elevação térmica.

Na ausência de informações fornecidas pelo fabricante, os parâmetros necessários podem ser obtidos em testes de campo sob condições reais de operação (KING *et al.*, 1997). Em aplicações de sistemas FV integrados a edificações (BIPV, do inglês *Building-Integrated Photovoltaics*), os módulos são posicionados a uma distância da fachada que cria um regime térmico distinto entre a face frontal e posterior. Nesses casos, o modelo simplificado baseado apenas em NMOT pode subestimar a  $T_{mod}$  em até 20 Kelvins (K) (DAVIS *et al.*, 2001), evidenciando a necessidade de calibração local dos coeficientes.

### 2.3.1 Classificação geral

A literatura especializada classifica os modelos de estimativa de  $T_{mod}$  em três categorias principais, diferenciadas pela complexidade matemática, demanda computacional e capacidade de representar fenômenos transientes (SKOPLAKI; PALYVOS, 2009; COSKUN *et al.*, 2017). Os modelos físicos estacionários assumem condições de regime permanente e são caracterizados por simplicidade e baixo custo computacional, sendo adequados para análises em escalas de tempo superiores a alguns minutos. Os modelos dinâmicos incorporam a capacitância térmica do módulo e equações diferenciais temporais, capturando variações rápidas de irradiância e vento, porém com maior complexidade e necessidade de parâmetros experimentais detalhados. Por fim, os métodos baseados em dados, especialmente técnicas de ML, representam abordagem empírica que dispensa modelagem fenomenológica explícita, adaptando-se aos padrões climáticos locais através de calibração estatística. Esta seção detalha as características, vantagens e limitações de cada abordagem, fornecendo contexto para a seleção dos modelos de referência implementados neste trabalho.

### 2.3.1.1 Modelos físicos estacionários

Os modelos estacionários são relativamente simples e caracterizados por baixos tempos computacionais, assumindo invariância instantânea das variáveis térmicas. No entanto, esses modelos não são flexíveis e podem superestimar ou subestimar o valor de  $T_{mod}$  sob condições climáticas variáveis (KAPLANIS; KAPLANI, 2018; KING *et al.*, 2004; LOBERA; VALKEALAHTI, 2013).

A intensidade de  $G$  incidente e outros parâmetros que afetam o desempenho dos módulos FV, em um curto período de tempo, são supostos constantes. Para intervalos de tempo de 1 minuto, uma abordagem de estado estacionário pode ser apropriada (JONES; UNDERWOOD, 2001); contudo, se a variação da transferência de calor entre o módulo FV e o ambiente for significativa, modelos mais complexos tornam-se necessários. Temperaturas em diferentes posições do módulo FV também são consideradas constantes para a condição de estado estacionário, uma simplificação que pode não refletir adequadamente os fenômenos térmicos transientes que ocorrem em condições reais de operação (ALY *et al.*, 2018).

Estudos comparativos demonstram que a diferença de temperatura prevista ( $\Delta T$ ) pelos modelos estacionários pode exceder limites aceitáveis quando os sistemas são submetidos a condições climáticas altamente variáveis, evidenciando a necessidade de modelos térmicos dinâmicos baseados no balanço total de energia para aplicações que demandam maior precisão (LOBERA; VALKEALAHTI, 2013).

### 2.3.1.2 Modelos físicos dinâmicos

Modelos dinâmicos podem expressar resultados mais realistas e precisos ao considerar a flutuação da irradiância solar e outras variações ambientais em curtos períodos, incluindo  $T_a$ ,  $G$  e  $V_w$  (GU *et al.*, 2019). Essa técnica é baseada na determinação de  $T_{mod}$  usando um balanço de energia detalhado. Assim, modelos dinâmicos descrevem a  $T_{mod}$  e seus processos térmicos em maior detalhe quando comparados aos modelos estacionários, permitindo análises paramétricas mais rigorosas e a captura de variações temporais em sistemas complexos (HANNOUCH *et al.*, 2024; KUO *et al.*, 2017). No entanto, esses modelos são relativamente complexos e exigem significativamente mais tempo computacional, custo e esforço de implementação (KUO *et al.*, 2017). O desenvolvimento dessas equações requer um cuidadoso balanceamento entre precisão e eficiência computacional, especialmente em aplicações de sistemas FV integrados a edifícios,

onde a velocidade de cálculo é crítica (LUO *et al.*, 2017).

Esses modelos introduzem a capacidade térmica efetiva do empilhamento (encapsulante, células, vidro e camada posterior) para capturar atrasos e a suavização das variações de  $G$ , aumentando a fidelidade em escalas de minutos a horas (DUFFIE; BECKMAN, 2013; BARROS *et al.*, 2021). A equação dinâmica incorpora a capacitância térmica da seguinte forma:

$$C_{th} \frac{dT_{mod}}{dt} = (1 - \rho - \eta)G - h_c(T_{mod} - T_a) - \epsilon\sigma(T_{mod}^4 - T_{sky}^4) \quad (2.4)$$

onde  $C_{th}$  é a capacitância térmica ( $J/K$ ) e  $\rho$  é a refletância. A solução exata dessa equação demanda o conhecimento de propriedades térmicas e construtivas do empilhamento, como a área efetiva de troca, a composição das camadas, a massa específica e os coeficientes de transferência de calor (BARROS *et al.*, 2021). Contudo, esses parâmetros raramente estão disponíveis nas fichas técnicas comerciais ou são facilmente mensuráveis em campo, o que dificulta a aplicação prática generalizada desses modelos (SANTOS *et al.*, 2024; COSKUN *et al.*, 2017).

### 2.3.1.3 Modelos baseados em parâmetros elétricos

Os métodos baseados em parâmetros elétricos fundamentam-se na correlação analítica entre a tensão de circuito aberto ( $V_{oc}$ ), ou a combinação entre a corrente de curto-circuito ( $I_{sc}$ ) e a tensão, e a temperatura da célula. Essa abordagem permite inferir a  $T_{mod}$  diretamente a partir das curvas I–V, o que possibilita a dispensa de sensores térmicos físicos instalados no dorso do módulo. Contudo, a aplicação dessas técnicas depende da realização de medições elétricas em tempo real e do conhecimento prévio de coeficientes de temperatura específicos, usualmente fornecidos nas fichas técnicas dos fabricantes (RODRIGO *et al.*, 2014; INTERNATIONAL ELECTROTECHNICAL COMMISSION, 2016).

### 2.3.1.4 Modelos baseados em dados

Nos últimos anos, esforços consideráveis têm sido dedicados pela comunidade científica ao desenvolvimento e aprimoramento de abordagens para o cálculo de  $T_{mod}$  a partir de variáveis climáticas mensuráveis (SANTOS *et al.*, 2021). Nesse contexto, as técnicas de aprendizado de máquina surgem como uma alternativa promissora. Diferentemente das abordagens físicas ou analíticas descritas anteriormente, os modelos baseados em dados priorizam a identificação de padrões estatísticos em grandes volumes de informação, não requerendo

modelagem fenomenológica detalhada e oferecendo maior adaptabilidade (ZOUINE *et al.*, 2019; CAMPOS *et al.*, 2020b).

Algoritmos como RF têm demonstrado desempenho preditivo competitivo quando comparados a arquiteturas de redes neurais profundas (JAIN *et al.*, 2022), com a vantagem de oferecerem maior simplicidade e interpretabilidade técnica (RENKEMA *et al.*, 2024b). O diferencial prático dessa abordagem, em relação aos modelos clássicos, é a capacidade de adaptar as predições às características climáticas e operacionais específicas de cada local de instalação, melhorando substancialmente a acurácia em comparação com correlações genéricas da literatura.

### 2.3.2 Síntese das correlações da literatura

Em revisão da literatura conduzida por (SANTOS *et al.*, 2021), foram analisadas 33 correlações para estimativa de  $T_{mod}$  publicadas a partir do ano 2000, com observações pertinentes sobre suas condições de aplicabilidade, como mostra a Tabela 1. Essas equações foram desenvolvidas considerando geometrias e níveis de integração específicos; portanto, a escolha do modelo mais adequado depende essencialmente da aplicação, da disponibilidade de dados e do grau de precisão exigido. Os autores propuseram uma contribuição original ao sintetizar todas essas correlações em apenas três formas funcionais gerais.

As correlações lineares, que incluem 20 modelos explícitos e o modelo ISFOC (quando seus termos entre parênteses são considerados constantes), podem ser expressas pela forma geral, conforme Equação 2.5:

$$T_{mod} = a_0 + a_1 T_a + a_2 G + a_3 V_w + a_4 T_b \quad (2.5)$$

onde os coeficientes  $a_0, a_1, a_2, a_3$  e  $a_4$  são específicos para cada modelo. Esta forma abrange modelos amplamente utilizados como (ROSS, 1976), e variações mais recentes como (KOEHL *et al.*, 2011a) e (FAIMAN, 2008).

Os modelos não lineares, incluindo o modelo Sandia (quando  $\Delta T = (T_{mod} - T_b)G_{ref}$  é assumido constante), são representados pela Equação 2.6:

$$T_{mod} = b_0 + b_1 T_a + C_1 \left( e^{d_1 + d_2 V_w} + e_1 \right) G + C_2 (e_2 + f_1 T_a)^{g_1} (e_3 + f_2 G)^{g_2} (e_4 + f_3 V_w)^{g_3} \quad (2.6)$$

Os 16 parâmetros desta forma permitem capturar comportamentos complexos, incluindo interações não lineares entre variáveis. Vale ressaltar que, o modelo de (KURTZ *et al.*, 2009) emerge como caso particular do modelo de (HORNUNG *et al.*, 2012).

Tabela 1 – Equações para  $T_{mod}$  a partir do ano 2000

Nº	Nome da correlação	Formulação
1	Fernández et al.	$T_{mod} = \frac{V_{oc} - c_1(G - c_3)}{c_2}$
2	Durisch et al.	$T_{mod} = T_a + kG$
3	Nordmann and Clavadetscher	Igual ao anterior
4	Krauter	Igual ao anterior
5	Mondol et al. I	$T_{mod} = T_a + 0,031G$
6	Hove	$T_{mod} = T_a + G \left( \frac{T_0 - 1}{U_L} \right)$
7	Tiwari	$T_{mod} = T_a + G \left( \frac{\alpha}{U_L} \right) \left[ 1 - \left( \frac{\eta}{T_a} \right) \right]$
8	Eicker	$T_{mod} = T_a + G \left( \frac{\alpha}{U_L} \right) \left[ 1 - \frac{\eta}{\alpha} \right]$
9	Standard	$T_{mod} = T_a + \left( \frac{G}{G_{NMOT}} \right) (T_{c,NMOT} - T_{a,NMOT})$
10	Davis	$T_{mod} = T_a + \left( \frac{G}{G_{NMOT}} \right) (T_{c,NMOT} - T_{a,NMOT}) \left[ 1 - \left( \frac{V_w}{V_0} \right) \right]$
11	Mondol et al. II	$T_{mod} = T_a + 0,031G - 0,058$
12	Tselepis	$T_{mod} = 30 + 0,0175(G - 150) + 1,14(T_a - 25)$
13	Tiwari and Sodha I	$T_{mod} = \frac{pG(T_a - 1) + (T_a + U_T T_h)}{U_i + U_T}$
14	Tiwari and Sodha II	$T_{mod} = \frac{T[acp + a + (1 - Bc)G - ncGBc + U_i T_a + U_T T_b]}{U_i + U_T}$
15	Almonacid	$T_{mod} = T_a + d_1 G + d_2 V_w$
16	Markvart	$T_{mod} = 0,943T_a + 4,3 + 0,028G - 1,528V_w$
17	Muzathik	$T_{mod} = 0,943T_a + 0,3529 + 0,0195G - 1,528V_w$
18	Akyuz et al.	$T_{mod} = 0,95T_a + 3,1 + 0,025G - 0,3V_w$
19	NMOT-1p model	$T_{mod} = T_a + \left( \frac{G}{G_{NMOT}} \right) (T_{c,NMOT} - T_{a,NMOT}) + a(V_w - V_{w,NMOT})$
20	NMOT-2p model	$T_{mod} = T_a + b \left( \frac{G}{G_{NMOT}} \right) (T_{c,NMOT} - T_{a,NMOT}) + c(V_w - V_{w,NMOT})$
21	ISFOC method	$T_{mod} = T_b + \left( \eta C_g \sum \frac{I_u}{I} \right) G$
22	Faiman	$T_{mod} = T_a + \frac{G}{U_0 + U_1 V_w}$
23	Skoplaki and Palyvos	$T_{mod} = T_a + \frac{0,32}{8,91 + 2V_w} G$
24	Skoplaki et al. I	$T_{mod} = T_a + \frac{0,25}{5,7 + 3,8V_w} G$
25	Duffie and Beckman I	$T_{mod} = T_a + \frac{G(T_{c,NMOT} - T_{a,NMOT})}{G_{NMOT}} \left( \frac{9,5}{5,7 + 3,8V_w} \right) \left[ 1 - \left( \frac{\eta}{\tau_0} \right) \right]$
26	Skoplaki et al. II	$T_{mod} = T_a + \frac{G(T_{c,NMOT} - T_{a,ref})}{G_{NMOT}} \frac{h_{NMOT}}{h} \left\{ 1 - \frac{\gamma_{STC}}{T_{mod}} (1 + \beta_{STC} T_{STC}) \right\}$
27	Chenni et al.	$T_{mod} = T_a + 0,0138G(1 + 0,031T_a)(1 - 0,042V_w)$
28	Kurtz et al.	$T_{mod} = T_a + G e^{-3,473 - 0,0594V_w}$
29	Hornung et al.	$T_{mod} = T_a + m \left[ e^{\left( \frac{-0,5V_w}{V_{w0}} \right)} + c \right] G$
30	Coskun et al.	$T_{mod} = 1,4T_a + 0,01(G - 500) - V_w^{0,8}$
31	Sandia's model	$T_{mod} = T_b + \left( \frac{G}{G_{ref}} \right) \Delta T, T_b = G(e^{a+bV_w}) + T_a$
32	Duffie and Beckman II	$T_{mod} = \frac{T_a + \frac{G}{G_{NMOT}} (T_{c,NMOT} - T_{a,ref}) \left\{ 1 - \frac{\gamma_{STC}}{T_{mod}} (1 + \beta_{STC} T_{STC}) \right\}}{1 - \frac{\gamma_{STC}}{T_{mod}} \frac{G}{G_{ref,NTE}} (T_{c,NMOT} - T_{a,ref})}$
33	Mattei	$T_{mod} = \frac{UT_a + G(T_a - \gamma_{STC} - \beta_{STC} \gamma_{STC} T_{STC})}{U - \beta_{STC} \gamma_{STC} G}$

Fonte: Adaptado de (SANTOS *et al.*, 2021).

Os modelos de estrutura racional, especificamente Duffie & Beckman II e Mattei et al., seguem a forma da Equação 2.7:

$$T_{mod} = \frac{h_1 T_a + h_2 G}{m_1 + n_1 G} \quad (2.7)$$

Esta estrutura captura o comportamento assintótico da temperatura com o aumento da irradiância, representando fisicamente a saturação térmica em altos níveis de radiação.

A análise revela que entre as 33 correlações,  $G$  aparece em 100%,  $T_a$  em 93,9%,  $V_w$  em 54,5%,  $\eta$  em 33,3%, absorção solar em 30,3% e transmitância do vidro em 24,2% dos modelos. A revisão demonstra que a precisão preditiva está mais correlacionada com a qualidade da calibração local dos coeficientes do que com a sofisticação matemática da forma funcional. Modelos simples bem calibrados frequentemente superam formulações complexas com parametrização genérica. Cada correlação foi desenvolvida para configurações específicas (estrutura livre, BIPV, sistemas híbridos FV-T), exigindo cautela na seleção e aplicação dos modelos.

Esta taxonomia em três formas gerais facilita não apenas a compreensão unificada do conjunto de modelos disponíveis, mas também orienta estratégias de calibração padronizadas e seleção criteriosa de estruturas candidatas para aplicações específicas ou hibridização com técnicas de aprendizado de máquina.

### 2.3.3 Modelos físicos de referência

A literatura consolida diversos modelos empíricos e semiempíricos para a estimativa da temperatura de operação de módulos FV. Dentre as formulações clássicas de maior destaque, encontram-se os modelos de *Ross*, de *Faiman* e a abordagem adotada pelo *software PVsyst*. Essas correlações representam diferentes níveis de complexidade na modelagem térmica e são amplamente empregadas como padrão de referência na avaliação de desempenho de sistemas FV (SKOPLAKI; PALYVOS, 2009; WHEELER *et al.*, 2021; IEA PVPS Task 13, 2017).

#### 2.3.3.1 Modelo de Ross

A formulação linear de *Ross* (ROSS, 1976) estabelece uma relação direta entre o incremento de temperatura do módulo e a irradiância solar incidente:

$$T_{mod} = T_a + r \cdot G \quad (2.8)$$

onde  $r$  é o coeficiente de *Ross* ( $(^{\circ}\text{C}\cdot\text{m}^2)/\text{W}$ ), que representa o incremento de temperatura por unidade de irradiância. Este coeficiente depende do tipo de instalação e do espaço de ventilação entre o módulo e a superfície de suporte (NORDMANN; CLAVADETSCHER, 2003). Valores típicos de  $r$  variam entre 0,02 e 0,05 ( $(^{\circ}\text{C}\cdot\text{m}^2)/\text{W}$ ) para instalações tipo *open rack*, podendo atingir 0,06 ( $(^{\circ}\text{C}\cdot\text{m}^2)/\text{W}$ ) em sistemas integrados a edificações (SKOPLAKI; PALYVOS, 2009).

A simplicidade do modelo de *Ross* o torna amplamente utilizado para estimativas preliminares, embora não capture explicitamente o efeito de resfriamento convectivo do vento.

### 2.3.3.2 Modelo de Faiman

O modelo proposto por *Faiman* (FAIMAN, 2008) incorpora explicitamente o efeito de resfriamento convectivo dependente da velocidade do vento, modelando a taxa de transferência de calor como função das condições ambientais:

$$T_{mod} = T_a + \frac{G}{U_0 + U_1 \cdot V_w} \quad (2.9)$$

onde  $U_0$  é o coeficiente de transferência de calor base ( $\text{W}/(\text{m}^2\cdot\text{K})$ ), representando perdas por convecção natural e radiação, e  $U_1$  é o coeficiente de transferência de calor dependente do vento ( $\text{W}\cdot\text{s}/(\text{m}^3\cdot\text{K})$ ). Valores reportados na literatura para instalações *open rack* indicam  $U_0 \approx 25 \text{ W}/(\text{m}^2\cdot\text{K})$  e  $U_1 \approx 6,84 \text{ W}\cdot\text{s}/(\text{m}^3\cdot\text{K})$  (FAIMAN, 2008).

Este modelo representa um avanço em relação ao modelo de *Ross* ao incorporar explicitamente o efeito da ventilação. Contudo, seus parâmetros foram calibrados em condições climáticas temperadas, sendo necessária verificação de sua aplicabilidade direta a climas semiáridos, onde as características de transferência de calor convectivo podem diferir significativamente devido à menor umidade relativa e maior amplitude térmica diária (KOEHL *et al.*, 2011b).

### 2.3.3.3 Modelo PVsyst

O modelo implementado no *software PVsyst* (MERMOUD, 2012) utiliza uma formulação de balanço de energia térmico análoga ao modelo de *Faiman*, mas incorpora explicitamente a parcela da irradiância que não é convertida em eletricidade ou refletida:

$$T_{mod} = T_a + \frac{\alpha \cdot G \cdot (1 - \eta)}{u_c + u_v \cdot V_w} \quad (2.10)$$

onde  $\alpha$  é o coeficiente de absorção do módulo e  $u_c$  e  $u_v$  são os coeficientes de perda térmica constante e convectiva ( $\text{W}/(\text{m}^2 \cdot \text{K})$ ) e ( $\text{W} \cdot \text{s}/(\text{m}^3 \cdot \text{K})$ ), respectivamente. Os valores desses fatores de perda térmica podem ser adotados a partir das recomendações padrão do *software* ou calibrados para condições locais específicas. A NMOT é definida em condições ambientais nominais:  $G_{NMOT} = 800 \text{ W/m}^2$ ,  $T_{a,NMOT} = 20^\circ\text{C}$ ,  $V_{w,NMOT} = 1 \text{ m/s}$ , em circuito aberto (ASTM INTERNATIONAL, 1996).

A versão simplificada sem dependência do vento ( $u_v = 0$ ) reduz-se a:

$$T_{mod} = T_a + \frac{G}{G_{NMOT}}(T_{NMOT} - T_{a,NMOT}) \quad (2.11)$$

Esta formulação é amplamente utilizada em softwares comerciais de simulação FV, oferecendo estimativas baseadas em parâmetros de catálogo dos fabricantes (STULTZ; WEN, 1977). Contudo, estudos demonstram que as condições padrão podem não ser representativas para diferentes regiões climáticas, evidenciando a necessidade de calibração local dos coeficientes  $u_c$  e  $u_v$  (HANIFAH *et al.*, 2018; PACHECO *et al.*, 2015). Em aplicações de sistemas FV integrados a edificações (BIPV), onde os módulos são instalados próximos à fachada, o modelo NMOT simples pode subestimar  $T_{mod}$  significativamente devido à redução da ventilação traseira e ao acoplamento térmico com a envoltória (AMODIO *et al.*, 2023).

Extensões modernas destes modelos incluem termos adicionais para correção de eficiência e ajustes empíricos localmente calibrados, aproximando-se de formulações híbridas que combinam fundamentos físicos com otimização estatística (MATTEI *et al.*, 2006).

#### 2.3.3.4 Limitações estruturais das correlações convencionais

As limitações estruturais das correlações estacionárias incluem: (i) necessidade de recalibração local dos coeficientes devido à fraca transferibilidade entre climas distintos (MOHAMMED *et al.*, 2019b; SANTOS *et al.*, 2021); (ii) supressão de efeitos transitórios rápidos de nuvens que induzem *overshoot* ou defasagens temporais (COSKUN *et al.*, 2017; BARROS *et al.*, 2021); (iii) sub-representação de variáveis latentes como emissividade de céu, umidade relativa, deposição de poeira e envelhecimento óptico (SKOPLAKI; PALYVOS, 2009; SANTOS *et al.*, 2024).

Modelos convencionais como *Ross* e *Faiman* apresentam correlação mínima entre variáveis em condições específicas, sendo inadequados para aplicações não-convencionais (PA-

TEL *et al.*, 2024a). O modelo NMOT, embora mais próximo aos valores experimentais, ainda demonstra limitações significativas na predição de perdas térmicas que podem atingir 6-27% da potência nominal (CHANDEL *et al.*, 2024).

Mesmo assim, sua interpretabilidade física e baixa exigência de dados asseguram relevância como *baseline* de referência, incluindo correlações de Ross, NMOT ajustado, Mattei e Faiman, nas avaliações comparativas com abordagens modernas (MATTEI *et al.*, 2006; FAIMAN, 2008; SANTOS *et al.*, 2021).

## 2.4 Modelos baseados em ML

A adoção de métodos de ML e de modelos híbridos físico-estatísticos amplia consideravelmente as capacidades preditivas em relação às abordagens térmicas convencionais. Essas técnicas permitem a captura de relações não lineares complexas sem a necessidade de parametrização física explícita. Além disso, possibilitam a inclusão flexível de variáveis adicionais, como ângulos solares e histórico temporal de dados, bem como a identificação e extração automática de interações entre as variáveis preditoras (JAIN *et al.*, 2022; RENKEMA *et al.*, 2024b).

### 2.4.1 Algoritmos comumente utilizados

**RL:** Serve como modelo *baseline* devido à sua simplicidade computacional e interpretabilidade direta dos coeficientes (SCHIELZETH, 2010), embora seja limitada para capturar relações não-lineares complexas presentes na modelagem térmica de células FV. A RL múltipla pode ser expressa como:

$$\hat{T}_{mod} = \beta_0 + \beta_1 G + \beta_2 T_a + \beta_3 V_w + \varepsilon \quad (2.12)$$

onde  $\beta_i$  são os coeficientes estimados por mínimos quadrados ordinários. Apesar de sua simplicidade, fornece uma base interpretável para comparação com modelos mais complexos e permite identificação rápida de variáveis significativas através de testes estatísticos. Sua principal limitação reside na incapacidade de modelar interações não-lineares entre variáveis meteorológicas e efeitos de saturação térmica.

**KNN:** Modelo não-paramétrico baseado na similaridade entre amostras no espaço das variáveis preditoras, utilizando a distância euclidiana ou outras métricas para identificar os  $k$  vizinhos mais próximos (TODESCHINI, 1989). A predição é calculada como:

$$\hat{T}_{mod} = \frac{1}{k} \sum_{i=1}^k T_{c,i} \quad (2.13)$$

onde  $T_{mod,i}$  são as temperaturas dos  $k$  vizinhos mais próximos. (CAMPOS *et al.*, 2020b) demonstraram sua eficácia para estimativas de temperatura utilizando variáveis como irradiância, temperatura ambiente e velocidade do vento. O KNN apresenta vantagens na captura de padrões locais complexos sem assumir forma funcional específica, sendo particularmente útil quando existem regimes operacionais distintos (baixa/alta irradiância, diferentes estações). Entretanto, sua performance é sensível à escolha de  $k$  e à presença de variáveis irrelevantes, além de apresentar alto custo computacional em datasets grandes devido à necessidade de calcular distâncias para todas as amostras de treinamento.

**RF:** Modelo *ensemble* baseado em múltiplas árvores de decisão construídas através de *bootstrap aggregating* (bagging) e seleção aleatória de variáveis em cada nó. A predição final é obtida pela média das predições individuais:

$$\hat{T}_{mod} = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (2.14)$$

onde  $B$  é o número de árvores e  $T_b(x)$  é a predição da  $b$ -ésima árvore. Apresenta boa capacidade de generalização e é naturalmente robusto ao *overfitting* devido ao processo de agregação (BREI-MAN, 2001). Permite análise de importância das variáveis através de medidas como diminuição da impureza ou *permutation importance*, aumentando significativamente sua interpretabilidade. Estudos comparativos demonstram que RF supera modelos lineares na captura de interações complexas entre variáveis meteorológicas, sendo particularmente eficaz em modelar efeitos de saturação térmica e não-linearidades em diferentes regimes de irradiância (LEDMAOUI *et al.*, 2023). Sua robustez a *outliers* e capacidade de lidar com variáveis correlacionadas o tornam adequado para dados meteorológicos ruidosos.

**XGBoost:** Modelo *ensemble* baseado em *gradient boosting* de árvores de decisão, que constrói sequencialmente árvores para corrigir erros das anteriores. A função objetivo otimizada inclui tanto o erro de predição quanto termos de regularização:

$$\mathcal{L} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (2.15)$$

onde  $l$  é a função de perda,  $\Omega(f_k)$  são termos de regularização das árvores individuais, e  $K$  é o número total de árvores. É conhecido por sua alta capacidade preditiva em problemas com relações não-lineares complexas e tem apresentado desempenho superior em séries temporais e problemas ambientais (CHEN; GUESTRIN, 2016). Estudos comparativos em sistemas FV demonstram que XGBoost supera consistentemente RF e outros algoritmos na previsão de potência e parâmetros térmicos (DIDAVI *et al.*, 2021; OBIORA *et al.*, 2021). Suas principais vantagens incluem: (i) capacidade de capturar interações complexas de alta ordem entre variáveis meteorológicas; (ii) robustez a dados desbalanceados e *outliers*; (iii) eficiência computacional através de otimizações algorítmicas; (iv) flexibilidade na definição de funções objetivo customizadas. A combinação de XGBoost com RF em arquiteturas *ensemble* empilhadas tem mostrado melhorias adicionais na precisão preditiva (BANIK; BISWAS, 2024), sugerindo complementaridade entre as abordagens de *bagging* e *boosting* para modelagem térmica de sistemas FV.

#### 2.4.2 Vantagens e limitações dos modelos ML

As principais vantagens dos modelos de ML incluem:

- Capacidade de capturar relações não-lineares e interações complexas entre variáveis meteorológicas
- Adaptabilidade aos dados específicos de cada local e condições operacionais
- Dispensam formulação física explícita detalhada dos processos térmicos
- Robustez a ruídos e *outliers* (especialmente modelos *ensemble*)
- Otimização automática de hiperparâmetros através de técnicas AutoML (BAKHT *et al.*, 2025)

As limitações incluem:

- Menor interpretabilidade física comparada aos modelos semi-empíricos (LIU *et al.*, 2022)
- Necessidade de grandes volumes de dados para treinamento adequado
- Risco de *overfitting* sem validação cruzada apropriada (TUCCI *et al.*, 2024; CANTILLO-LUNA *et al.*, 2023)
- Dificuldade de extrapolação para condições não observadas nos dados de treino
- Tendência à memorização de padrões específicos sem capturar relações fundamentais

### 2.4.3 Interpretabilidade e coerência físico-estatística

Resultados recentes confirmam convergência entre a hierarquia física de relevância (predomínio de  $G$  e  $T_a$ , participação moderada de  $V_w$ ) e as importâncias obtidas em modelos ML, reforçando a coerência físico-estatística (SANTOS *et al.*, 2021; RENKEMA *et al.*, 2024b). Técnicas de interpretabilidade como SHAP (*SHapley Additive exPlanations*) permitem identificar a contribuição de cada parâmetro ambiental nas previsões, melhorando a transparência dos modelos (BAKHT *et al.*, 2025).

A avaliação adequada desses modelos requer comparação estruturada contra correlações semi-empíricas calibradas, evitando superestimação de ganhos decorrentes de *splits* aleatórios sem respeito à autocorrelação temporal (MOHAMMED *et al.*, 2019b; RENKEMA *et al.*, 2024b). Taxas de aprendizado adequadas são cruciais para equilibrar velocidade de convergência e capacidade de generalização (TUCCI *et al.*, 2024).

## 2.5 Medição de $T_{mod}$ : técnicas e limitações

A medição de  $T_{mod}$  enfrenta limitações instrumentais e conceituais: sensores internos sob a célula fornecem referência de alta precisão ( $\pm 1$  °C), porém são intrusivos e raros em operação contínua (YANDT *et al.*, 2012; INTERNATIONAL ELECTROTECHNICAL COMMISSION, 2016). A abordagem com termopares no dorso, distribuídos para capturar heterogeneidade espacial (tipicamente variação de até 5 °C entre centro e bordas), constitui prática padrão e balanceia custo e precisão (KING *et al.*, 2004; FAIMAN, 2008).

Termografia infravermelha auxilia na identificação de *hotspots* e dispersão térmica, mas depende de emissividade calibrada e controle de reflexões (SKOPLAKI; PALYVOS, 2009). Métodos elétricos  $I_{sc}$ - $V_{oc}$  exploram coeficientes de temperatura de  $V_{oc}$  e  $P_{max}$  para inferir  $T_{mod}$ , exigindo aquisição de curvas I-V de alta qualidade e correção espectral (RODRIGO *et al.*, 2014; INTERNATIONAL ELECTROTECHNICAL COMMISSION, 2016).

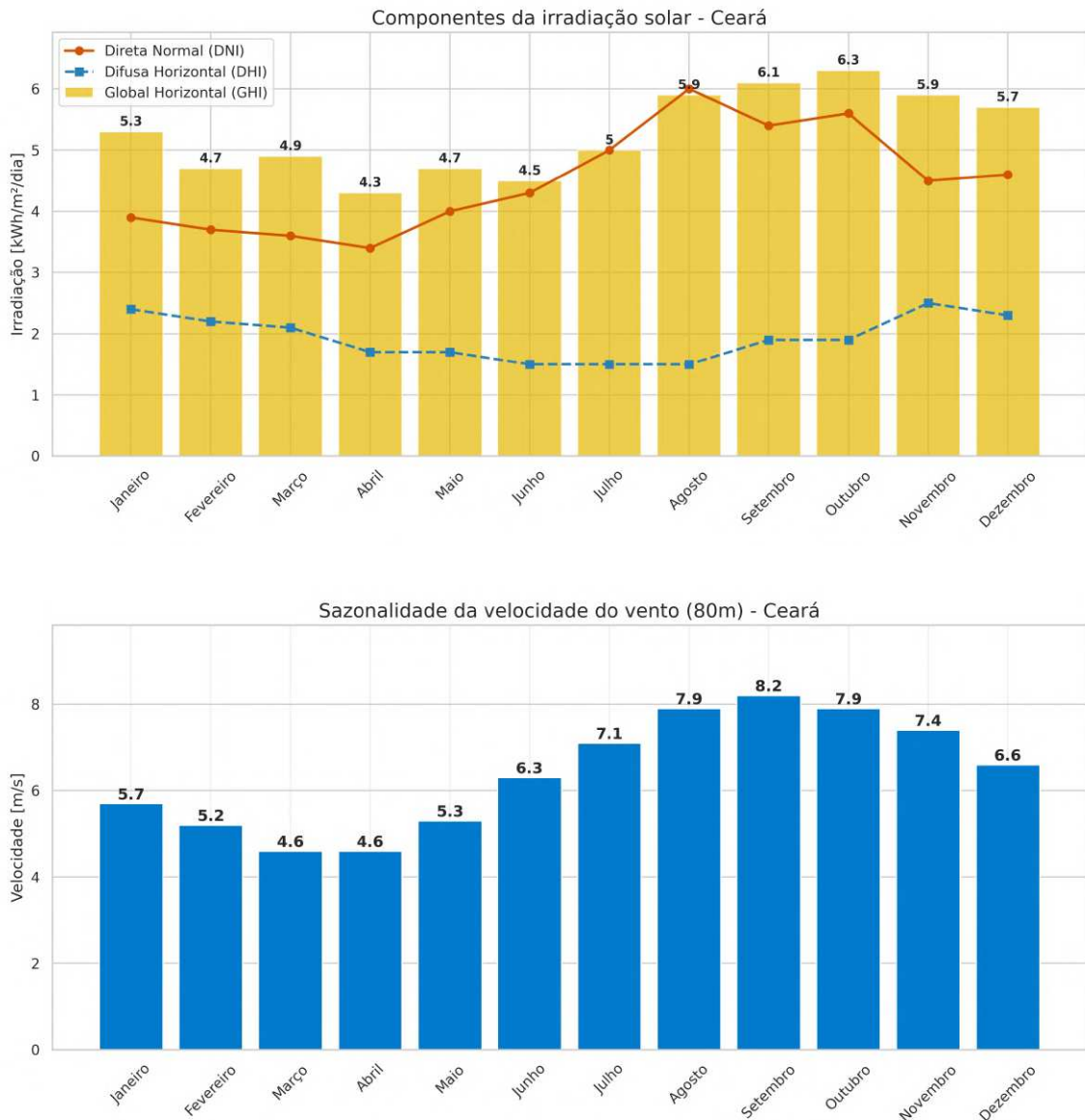
Em todas as abordagens, discrepâncias entre o ângulo de visão do sensor de irradiação e o plano efetivo das células, além de diferenças de resposta espectral, compõem fontes de incerteza que propagam para o erro de  $T_{mod}$  (KING *et al.*, 2004; RAHMAN *et al.*, 2015).

## 2.6 Considerações para clima semiárido

O clima semiárido brasileiro impõe desafios singulares à modelagem FV, frequentemente expondo as limitações de correlações desenvolvidas para regiões temperadas. Estudos como o de (BUSSON *et al.*, 2018) demonstram que modelos clássicos exibem desvios sistemáticos quando aplicados a essas condições, exigindo abordagens mais adaptativas.

A região do Ceará caracteriza-se por elevados índices de  $G$  e regimes de vento com forte sazonalidade, conforme ilustrado na Figura 1. A combinação de alta incidência solar e  $T_a$  pode levar  $T_{mod}$  a patamares superiores a  $70\text{ }^{\circ}\text{C}$ , resultando em perdas de eficiência significativas, que podem exceder 20% em picos de verão (BUSSON *et al.*, 2021; NASCIMENTO *et al.*, 2020).

Figura 1 – Sazonalidade dos recursos climáticos no Ceará



Fonte: Elaborado pelo autor com dados de ATLAS SOLAR E EÓLICO (2019).

Observa-se na Figura 1 a correlação sazonal onde o período de maior disponibilidade solar (segundo semestre) coincide com a intensificação dos ventos alísios. O segundo semestre também apresenta os maiores índices de  $G$ , coincidindo com a intensificação de  $V_w$ . Esse acoplamento é crucial para o resfriamento convectivo natural, introduzindo não-linearidades que modelos lineares simples muitas vezes falham em capturar, especialmente em eventos de vento nulo que eliminam a convecção forçada (MEDEIROS *et al.*, 2021).

Além disso, fatores como a baixa umidade relativa e a deposição severa de sujeira (*soiling*) alteram as propriedades térmicas superficiais dos módulos, criando pontos quentes localizados (COSTA *et al.*, 2021; SOUZA *et al.*, 2025). Pesquisas recentes indicam que a incorporação de variáveis como direção do vento e precipitação é crítica para a precisão preditiva em climas áridos, reduzindo erros durante transições sazonais (FILHO *et al.*, 2025; BARROS, 2022).

Nesse contexto, a validação experimental presente na literatura aponta que modelos baseados em dados e algoritmos *ensemble* (como a RF e o XGBoost) oferecem maior robustez para lidar com essas complexidades climáticas (SANTOS *et al.*, 2021; SANTOS *et al.*, 2024). Essa fundamentação teórica justifica a abordagem metodológica adotada neste trabalho, cujo foco central é avaliar e comparar o desempenho preditivo desses algoritmos na estimativa de  $T_{mod}$  sob as condições ambientais severas do semiárido local, buscando superar as limitações das correlações físicas estacionárias.

## 2.7 Lacunas na literatura e justificativa do estudo

Apesar dos avanços na modelagem de temperatura de módulos FV, persistem lacunas importantes:

1. **Comparações limitadas:** Muitos estudos comparam apenas um modelo físico/empírico de referência com um algoritmo de ML, sem explorar múltiplas correlações semi-empíricas.
2. **Ausência de validação temporal rigorosa:** Divisões aleatórias de dados podem levar ao vazamento temporal (*temporal leakage*), superestimando o desempenho dos modelos.
3. **Falta de quantificação de incerteza:** Modelos ML raramente fornecem intervalos de predição, limitando sua aplicabilidade operacional.
4. **Interpretabilidade limitada:** Poucos estudos verificam a coerência física entre importâncias de variáveis em ML e coeficientes de modelos lineares.
5. **Generalização climática:** Correlações calibradas em ambientes temperados podem apre-

sentar perda de acurácia em regiões semiáridas.

Este trabalho busca preencher essas lacunas por meio de uma abordagem comparativa entre modelos físico-semiempíricos e algoritmos de ML, com foco específico nas condições do clima semiárido, englobando a avaliação do desempenho preditivo e a interpretabilidade físico-estatística dos resultados.

### 3 METODOLOGIA

Este capítulo apresenta a metodologia empregada para a implementação e avaliação de modelos de ML aplicados à estimativa de  $T_{mod}$ . A abordagem metodológica foi estruturada em duas etapas principais: o tratamento e preparação dos dados experimentais coletados no LEA-UFC, e a implementação e comparação de diferentes algoritmos de ML.

A metodologia adotada inicia com a seleção e filtragem dos dados baseada em critérios físicos e meteorológicos, seguida pela aplicação de técnicas de pré-processamento para garantir a qualidade e consistência do conjunto de dados. Na sequência, são configurados e ajustados quatro algoritmos distintos: RL, KNN, RF e XGBoost, cada um submetido a estratégias específicas de calibração de hiperparâmetros.

Para fins de comparação com as abordagens de ML, três modelos físicos clássicos foram selecionados como referência (*benchmarks*), representando diferentes níveis de complexidade na modelagem da temperatura de módulos FV.

O processo de avaliação contempla métricas estatísticas e análises gráficas complementares, permitindo uma comparação entre os modelos e a identificação da abordagem mais adequada para a estimativa de  $T_{mod}$  nas condições climáticas de Fortaleza, Ceará. A metodologia proposta visa não apenas alcançar precisão preditiva, mas também fornecer insights sobre a importância relativa das variáveis meteorológicas na determinação da temperatura de operação dos módulos FV.

O presente capítulo está organizado da seguinte forma: a Seção 3.1 descreve a caracterização do sistema experimental utilizado, incluindo localização, equipamentos e instrumentação; a Seção 3.2 apresenta os procedimentos de tratamento dos dados, abrangendo filtragem, detecção de anomalias e normalização; a Seção 3.3 discute a seleção das variáveis preditivas utilizadas nos modelos; a Seção 3.5 aborda a modelagem por ML e as justificativas para a escolha dos algoritmos; a Seção 3.7 detalha as estratégias de validação e otimização empregadas, incluindo divisão temporal dos dados e ajuste de hiperparâmetros; a Seção 3.4 define as métricas de avaliação e interpretabilidade utilizadas para comparação dos modelos; e, finalmente, a Seção 3.6 apresenta a abordagem de modelagem híbrida, incluindo otimização de modelos físicos, calibração de correlações semi-empíricas e validação comparativa das correlações híbridas.

### 3.1 Caracterização do sistema experimental

Os dados analisados neste trabalho foram coletados na plataforma experimental do LEA-UFC, situado em Fortaleza (Latitude  $03^{\circ}43'$  S, Longitude  $38^{\circ}32'$  O, elevação de 21 m). A localização, próxima à Linha do Equador, confere à região uma dinâmica climática distinta, com um primeiro semestre tipicamente chuvoso e um segundo semestre caracterizado por estiagem, ventos intensos e elevados índices de irradiância solar.

A planta FV analisada do LEA-UFC é constituída por 12 módulos de silício policristalino (p-Si) com potência nominal de 330 Wp cada. Os módulos estão instalados em uma estrutura fixa do tipo *open rack*, orientada para o Norte geográfico com uma inclinação de  $10^{\circ}$  em relação à horizontal. A disposição física do arranjo é ilustrada na Figura 2, que apresenta: (a) a vista aérea da localização do laboratório; (b) o arranjo dos 12 módulos; e (c) o detalhe do posicionamento dos sensores de temperatura no dorso dos módulos. Conforme detalhado por (SOUZA *et al.*, 2025), o monitoramento térmico é realizado por sensores posicionados na face posterior (*backsheet*) de cada módulo, com um sensor central instalado em todos eles. Adicionalmente, os módulos situados nas extremidades do arranjo são equipados com sensores suplementares localizados nas bordas superior e inferior, permitindo capturar de forma mais precisa o gradiente térmico da planta.

O monitoramento do sistema completo e a aquisição das variáveis meteorológicas são realizados por uma arquitetura escalável projetada para acompanhamento em tempo real. Esta arquitetura utiliza microcontroladores ESP32 equipados com módulos de comunicação *Long-Range* (LoRa) para a coleta e transmissão das informações, com armazenamento integrado na plataforma IoT ThingSpeak (ASSIS *et al.*, 2025). Este sistema registra  $G$  (tanto no plano horizontal quanto no plano do arranjo), medida por um piranômetro *Hukseflux* LP02;  $T_a$ , coletada por um sensor termorresistivo PT100 abrigado em escudo de radiação; e  $V_w$ , aferida por um anemômetro NRG #40C ao nível dos painéis. Para garantir a fidelidade dos dados do perfil térmico, o monitoramento específico de  $T_{mod}$  é gerenciado por um *datalogger* CR1000 (ARARIPE *et al.*, 2024), que adquire as leituras de sensores digitais de temperatura DS18B20 fixados na região posterior das placas.

### 3.2 Tratamento dos dados

Para o tratamento dos dados, foram aplicadas as seguintes etapas sequenciais:

Figura 2 – Instalação experimental no LEA-UFC  
(a)



(b)



(c)

**Sensor de Temperatura**

- Pasta Térmica
- Silicone Adesivo

ST2		ST1	
6ST2S	→	6ST1S	
5ST2C	→	5ST1C	
4ST2C	→	4ST1C	
3ST2C	→	3ST1C	
2ST2C	→	2ST1C	
1ST2C	→	1ST1S	
1ST2S	→	1ST1S	

**Sensor de Temperatura**  
Inseridos Entre o vidro do painel e a estrutura inferior

**Anemômetro**

**Sensor DS18B20 Inserido**

**Sensor Isolado Termicamente**

**Sensor de Radiação**

Fonte: Elaborado pelo autor.

1. **Importação e organização:** Os dados foram importados de um arquivo CSV contendo registros meteorológicos e elétricos de um sistema FV instalado no LEA-UFC.
2. **Cálculo da posição solar:** Utilizou-se a função `get_solarposition` da biblioteca `PVlib` para calcular os ângulos zenital e azimutal a partir da latitude, longitude e altitude do local.
3. **Filtro de dados diurnos:** Apenas as medições realizadas quando o Sol estava acima do horizonte (ângulo zenital  $< 90^\circ$ ) foram consideradas, assegurando a presença efetiva de irradiação solar.
4. **Remoção de outliers:** Foram definidos limites físicos realistas para as variáveis meteorológicas, com base em dados históricos da região de Fortaleza (INMET, 2023):
  - Irradiância:  $10 \text{ W/m}^2 \leq G \leq 1250 \text{ W/m}^2$ ;
  - Temperatura ambiente:  $21,7^\circ\text{C} \leq T_a \leq 39,0^\circ\text{C}$ ;
  - Velocidade do vento:  $0 \text{ m/s} \leq V_w \leq 15 \text{ m/s}$ .
5. **Cálculo da variável resposta:**  $\bar{T}_{mod}$  foi determinada com base nas médias das medições de 12 sensores instalados nos módulos FV da planta.
6. **Seleção de variáveis:** Foram selecionados cinco preditores:  $G$ ,  $T_a$ ,  $V_w$ , ângulo zenital (*zenith*) e ângulo azimutal (*azimuth*).
7. **Divisão temporal treino-teste:** Visando avaliar a capacidade de generalização temporal dos modelos, os dados foram divididos em 50% para treinamento e 50% para teste (136.065 amostras). Esta abordagem é crucial para séries temporais, evitando vazamento de informação futura (*data leakage*) e simulando a aplicação prática dos modelos.
8. **Normalização condicional:** Para os modelos sensíveis à escala (KNN e RL), foi aplicado `StandardScaler` através de pipelines do `Scikit-learn`. O ajuste (*fit*) foi realizado exclusivamente no conjunto de treinamento, com posterior transformação de ambos os conjuntos. Modelos baseados em árvores (RF e XGBoost) não requerem normalização.

O conjunto de dados final compreende o período de 22 de maio de 2023 a 30 de maio de 2025, com registros em intervalos de 1 minuto. Após a filtragem, foram mantidas 273.310 observações válidas das 946.618 originais, representando uma taxa de aproveitamento de 28,9%. Em relação aos dados diurnos, a taxa de aproveitamento foi de 99,8%.

### 3.3 Seleção das variáveis preditivas

A seleção das variáveis preditivas utilizadas nos modelos de ML foi orientada tanto pelos fundamentos teóricos de transferência de calor quanto por evidências empíricas da literatura

especializada. O conjunto de entradas definido para o treinamento dos algoritmos compreende cinco variáveis principais:  $G$ ,  $T_a$ ,  $V_w$ ,  $zenith$  e  $azimuth$ .

A inclusão dos parâmetros geométricos ( $zenith$  e  $azimuth$ ) justifica-se pela correlação temporal que essas métricas apresentam com a dinâmica de aquecimento diário do arranjo. Embora não sejam fontes termodinâmicas diretas de calor, as coordenadas solares atuam como marcadores temporais e determinam o ângulo de incidência da radiação. Essa configuração afeta a fração de energia efetivamente absorvida pelo painel e a dinâmica convectiva ao redor das estruturas (DUFFIE; BECKMAN, 2013; KLEISSL; WEST, 2015). Trabalhos recentes, como o de (CAMPOS, 2020), também utilizaram parâmetros solares como variáveis de entrada em modelos de estimativa de  $T_{mod}$ , evidenciando ganhos em acurácia preditiva.

### 3.4 Métricas de avaliação e interpretabilidade

Para avaliar o desempenho dos modelos, foram implementadas as métricas descritas a seguir. Nas equações,  $n$  representa o número total de amostras,  $T_{mod,i}$  refere-se à temperatura medida (alvo),  $\hat{T}_{mod,i}$  à temperatura estimada pelo modelo, e  $\bar{T}_{mod}$  à média aritmética de todas as observações medidas.

**Mean Absolute Error (Erro Absoluto Médio – MAE)** - Quantifica a magnitude média dos erros sem considerar sua direção (se é para mais ou para menos):

$$MAE = \frac{1}{n} \sum_{i=1}^n |T_{mod,i} - \hat{T}_{mod,i}| \quad (3.1)$$

**Root Mean Square Error (Raiz do Erro Quadrático Médio – RMSE)** - Penaliza erros maiores de forma mais severa do que o MAE, devido à elevação das diferenças ao quadrado antes da média:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (T_{mod,i} - \hat{T}_{mod,i})^2} \quad (3.2)$$

#### RMSE normalizado (nRMSE)

Permite a comparação do erro entre diferentes estudos independentemente da escala dos dados, sendo normalizado pela média das observações.

$$nRMSE = \frac{RMSE}{\bar{T}_{mod}} \times 100\% \quad (3.3)$$

**Mean Bias Error (Erro de Viés Médio – MBE)** - Indica se o modelo possui uma tendência

sistemática de superestimar ( $MBE > 0$ ) ou subestimar ( $MBE < 0$ ) os valores reais.

$$MBE = \frac{1}{n} \sum_{i=1}^n (\hat{T}_{mod,i} - T_{mod,i}) \quad (3.4)$$

**Coefficiente de Determinação ( $R^2$ )** - Indica a proporção da variância dos dados reais que é explicada pelo modelo, variando de  $-\infty$  a 1 (onde 1 é o ajuste perfeito).

$$R^2 = 1 - \frac{\sum_{i=1}^n (T_{mod,i} - \hat{T}_{mod,i})^2}{\sum_{i=1}^n (T_{mod,i} - \bar{T}_{mod})^2} \quad (3.5)$$

**Percentil 95 do erro absoluto ( $P_{95}$ )** - Representa o valor abaixo do qual se concentram 95% dos erros absolutos cometidos pelo modelo. É uma métrica estatística robusta para avaliar o limite de confiança das estimativas e a presença de erros extremos (*outliers*) nas predições:

$$P_{95} = \text{percentil}(|T_{mod,i} - \hat{T}_{mod,i}|, 95) \quad (3.6)$$

Além das métricas quantitativas de erro, a interpretabilidade dos modelos de ML foi avaliada para compreender o peso físico de cada variável meteorológica nas predições. Para os algoritmos baseados em árvores de decisão, utilizou-se a métrica de *Feature Importance*.

Essa técnica quantifica a contribuição relativa de cada variável de entrada para a redução da incerteza do modelo durante o treinamento. A importância relativa permite não apenas atestar a precisão preditiva da modelagem, mas também validar fisicamente o comportamento do algoritmo frente à dinâmica térmica real dos módulos FV no microclima estudado.

### 3.5 Modelagem por ML e justificativas

A seleção dos modelos de ML seguiu critérios metodológicos estabelecidos na literatura para garantir representatividade de diferentes paradigmas de aprendizado e permitir análise comparativa abrangente (JAMES *et al.*, 2013; HASTIE *et al.*, 2009; MELLIT *et al.*, 2021). Os quatro modelos foram escolhidos com base nos seguintes critérios:

- **RL:** Escolhida por sua simplicidade, facilidade de interpretação e por servir como modelo *baseline* (*white-box*). Apesar de suas limitações em capturar relações não-lineares, é importante como referência comparativa inicial (RENKEMA, 2024).
- **KNN:** Modelo não paramétrico e relativamente interpretável, baseado na similaridade entre amostras no espaço das variáveis preditoras (HASTIE *et al.*, 2009). Tem sido aplicado com sucesso em estimativas de  $T_{mod}$  em estudos como Campos (2020).

- **RF:** Um modelo do tipo *ensemble* baseado em múltiplas árvores de decisão. Embora considerado um modelo de *caixa-preta* (*black-box*), permite análise de importância das variáveis, o que aumenta sua interpretabilidade. Apresenta boa capacidade de generalização e é robusto a *overfitting* (BREIMAN, 2001).
- **XGBoost:** Outro modelo *ensemble*, baseado em *boosting* de árvores, conhecido por sua alta capacidade preditiva em problemas com relações não-lineares complexas. Embora também classificado como *black-box*, permite análise de *feature importance*. Sua escolha foi motivada pelo bom desempenho relatado em séries temporais e problemas ambientais (CHEN; GUESTRIN, 2016).

A seleção alinha-se com trabalhos consolidados na área, onde o RF é amplamente utilizado em modelagem de sistemas de energia renovável devido ao equilíbrio entre performance e interpretabilidade (VOYANT *et al.*, 2017; NESPOLI *et al.*, 2019), enquanto o XGBoost tem crescente adoção em aplicações FV por sua alta acurácia em problemas de regressão com dados meteorológicos (AHMED *et al.*, 2020; LI *et al.*, 2020). O KNN constitui método de referência para captura de padrões locais em séries temporais meteorológicas (ANTONANZAS *et al.*, 2016), e RL representa baseline obrigatório para quantificar ganhos de métodos mais complexos (SKOPLAKI; PALYVOS, 2009). A diversidade entre modelos lineares, baseados em vizinhança e baseados em árvores foi proposital, visando explorar diferentes formas de relacionamento entre as variáveis.

### 3.6 Modelagem híbrida

Para complementar a análise comparativa entre diferentes abordagens de modelagem, os modelos físicos clássicos (*Ross, Faiman e PVsyst*) foram recalibrados utilizando o conjunto de treinamento para adaptação às condições climáticas específicas do semiárido de Fortaleza. Adicionalmente, correlações baseadas nas três formas gerais apresentadas na fundamentação teórica (2.5, 2.6 e 2.7) foram implementadas e otimizadas. Esta seção descreve os procedimentos de calibração, restrições aplicadas e critérios de validação para os modelos híbridos.

#### 3.6.1 Otimização de modelos físicos

A recalibração dos modelos físicos foi realizada através de técnicas de otimização numérica aplicadas ao conjunto de treinamento, com o objetivo de minimizar o RMSE entre

as temperaturas estimadas e medidas. Para o modelo de *Ross* (Equação 2.8), empregou-se o método de mínimos quadrados restringidos (*constrained least squares*) implementado através da biblioteca `scipy.optimize`, com o parâmetro  $r$  restrito ao intervalo  $[0.001, 0.1] \frac{^\circ\text{C}\cdot\text{m}^2}{\text{W}}$ . Esta restrição garante que o coeficiente de resistência térmica permaneça dentro de limites fisicamente realistas para módulos de silício cristalino em condições de instalação tipo *open rack*.

Para os modelos de *Faiman* e *PVsyst* (Equações 2.9 e 2.10), que envolvem múltiplos parâmetros não lineares, foi aplicado o algoritmo de evolução diferencial (*differential evolution*). Este método de otimização global baseado em população é particularmente adequado para problemas com múltiplos ótimos locais e superfícies de resposta complexas. Os intervalos de busca foram estabelecidos com base em valores reportados na literatura e considerações físicas sobre mecanismos de transferência de calor:

- Modelo de *Faiman*:  $U_0 \in [1.0, 100.0] \text{ W}/(\text{m}^2 \cdot \text{K})$  e  $U_1 \in [0.1, 50.0] \text{ W} \cdot \text{s}/(\text{m}^3 \cdot \text{K})$ ;
- Modelo *PVsyst*: Conforme apontado na literatura de modelagem preditiva, para fins de otimização de parâmetros puramente baseada em dados meteorológicos, adotou-se a formulação simplificada do balanço de energia, estruturalmente análoga ao modelo de *Faiman*. Nesta abordagem, os termos referentes à absorvância e eficiência do módulo ( $\alpha(1 - \eta)$ ) são suprimidos do numerador, sendo seus efeitos físicos absorvidos diretamente na calibração dos coeficientes de perda térmica. Os limites de busca estabelecidos foram:  $u_c \in [10, 0, 50, 0] \text{ W}/(\text{m}^2 \cdot \text{K})$  e  $u_v \in [0, 0, 20, 0] \text{ W} \cdot \text{s}/(\text{m}^3 \cdot \text{K})$ .

Os limites inferiores de  $U_0$  e  $u_c$  foram estabelecidos considerando perdas mínimas por convecção natural e radiação, enquanto os limites superiores evitam coeficientes irrealisticamente elevados que resultariam em respostas térmicas instantâneas. Para os coeficientes dependentes do vento ( $U_1$  e  $u_v$ ), os limites refletem a gama de intensidades de resfriamento convectivo observadas em diferentes configurações de montagem.

A função objetivo para todos os modelos foi definida como:

$$\text{RMSE}_{\text{treino}} = \sqrt{\frac{1}{n_{\text{treino}}} \sum_{i=1}^{n_{\text{treino}}} (T_{\text{mod},i} - \hat{T}_{\text{mod},i}(\theta))^2} \quad (3.7)$$

onde  $\theta$  representa o vetor de parâmetros a ser otimizado,  $T_{\text{mod},i}$  são as temperaturas medidas,  $\hat{T}_{\text{mod},i}$  são as estimativas do modelo, e  $n_{\text{treino}}$  é o número de amostras no conjunto de treinamento. Os critérios de convergência incluíram tolerância de  $10^{-6}$  para a variação da função objetivo entre iterações consecutivas e um máximo de 1.000 iterações. Para o algoritmo de evolução

diferencial, o tamanho da população foi fixado em 15 vezes o número de parâmetros, garantindo exploração adequada do espaço de busca.

### 3.6.2 Calibração de correlações semi-empíricas

As três formas gerais de correlações semi-empíricas (Linear, Não-Linear e Racional) foram calibradas utilizando procedimentos de otimização específicos adequados à estrutura matemática de cada modelo. Para a forma linear geral (Equação 2.5), aplicou-se o método de mínimos quadrados ordinários com pré-processamento dos dados.  $G$  foi normalizado dividindo-se por  $1.000 \text{ W/m}^2$  ( $G' = G/1000$ ) para melhorar o condicionamento numérico da matriz de design e reduzir discrepâncias de escala entre variáveis preditoras. Os coeficientes  $a_0, a_1, a_2, a_3$  foram determinados através da solução do sistema de equações normais.

A forma não-linear geral (Equação 2.6) utiliza uma estrutura multi-termo com 11 parâmetros ( $b_0, b_1, C_1, d_1, d_2, e_1, C_2, e_2, f_1, f_2, f_3$ ), combinando efeitos lineares e interações entre variáveis meteorológicas. Empregou-se o algoritmo *Trust Region Reflective* (padrão do `scipy.optimize.curve_fit` para problemas com restrições) para minimizar o erro quadrático médio, com irradiância normalizada ( $G' = G/1000$ ) para estabilidade numérica.

- **Restrição de parâmetros:** Aplicação de limites rígidos (*bounds*) aos coeficientes ( $b_0, \dots, f_3$ ) para garantir consistência física e evitar valores irrealistas;
- **Simplificação estrutural:** Linearização dos expoentes ( $g_i = 1$ ) e remoção de termos exponenciais complexos para reduzir a dimensionalidade e aumentar a estabilidade numérica;
- **Controle de *overfitting*:** O uso de uma estrutura simplificada com limites restritos dispensou a necessidade de regularização L2 (Tikhonov), priorizando a convergência do ajuste numérico.

A forma racional geral (Equação 2.7) foi calibrada utilizando ajuste de curva não linear com atenção especial à prevenção de singularidades no denominador. Os limites estabelecidos foram:  $h_1 \in [0, 1; 3, 0]$  (multiplicador da temperatura ambiente),  $h_2 \in [0, 001; 0, 1]$  ( $^\circ\text{C}\cdot\text{m}^2/\text{W}$  (efeito da irradiância)),  $m_1 \in [0, 1; 10, 0]$  (termo constante do denominador) e  $n_1 \in [-500; 500]$  (termo dependente da irradiância no denominador). O limite inferior de  $m_1$  garante que o denominador nunca se anule, enquanto o intervalo de  $n_1$  permite tanto comportamento linear ( $n_1 \approx 0$ ) quanto assintótico ( $n_1 > 0$ ). Todas as calibrações foram realizadas no conjunto de treinamento.

### 3.6.3 Validação e comparação dos modelos

Após a fase de calibração, todas as abordagens preditivas (modelos físicos, algoritmos de ML e híbridos) foram avaliadas no conjunto de teste não visto (136.065 amostras). O desempenho foi quantificado de forma unificada utilizando as métricas estatísticas de erro previamente definidas na Seção 3.4. Esta padronização metodológica permite uma comparação direta e rigorosa entre as diferentes topologias. Os parâmetros e hiperparâmetros calibrados na fase de treinamento foram estritamente fixados, e nenhum ajuste adicional foi realizado durante a avaliação no conjunto de teste, garantindo assim uma estimativa não enviesada da capacidade de generalização dos modelos.

Para quantificar o ganho de desempenho proporcionado pela otimização computacional, os modelos físicos foram avaliados em duas configurações distintas: (i) com os parâmetros originais reportados na literatura ou derivados de condições padronizadas de ensaio; e (ii) com os parâmetros recalibrados localmente. A comparação entre estas configurações atesta a adaptabilidade e a real necessidade de calibração das equações empíricas clássicas quando submetidas à dinâmica climática específica do semiárido local.

### 3.7 Estratégias de validação e otimização

Para garantir a robustez e a capacidade de generalização dos modelos, foi aplicada uma estratégia de validação cruzada do tipo *K-Fold*, com cinco subdivisões ( $K = 5$ ), utilizando embaralhamento dos dados e uma semente fixa (`random_state=42`). Essa abordagem garante que os resultados obtidos sejam reprodutíveis e não dependam de uma única divisão dos dados (HASTIE *et al.*, 2009).

A otimização dos hiperparâmetros foi realizada por meio de duas abordagens principais:

- **GridSearchCV:** Utilizada para os modelos RF e KNN, explorando uma grade de combinações de hiperparâmetros, conforme especificado na Tabela 3.
- **RandomizedSearchCV:** Aplicada ao modelo XGBoost, considerando o grande número de hiperparâmetros possíveis e buscando reduzir o custo computacional.

Para a RL, não foram necessários processos de otimização, pois o modelo não depende de hiperparâmetros ajustáveis. A Tabela 2 resume as características e interpretabilidade dos modelos aplicados e a Tabela 3 as configurações de normalização e hiperparâmetros utilizados

nos modelos.

Tabela 2 – Características gerais dos modelos aplicados

Modelo	Interpretabilidade	Descrição resumida
<i>Ross</i>	Alta	Relação linear $T_{mod} - T_a$ vs $G$ (um parâmetro).
<i>Faiman</i>	Alta	Inclusão de termo convectivo dependente de $V_w$ .
<i>PVsyst</i>	Alta	Balço térmico em regime estacionário com correção de vento.
RL	Alta	Modelo estatístico multivariado com interpretação direta dos coeficientes.
KNN	Alta	Predição transparente baseada na similaridade com amostras históricas rastreáveis.
RF	Baixa	Conjunto ( <i>bagging</i> ) de múltiplas árvores, o que dilui a rastreabilidade individual.
XGBoost	Baixa	Múltiplas árvores sequenciais otimizadas por gradiente, configurando arquitetura caixa-preta.

Fonte: Adaptado de Linardatos *et al.* (2021).

Observação: Interpretabilidade relativa à transparência intrínseca dos mecanismos internos do modelo e à rastreabilidade de suas predições individuais.

Tabela 3 – Configurações e hiperparâmetros explorados

Modelo	Normalização	Busca	CV	Hiperparâmetros (intervalos / conjunto)
<i>Ross</i>	Não	–	–	$r$ .
<i>Faiman</i>	Não	–	–	$U_0, U_1$ via ajuste não linear.
<i>PVsyst</i>	Não	–	–	$u_c, u_v$ via algoritmo de evolução diferencial.
RL	Sim	–	–	Sem hiperparâmetros ajustáveis.
KNN	Sim	Grid	5-fold	$k \in \{3, 5, 7, 9, 11, 15\}$ ; métricas $\in \{\text{euclidean, manhattan}\}$ ; pesos $\in \{\text{uniform, distance}\}$ .
RF	Não	Grid	5-fold	$n\_estimators \in [100, 200, 300]$ ; $\text{max\_depth} \in \{\text{None}, 10, 20, 30\}$ ; $\text{min\_samples\_split} \in \{2, 5, 10\}$ ; $\text{min\_samples\_leaf} \in \{1, 2, 4\}$ ; $\text{max\_features} \in \{\text{auto, sqrt}\}$ .
XGBoost	Não	Random	5-fold	$n\_estimators \in [200, 400, 600, 800]$ ; $\text{learning\_rate} \in [0.01, 0.05, 0.1, 0.2]$ ; $\text{max\_depth} \in [3, 4, 5, 6, 8]$ ; $\text{subsampling} \in [0.6, 0.8, 1.0]$ ; $\text{colsample\_bytree} \in [0.6, 0.8, 1.0]$ ; $\text{reg\_lambda} \in [0, 1, 5, 10]$ ; $\text{reg\_alpha} \in [0, 0.1, 0.5]$ .

Fonte: Elaborado pelo autor.

Nota: Normalização aplicada via Pipeline apenas para modelos sensíveis à escala (RL e KNN), evitando vazamento de dados (*data leakage*).

As etapas metodológicas descritas buscaram estabelecer uma comparação a mais justa possível entre as abordagens de naturezas distintas, reconhecendo as assimetrias inerentes aos conjuntos de variáveis de entrada exigidos por cada equação ou algoritmo. Os valores numéricos finais, como os hiperparâmetros, as métricas de desempenho e os parâmetros físicos

calibrados, são consolidados na subseção de resultados comparativos (*vide* Tabela 8). Pequenas variações nas métricas de erro (como  $RMSE \pm 0,02 \text{ } ^\circ\text{C}$  e  $MAE \pm 0,01 \text{ } ^\circ\text{C}$ ) podem ocorrer em função de diferenças na semente aleatória de inicialização dos algoritmos ou na partição temporal dos dados.

Esta fundamentação permite, na próxima seção, apresentar e discutir quantitativamente o desempenho comparativo dos modelos e a análise interpretativa.

## 4 RESULTADOS

Este capítulo apresenta e discute os resultados obtidos no presente estudo. A organização segue a seguinte forma: a Seção 4.1 apresenta a análise dos dados coletados, incluindo caracterização estatística e comportamento temporal das variáveis meteorológicas e térmicas monitoradas; a Seção 4.2 avalia o desempenho dos modelos de ML implementados para estimativa de  $T_{mod}$ ; a Seção 4.3 discute os resultados dos modelos físicos de referência utilizados como baseline; a Seção 4.4 apresenta os resultados da modelagem híbrida, combinando formulações físicas com ajuste baseado em dados; e, finalmente, a Seção 4.5 realiza uma análise comparativa de desempenho entre todas as abordagens, identificando fatores determinantes do desempenho relativo dos modelos e implicações para aplicações práticas em clima semiárido.

### 4.1 Análise dos dados coletados

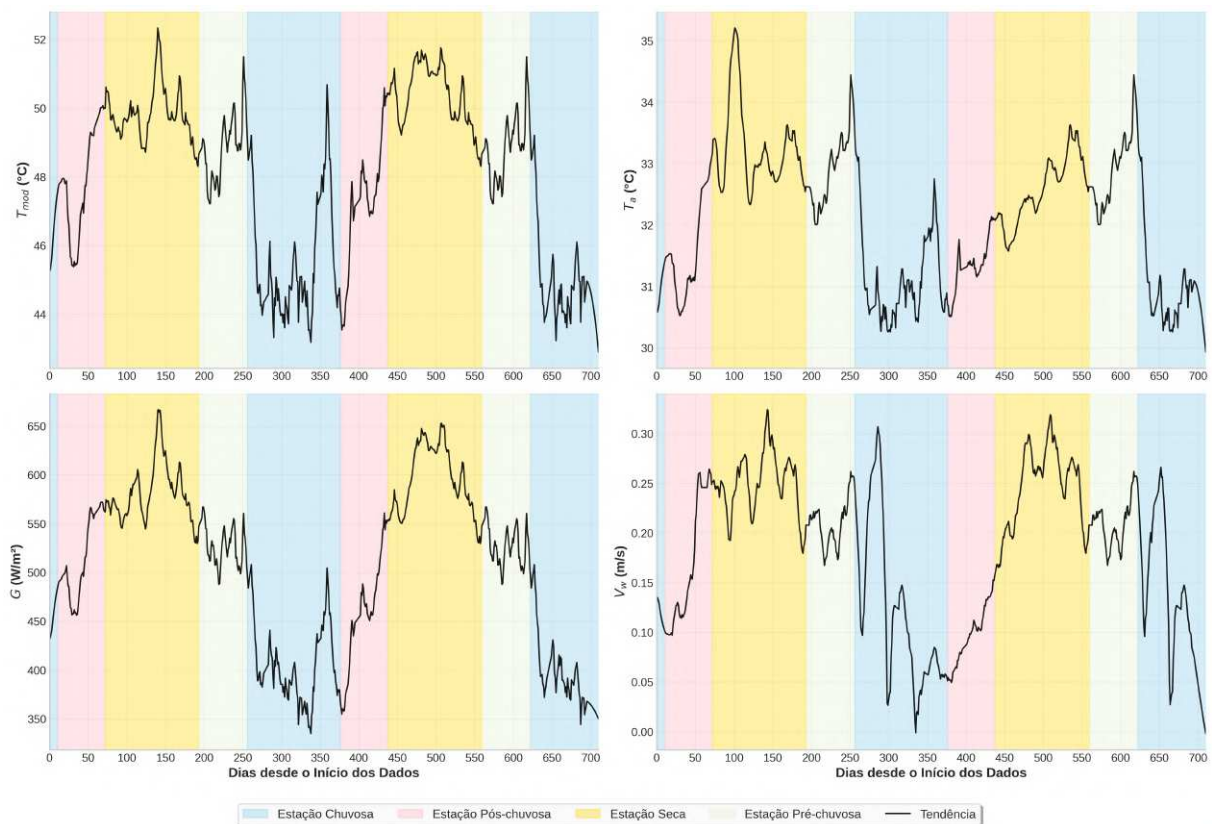
Os resultados referentes à caracterização dos dados de  $G$ ,  $T_a$ ,  $T_{mod}$ , *azimuth*, *zenith* e  $V_w$  obtidos no LEA-UFC ao longo de dois anos de monitoramento contínuo (22 de maio de 2023 a 30 de maio de 2025) estão expostos a seguir. Analisando os dados ao longo do período completo, observa-se variação sazonal pronunciada nas medições, com padrões distintos entre as estações chuvosa e seca, refletindo as características climáticas típicas do semiárido cearense.

Fortaleza está localizada em região de clima BSh<sup>1</sup>, caracterizado por sazonalidade pluviométrica bem definida (MAGALHÃES; ZANELLA, 2011; SOUZA; CARVALHO, 2025). O ciclo hidrológico anual pode ser subdividido em quatro períodos distintos (SOUZA; CARVALHO, 2025): (i) estação chuvosa, concentrada no primeiro quadrimestre (fevereiro a maio), quando ocorrem 80–90% da precipitação anual (600–800 mm), resultando em aumento de nebulosidade e redução de 20–30% na irradiância global incidente; (ii) estação pós-chuvosa (junho a julho), caracterizada por transição gradual com redução progressiva da precipitação e aumento da disponibilidade solar; (iii) estação seca (agosto a janeiro), com precipitação praticamente nula (0–50 mm/mês), céu predominantemente limpo (>85% de dias sem nuvens) e máximos de  $G$ , com médias diárias superiores a 900 W/m<sup>2</sup>; e (iv) estação pré-chuvosa (final de janeiro e fevereiro), marcada por aumento gradual da umidade atmosférica e retorno da atividade convectiva. Esta forte sazonalidade impacta diretamente o comportamento térmico dos módulos FV, conforme evidenciado nas análises subsequentes (ALFARO *et al.*, 2023; FUNCEME, 2019).

<sup>1</sup> Segundo a classificação climática de Köppen–Geiger, o tipo BSh corresponde ao clima semiárido quente: B (climas áridos), BS (estepe) e h (temperatura média anual  $\geq 18^\circ\text{C}$ ) (PEEL *et al.*, 2007).

A Figura 3 apresenta a evolução temporal das quatro variáveis principais ( $G$ ,  $T_{mod}$ ,  $T_a$  e  $V_w$ ) ao longo dos 730 dias de medição, com marcação visual das quatro estações sazonais: chuvosa (fundo azul claro, dias 1–10, 256–376 e 622–730), pós-chuvosa (fundo rosa claro, dias 11–71 e 377–437), seca (fundo amarelo, dias 72–193 e 438–559) e pré-chuvosa (fundo cinza claro, dias 194–255 e 560–621). Para evitar a sobrecarga visual decorrente da alta densidade de dados intradiários e destacar o comportamento macroscópico de longo prazo, os registros brutos foram previamente agregados em médias diárias. Sobre estes valores agregados, aplicou-se o filtro de suavização *Savitzky-Golay*<sup>2</sup> (utilizando uma janela móvel de 31 dias e polinômio de segunda ordem), resultando nas tendências sazonais contínuas representadas pelas linhas pretas. Observa-se que  $G$  e  $T_{mod}$  apresentam forte acoplamento temporal, com médias suavizadas de 650 W/m<sup>2</sup> e 50,0°C durante a estação seca, reduzindo para 400 W/m<sup>2</sup> e 45,0°C na estação chuvosa, respectivamente.  $T_a$  mantém-se relativamente estável (30,0–34,0°C) ao longo do ano, enquanto  $V_w$  exibe baixa variabilidade sazonal devido à configuração urbana do local de medição.

Figura 3 – Variação sazonal das variáveis meteorológicas e  $T_{mod}$  ao longo do período de medição (22/mar/2023–30/mar/2025) no clima semiárido de Fortaleza/CE.



Fonte: Elaborado pelo autor.

<sup>2</sup> Filtro digital utilizado para suavização de dados temporais que minimiza o ruído mantendo a forma e a amplitude dos picos e vales do sinal original.

A Figura 4 apresenta uma matriz de correlação entre as diferentes variáveis analisadas. Na diagonal principal, encontram-se histogramas de frequência que mostram a distribuição dos dados para cada variável. Os gráficos de dispersão, localizados abaixo da diagonal principal, exibem a relação entre pares de variáveis, com uma linha de suavização em vermelho que auxilia na visualização das tendências. Na parte superior da diagonal, estão dispostos os coeficientes de correlação, representados em azul claro, cujo tamanho indica a intensidade da correlação.

A análise dos coeficientes de correlação apresentada na Figura 4 revela que  $T_{mod}$  apresenta duas correlações significativas com diferentes variáveis. O preditor  $G$  mostra uma correlação mais pronunciada, com coeficiente de 0,84, indicando forte relação positiva entre irradiância incidente e aquecimento do módulo. Similarmente,  $T_a$  correlaciona-se fortemente com  $T_{mod}$ , apresentando coeficiente de 0,73, refletindo a influência da temperatura ambiente no balanço térmico do sistema FV.

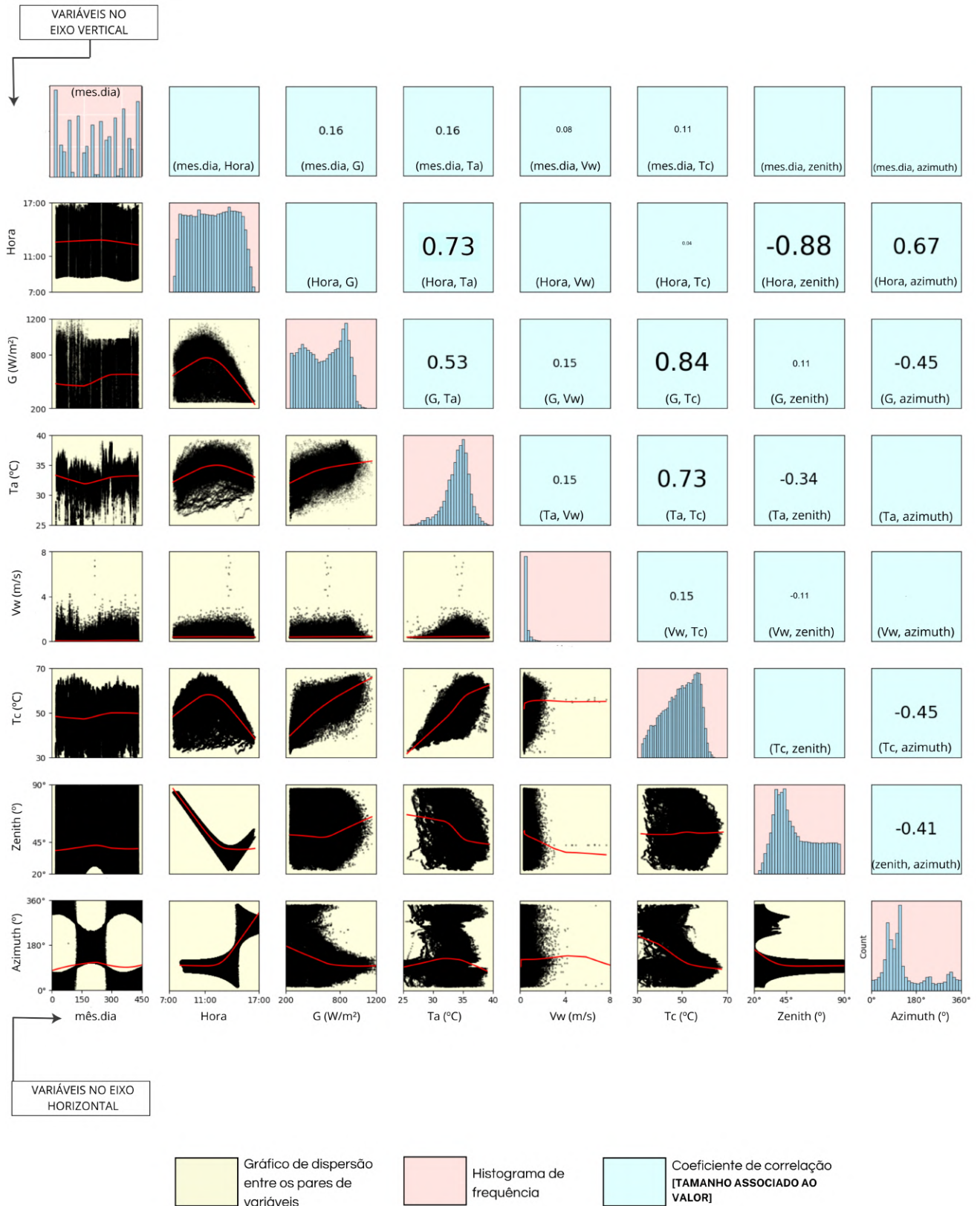
As variáveis secundárias demonstram influências menos expressivas.  $V_w$  apresenta correlação relativamente baixa de 0,15, sugerindo impacto limitado na temperatura do módulo, comportamento esperado em área urbanas com predominância de baixas velocidades de vento (média observada de 1,2 m/s). Os ângulos solares revelam padrões relevantes: o ângulo zenital mostra correlação negativa muito sutil, enquanto o ângulo azimutal apresenta correlação negativa moderada de -0,45, capturando o efeito da orientação solar sobre a incidência da irradiância. A distribuição de  $T_{mod}$  segue aproximadamente uma distribuição normal, concentrando-se predominantemente entre 45°C e 55°C, o que sugere estabilidade térmica relativa do módulo FV nas condições estudadas, conforme Figura 4.

#### 4.1.1 Irradiância solar

$G$  apresenta ampla variação ao longo do período estudado (22 de maio de 2023 a 30 de maio de 2025), com padrões característicos tanto diários quanto sazonais fortemente influenciados pela dinâmica climática do semiárido cearense. Os valores medidos variam de 10,0 W/m<sup>2</sup> (mínimo, registrado no início e final do dia) a 1249,5 W/m<sup>2</sup> (máximo, em condições de céu claro durante estação seca), com média de 523,1 W/m<sup>2</sup> e mediana de 554,6 W/m<sup>2</sup>, refletindo a dinâmica natural de  $G$  sobre módulos FV em condições reais de operação.

A Figura 5a (gráfico inferior azul) mostra a caracterização horária da irradiância, onde o eixo horizontal indica a hora do dia e o eixo vertical indica  $G$  medida, evidenciando o padrão diário típico de  $G$  com pico próximo ao meio-dia solar. O padrão diário de irradiância

Figura 4 – Histogramas de frequência e correlações entre os pares de variáveis do banco de dados do LEA-UFC (22 de maio de 2023 a 30 de maio de 2025).

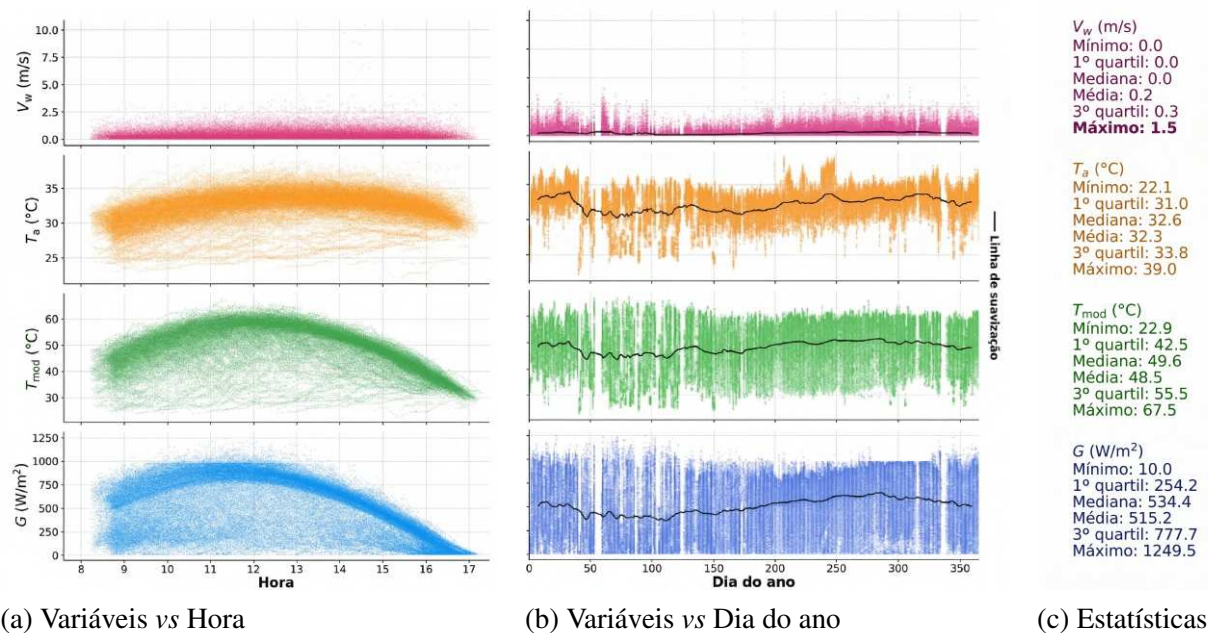


Fonte: Elaborado pelo autor.

Nota: A variável  $T_c$  apresentada na figura refere-se a  $T_{mod}$ .

é bem definido: a partir das 8:00h, observa-se comportamento crescente até atingir valores máximos entre 11:00h e 13:00h, período no qual as medianas aproximam-se de 554,6 W/m<sup>2</sup> com picos ocasionais chegando a 1249,5 W/m<sup>2</sup>, conforme a Figura 5a, e Figura 6, gráfico superior direito. Após as 14:00h, os valores decrescem significativamente, caracterizando o período vespertino.

Figura 5 – Gráficos de caracterização de  $V_w$ ,  $T_a$ ,  $T_{mod}$  e  $G$ : (a) ao longo do dia (b) ao longo do ano (c) estatísticas descritivas.

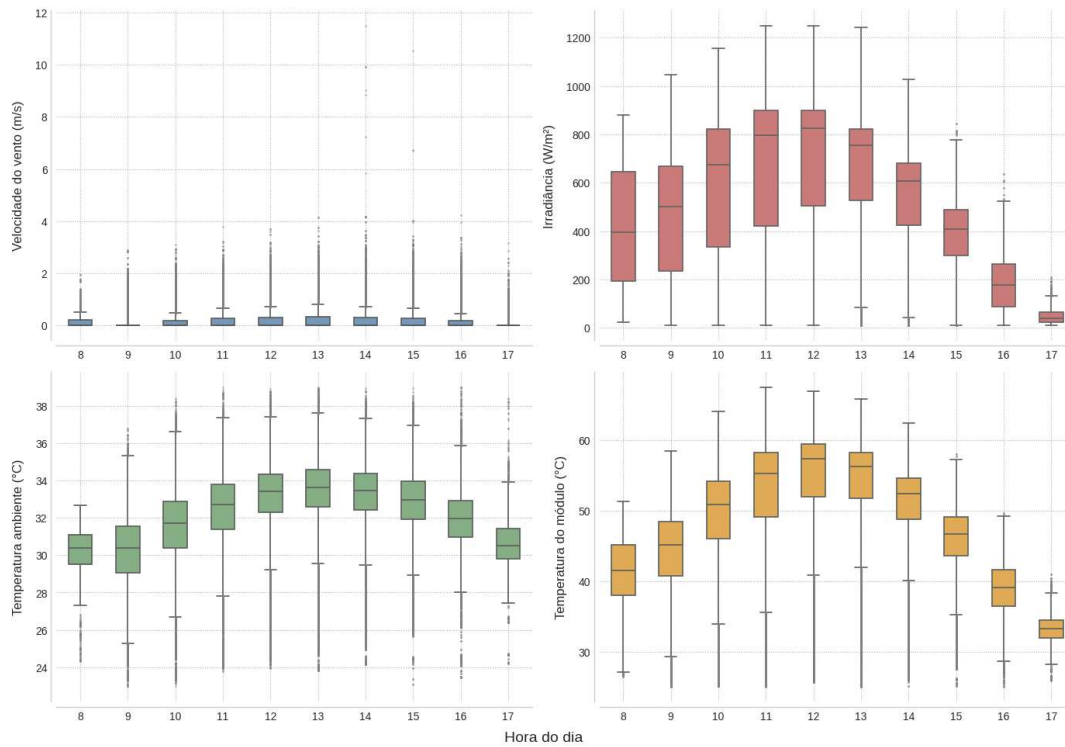


Fonte: Elaborado pelo autor.

A Figura 5b (gráfico inferior azul) apresenta a distribuição temporal de  $G$  ao longo do período, evidenciando a forte sazonalidade característica do clima semiárido. Observam-se dois padrões distintos: (i) estação chuvosa (aproximadamente dias 01–20, 256–376 e 622–730 do ano), com redução de irradiância devido ao aumento de nebulosidade, concentrando valores predominantemente entre 350–500 W/m<sup>2</sup>; e (ii) estação seca (aproximadamente dias 21–255 e 377–621), caracterizada por céu predominantemente limpo e valores sustentados entre 600–800 W/m<sup>2</sup>, com picos frequentes acima de 900 W/m<sup>2</sup>.

A Figura 5c (seção inferior azul) apresenta as estatísticas descritivas de  $G$ : valores mínimo (10,0 W/m<sup>2</sup>) e máximo (1249,5 W/m<sup>2</sup>); média (515,2 W/m<sup>2</sup>) e mediana (534,4 W/m<sup>2</sup>); e os quartis 1º (254,0 W/m<sup>2</sup>) e 3º (777,7 W/m<sup>2</sup>). O 1º quartil representa o valor abaixo do qual se encontram 25% das observações, enquanto o 3º quartil delimita os 75% inferiores da distribuição.

Figura 6 – Boxplot de distribuição das variáveis por hora do dia.



Fonte: Elaborado pelo autor.

#### 4.1.2 Temperatura ambiente

Devido à localização geográfica da região e às características do clima semiárido local, observa-se variação moderada de  $T_a$  ao longo do período estudado. As medições foram realizadas utilizando o sensor PT100, o qual possui uma precisão nominal de  $\pm 0,14^\circ\text{C}$  de acordo com as especificações do fabricante. A partir desses dados, nota-se uma amplitude térmica anual relativamente pequena, mas com oscilações diárias significativas. Os valores mínimos de temperatura ambiente são observados principalmente no início da manhã, com médias em torno de  $28,0\text{--}30,0^\circ\text{C}$ , sendo registrado o valor mínimo absoluto de  $22,1^\circ\text{C}$ .

Já os valores máximos ocorrem durante o período de maior incidência solar (12:00h–14:00h), com médias chegando a  $34^\circ\text{C}$  e pico máximo de  $39,0^\circ\text{C}$  durante a estação seca, quando a reduzida umidade relativa (40–50%) e o céu claro favorecem maior aquecimento da superfície, conforme observado nas estatísticas descritivas (Figura 5c, segunda seção).

Essa variação de  $T_a$  ao longo do ano pode ser vista na Figura 5b (gráfico laranja), onde se observa distribuição relativamente estável ao longo do período analisado, com oscilações típicas de variação diária e leve sazonalidade associada aos regimes de precipitação. A amplitude térmica anual é pequena (aproximadamente  $3\text{--}4^\circ\text{C}$  entre médias mensais da estação chuvosa e

seca), característica de regiões de baixa latitude onde a variação do ângulo de incidência solar ao longo do ano é menor comparada a regiões temperadas.

A Figura 5a (gráfico laranja) também mostra caracterização das temperaturas horárias. As temperaturas são crescentes a partir das primeiras horas da manhã (6:00h–7:00h) até atingirem seus valores máximos entre 12:00h e 14:00h, com medianas próximas a 32,5°C e valores de 3º quartil de 33,7°C neste período, conforme Figura 6. A Figura 5c (segunda seção) apresenta as estatísticas descritivas de  $T_a$ , indicando média geral de 32,3°C e mediana de 32,6°C para todo o período estudado.

A análise estatística dos dados de temperatura ambiente revela valores significativos para o período estudado: temperatura mínima de 23,00°C e máxima de 39,00°C. A mediana de 32,56°C e média de 32,23°C indicam distribuição aproximadamente simétrica. Os quartis demonstram a dispersão dos dados, com primeiro quartil em 31,12°C e terceiro quartil em 34,02°C, evidenciando variação relativamente pequena na distribuição das temperaturas ao longo do período analisado.

#### 4.1.3 Temperatura do módulo

A distribuição horária de  $T_{mod}$ , mostrada na Figura 6 (boxplot em verde), exhibe um padrão claro de aquecimento durante o dia, com as temperaturas mais elevadas ocorrendo entre 11:00h e 13:00h, onde as medianas alcançam 49,0°C. As temperaturas mais baixas do período de geração são registradas no início da manhã (08:00h), com medianas de 34,0°C, e no final da tarde (17:00h), com medianas de 34,5°C.

É notável a forte correlação entre a  $T_{mod}$  e  $G$ , com ambas apresentando picos simultâneos durante as horas centrais do dia.  $T_{mod}$  tende a ser significativamente superior à temperatura ambiente, com uma diferença que atinge entre 15,0°C e 18,0°C nos momentos de maior  $G$ . Esta diferença é particularmente evidente quando se compara os valores máximos diários: enquanto a temperatura do módulo frequentemente ultrapassa 50,0°C, chegando ao máximo absoluto de 67,5°C, a temperatura ambiente atinge seu pico em 39,0°C.

A Figura 5c (terceira seção) apresenta as estatísticas descritivas de  $T_{mod}$ : temperatura mínima de 22,9°C, máxima de 67,5°C, mediana de 49,6°C e média de 48,5°C. A Figura 5a (gráfico verde) e a Figura 5b (gráfico verde) mostram a variação horária e a série temporal de  $T_{mod}$ , respectivamente.

#### 4.1.4 *Velocidade do vento*

A análise dos dados mostra que  $V_w$  apresenta variabilidade ao longo do período analisado, como evidenciado pela série temporal (Figura 5b, gráfico superior), porém com predominância de valores baixos. Os valores permanecem consistentemente baixos, com mediana de 0,0 m/s e média de 0,2 m/s, variando principalmente entre 0,0 e 0,3 m/s (3º quartil), com máximo registrado de 1,5 m/s nas estatísticas gerais do período.

Esta característica de baixas velocidades é atribuída à configuração experimental do sistema de monitoramento, conforme a Figura 2c: o sensor anemométrico está instalado próximo aos módulos FV, em área urbana rodeada por edifícios e vegetação arbórea, resultando em redução substancial das velocidades medidas devido ao efeito de barreira física e rugosidade superficial elevada. Esta condição de microclima local tem implicações para o balanço térmico dos módulos: a reduzida convecção forçada limita o efeito de resfriamento do vento, justificando sua baixa correlação (0,15) com  $T_{mod}$  observada na matriz de correlações (Figura 4) e sua contribuição reduzida (<2%) nos modelos de ML.

Na distribuição horária (Figura 6, gráfico superior esquerdo), observa-se que, embora as medianas permaneçam consistentemente baixas ao longo do dia (0,0 a 0,5 m/s), ocorrem picos ocasionais em diversos horários, com máximo atingindo até 1,5 m/s, particularmente registrados entre 13h e 15h, possivelmente associados a brisas térmicas locais. A dispersão dos dados é evidente tanto na variação horária (Figura 5a, gráfico superior) quanto na temporal (Figura 5b, gráfico superior), refletindo a natureza intermitente dos ventos na região.

#### 4.1.5 *Ângulos solares*

O ângulo zenital solar apresenta variação característica ao longo do dia, iniciando com valores elevados no início da manhã (aproximadamente 80,0° às 08:00h) e diminuindo progressivamente até atingir seu mínimo no meio-dia solar (aproximadamente 11:40h). Devido à baixa latitude de Fortaleza (3,7° S), o valor mínimo do zênite oscila entre 0,0° e 27,0° ao longo das estações, indicando que o sol atinge posições muito próximas à verticalidade em dois períodos do ano. Esse comportamento é demonstrado na matriz de dispersão da Figura 4. A série temporal anual mostra oscilações em torno de 40,0°–45,0°, refletindo a variação da declinação solar.

O ângulo azimutal apresenta um padrão diário crescente, iniciando com valores

próximos a  $100,0^\circ$  às 08:00h (orientação leste) e aumentando até aproximadamente  $300,0^\circ$  às 17:00h (orientação oeste). A série temporal anual mostra uma variação sazonal em que os valores médios diários oscilam predominantemente entre  $120,0^\circ$  e  $160,0^\circ$ , com picos isolados acima de  $200,0^\circ$  em determinados períodos do ano, conforme ilustrado no gráfico de dispersão da Figura 4.

#### 4.1.6 Variabilidade diária e padrões sazonais

A análise revela padrões distintos de variabilidade diária e tendências sazonais fortemente acopladas ao regime de precipitação característico do semiárido, conforme caracterizado na Seção 4.1. A variabilidade diária é particularmente pronunciada em  $T_{mod}$  e  $G$ , com oscilações de até  $30^\circ\text{C}$  e  $800\text{ W/m}^2$ , respectivamente, refletindo os ciclos diurnos característicos. A variável  $T_a$  apresenta amplitude diária moderada (aproximadamente  $13^\circ\text{C}$ ), enquanto  $V_w$  exibe a menor variabilidade (até  $0,8\text{ m/s}$ ).

Ao observar a série temporal completa, as variações evidenciam comportamentos sazonais cíclicos distintos, diretamente associados às estações chuvosa e seca:

- $G$ : Apresenta a sazonalidade mais pronunciada e bem definida entre todas as variáveis. Durante a estação seca (aproximadamente dias 200–400 no primeiro ano e 565–730 no segundo ano), os valores médios diários atingem  $650\text{--}660\text{ W/m}^2$ , refletindo condições de céu predominantemente limpo e alta disponibilidade solar. Em contraste, durante a estação chuvosa (aproximadamente dias 50–150 no primeiro ano e 415–515 no segundo ano), a irradiância média reduz para aproximadamente  $350\text{--}400\text{ W/m}^2$  devido ao aumento de nebulosidade associado ao período de máxima precipitação.
- $T_{mod}$ : Segue diretamente o padrão de  $G$ , registrando médias mais amenas, entre  $44\text{--}45^\circ\text{C}$ , durante a estação chuvosa (dias 50–150 e 415–515), e atingindo os maiores patamares, entre  $50\text{--}51^\circ\text{C}$ , durante a estação seca (dias 200–400 e 565–730), com picos ocasionais acima de  $60^\circ\text{C}$  em condições de alta irradiância e baixa ventilação. Esta forte correlação visual entre as tendências de  $T_{mod}$  e  $G$  confirma a dependência direta da temperatura do módulo em relação à irradiância incidente, conforme esperado fisicamente e validado pelo coeficiente de correlação de  $0,84$  (Figura 4).
- $T_a$ : Mantém-se relativamente estável ( $30\text{--}34^\circ\text{C}$ ), com variação sazonal atenuada, o que é característico do clima semiárido de baixa latitude, onde a amplitude térmica anual é significativamente menor que a amplitude diária. Observa-se uma leve redução (aproximadamente  $1\text{--}2^\circ\text{C}$ ) durante a estação chuvosa em relação à estação seca.

madamente 2–3°C) durante os períodos chuvosos devido ao aumento de nebulosidade e umidade relativa (75–85%), se comparado aos períodos secos de umidade reduzida (40–50%).

- $V_w$ : Apresenta uma leve tendência decrescente em sua linha de base ao longo dos dois anos de monitoramento, variando de aproximadamente 0,27 m/s no início do período para 0,05 m/s ao final do segundo ano, exibindo a menor variabilidade sazonal dentre as variáveis. A predominância de valores muito baixos, com uma média global de apenas 0,2 m/s (como mencionado anteriormente na Seção 4.1.4), é atribuída à configuração experimental do sistema em ambiente com alta rugosidade. Esta condição de microclima local é representativa de instalações FV em ambientes urbanos e explica a baixa influência desta variável (<2%) nos modelos preditivos de  $T_{mod}$  desenvolvidos nesta dissertação.

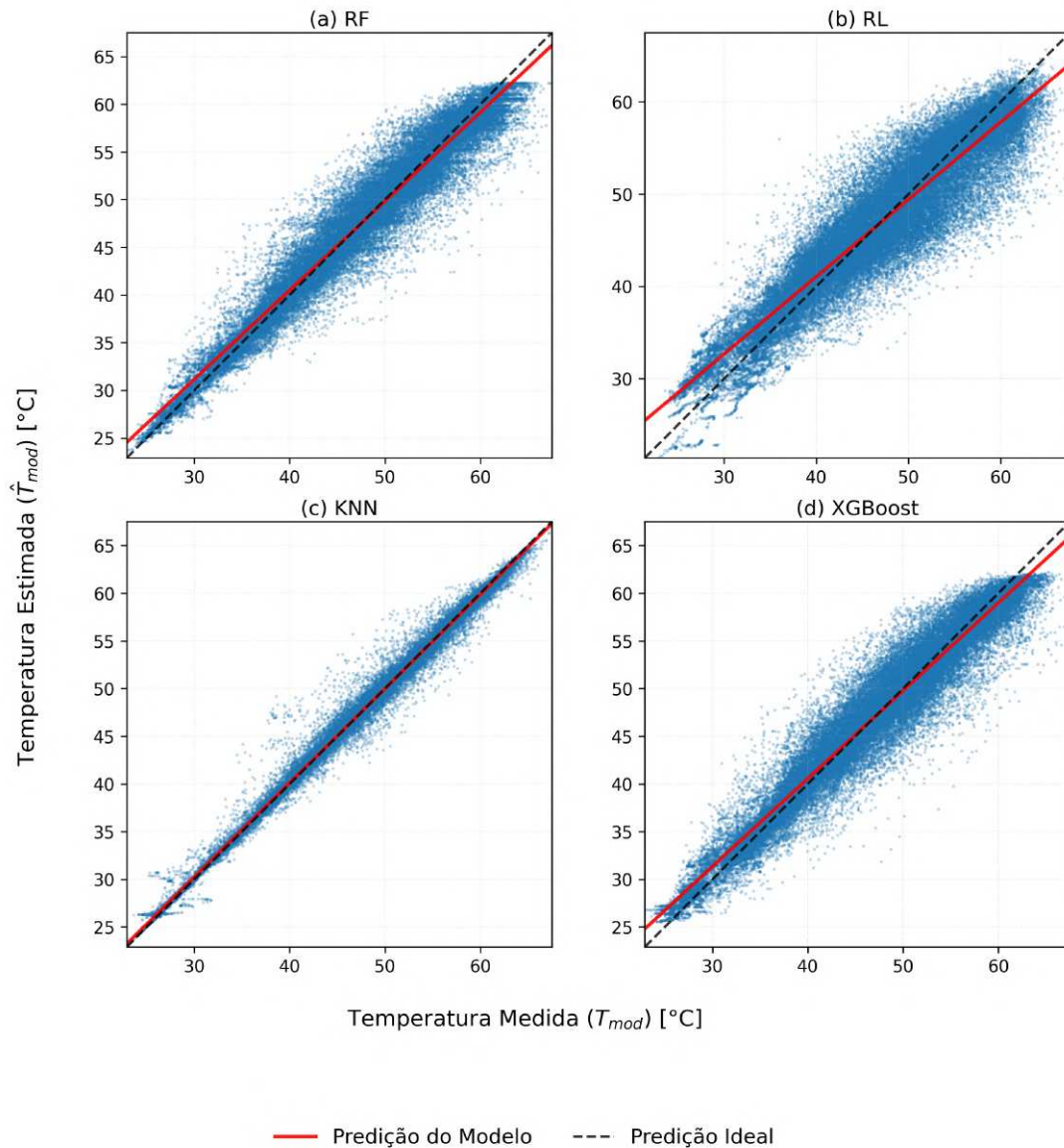
Esses dois ciclos sazonais completos capturados no período de monitoramento (maio/2023–maio/2025) garantem representatividade estatística robusta para o treinamento e a validação dos modelos de ML. Dessa forma, os algoritmos são capazes de mapear tanto as variações diárias quanto as oscilações sazonais típicas do semiárido local, um fator fundamental para a generalização adequada das previsões de temperatura de módulos FV.

## 4.2 Análise de desempenho dos modelos de ML

A  $\hat{T}_{mod}$  estimada é representada na Figura 7, expondo os resultados no conjunto de teste dos modelos RF, KNN, *XGBoost* e RL, respectivamente, na forma de gráficos de dispersão Predito vs. Real. Este tipo de visualização permite avaliar o grau de precisão com que o modelo de regressão realiza estimativas para diferentes valores de resposta. A distância vertical da reta diagonal (estimativa perfeita) a qualquer ponto representa o erro da estimativa para aquela observação.

Nos quatro gráficos de dispersão, os pontos distribuem-se ao longo da reta de correlação entre os valores preditos e reais, com densidade variável de *outliers* dependendo do modelo. Observa-se que o modelo KNN (Figura 7c) apresenta a distribuição mais concentrada ao redor da reta diagonal, com dispersão lateral mínima e raros *outliers*, refletindo seu desempenho superior (RMSE = 1,44°C,  $R^2 = 0,9703$ ). Os modelos RF (Figura 7a) e *XGBoost* (Figura 7d) exibem comportamento intermediário, com dispersão lateral moderada e simetria adequada ao redor da linha de predição perfeita, consistente com seus valores de RMSE de 2,11°C e 2,29°C, respectivamente. Em contraste, o modelo RL (Figura 7b) apresenta dispersão lateral acentuada,

Figura 7 – Comparação entre valores preditos ( $\hat{T}_{mod}$ ) e reais ( $T_{mod}$ ) para os modelos ML



Fonte: Elaborado pelo autor.

particularmente na faixa de temperaturas intermediárias (35–45°C), com maior densidade de *outliers* e desvios sistemáticos, resultando no maior RMSE (3,24°C) entre os modelos avaliados.

Para uma análise mais detalhada da capacidade preditiva sob diferentes dinâmicas atmosféricas, a Figura 8 apresenta a comparação entre os valores estimados pelos algoritmos (KNN e RF) e os valores reais medidos de  $T_{mod}$ . O recorte temporal compreende os cinco primeiros dias do conjunto de teste (iniciado em 20 de maio). Do ponto de vista climatológico, este período corresponde à transição entre o final da quadra chuvosa e o início da estação seca no semiárido cearense. Esta janela temporal é particularmente representativa para a validação, pois exibe um cenário misto: contém tanto dias de céu predominantemente limpo (caracterizados

por curvas térmicas suaves e contínuas) quanto dias com alta intermitência de nebulosidade (evidenciados por quedas abruptas de temperatura nas horas centrais). Pode-se observar que o modelo KNN, de melhor desempenho geral, acompanha de perto os padrões de variação intradiária de  $T_{mod}$ , capturando com alta precisão tanto os picos de aquecimento diário quanto os vales bruscos de resfriamento causados por eventos transientes.

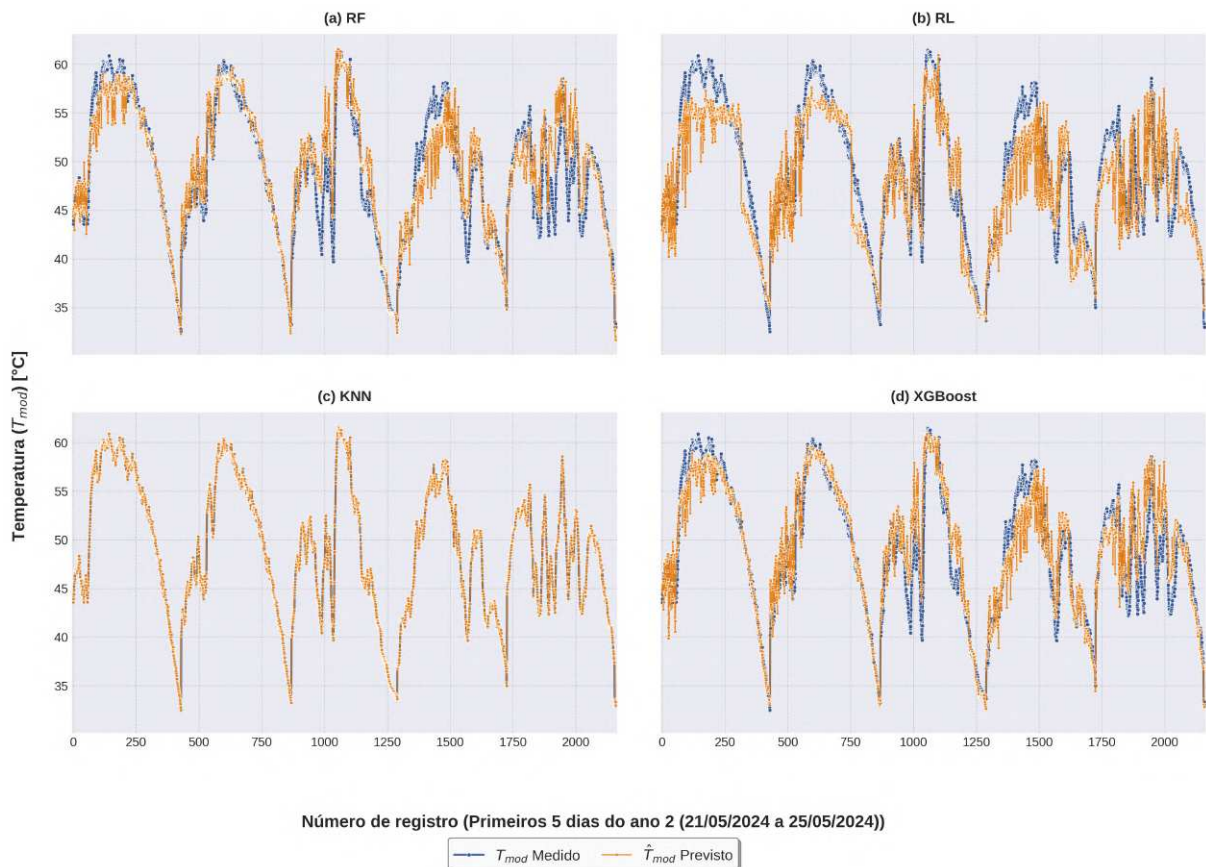
A fim de avaliar o desempenho dos modelos propostos, a Tabela 4 mostra os principais resultados dos modelos de ML no conjunto de teste.

Tabela 4 – Principais resultados dos modelos de ML

Modelo	MAE (°C)	RMSE (°C)	MBE (°C)	P95 (°C)	nRMSE (%)	$R^2$
KNN	0,8318	1,4361	-0,1410	3,13	3,20	0,9703
RF	1,5465	2,1113	-0,2693	4,44	4,75	0,9358
XGBoost	1,6967	2,2939	-0,3884	4,77	5,16	0,9242
RL	2,5506	3,2375	-0,7013	6,35	7,27	0,8491

Fonte: Elaborado pelo autor.

Figura 8 – Comparação entre  $T_{mod}$  média medida e  $\hat{T}_{mod}$  estimada para os modelos ao longo dos primeiros 5 dias do conjunto de teste



Fonte: Elaborado pelo autor.

Os resultados da Tabela 4 indicam que o modelo KNN apresenta os melhores recursos para estimar  $T_{mod}$ , com os menores valores de erro (RMSE de 1,44°C e MAE de 0,83°C) e o melhor  $R^2$  (0,97). Este resultado demonstra que a magnitude média do erro do modelo em estimar  $T_{mod}$  é de aproximadamente 1,44°C, com o modelo capaz de explicar 97,03% da variância de  $T_{mod}$ .

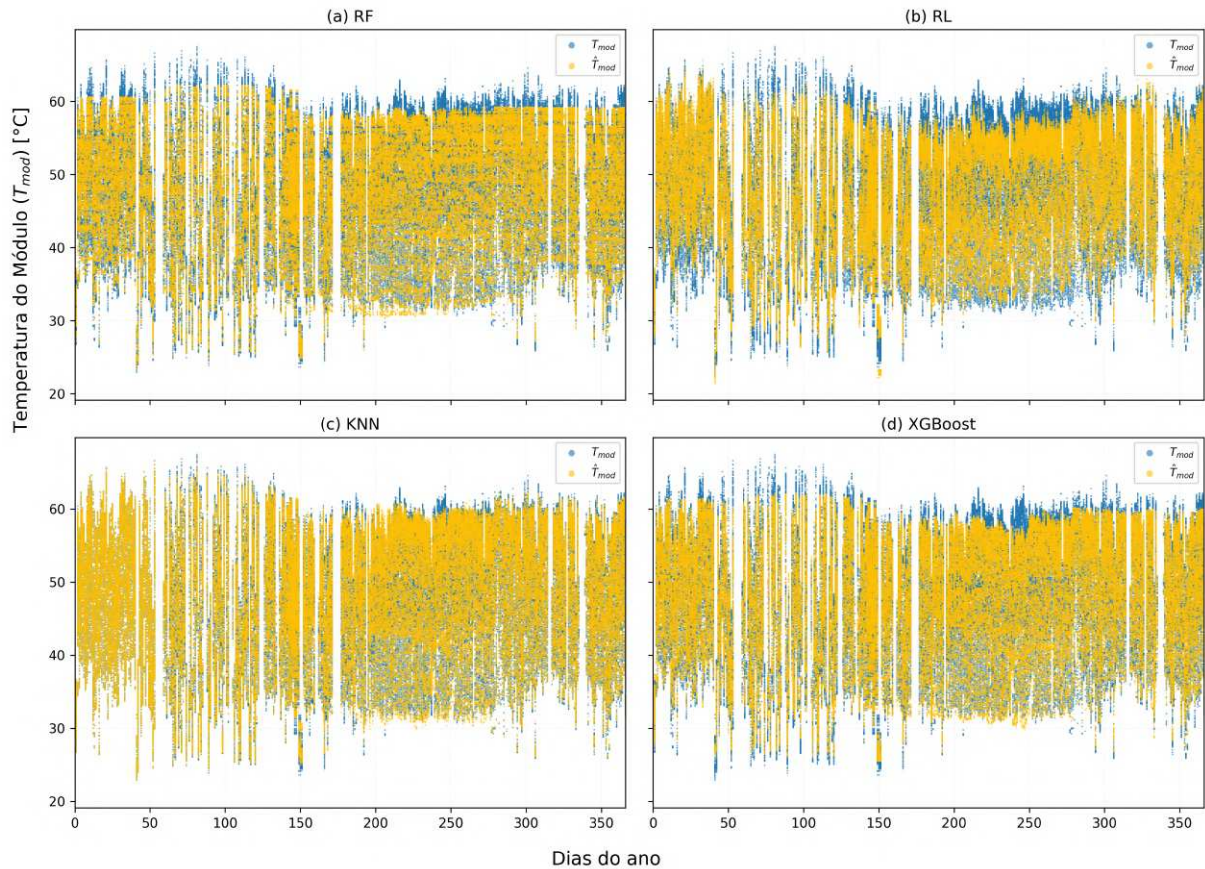
A Figura 9 exibe as previsões ( $\hat{T}_{mod}$ ) geradas pelas técnicas RF, RL, KNN e *XGBoost*. A resposta estimada (em amarelo) é sobreposta aos valores reais medidos de  $T_{mod}$  do conjunto de teste (em azul) em função dos dias contínuos de monitoramento. Comparando os painéis da Figura 9, observa-se que o modelo KNN apresenta uma cobertura visualmente superior dos dados de teste ao longo de todo o período, capturando adequadamente tanto as tendências sazonais quanto as variações intradiárias. Embora o KNN apresente o melhor desempenho quantitativo global, nota-se uma ligeira dificuldade comum aos algoritmos em atingir a amplitude total de alguns valores extremos de temperatura (picos de aquecimento típicos da estação seca e quedas abruptas de transição). O modelo de RL, por sua vez, mostra limitações claras e severas, especialmente em prever os picos de temperatura acima de 60,0°C, subestimando sistematicamente o estresse térmico máximo e sendo destacadamente o de pior desempenho nessa faixa operacional.

O modelo RF apresentou o segundo melhor desempenho, com RMSE de 2,11°C e  $R^2$  de 0,94, demonstrando boa capacidade preditiva. O MAE de 1,55°C indica erros médios absolutos relativamente baixos, consolidando o RF como uma alternativa robusta para a estimativa de  $T_{mod}$  em aplicações práticas. O MBE de -0,27°C revela uma leve tendência de subestimação da temperatura, comportamento comum em modelos *ensemble* baseados em árvores de decisão.

O modelo *XGBoost* obteve desempenho muito próximo ao RF, com RMSE de 2,29°C e  $R^2$  de 0,92 no conjunto de teste (Tabela 4), confirmando que métodos baseados em árvores de decisão são altamente eficazes para esta aplicação. O MAE de 1,70°C e MBE de -0,39°C indicam comportamento similar ao RF, com ligeira tendência de subestimação das temperaturas mais elevadas.

A RL apresentou o pior desempenho entre todos os modelos testados, com RMSE de 3,24°C, MAE de 2,55°C e  $R^2$  de 0,85 no conjunto de teste. Este resultado era esperado, considerando que a relação entre as variáveis preditoras e a temperatura do módulo apresenta características não-lineares que não podem ser adequadamente capturadas por um modelo linear simples. O MBE de -0,70°C, o mais negativo entre todos os modelos, indica maior tendência

Figura 9 – Estimativa de  $\hat{T}_{mod}$  (amarelo) sobreposta aos dados reais (azul) vs. dia do ano para os modelos: (a) RF, (b) RL, (c) KNN, (d) *XGBoost*.



Fonte: Elaborado pelo autor.

de subestimação sistemática da temperatura, particularmente durante períodos de alta irradiância.

### 4.3 Modelos físicos de referência

Para estabelecer *benchmarks* científicos robustos e quantificar adequadamente os ganhos proporcionados pelos métodos de ML, foram implementados três modelos físico-semiempíricos consolidados na literatura: *Ross*, *Faiman* e *PVsyst*. Estes modelos apresentam complexidade crescente na representação dos mecanismos de transferência de calor e serviram como linha de base para avaliar sistematicamente o desempenho dos algoritmos de ML na estimativa de  $T_{mod}$ .

Os três modelos foram otimizados utilizando o conjunto de treinamento através de técnicas de minimização. O modelo de *Ross* foi calibrado por mínimos quadrados restritos<sup>3</sup>,

<sup>3</sup> O método de mínimos quadrados restritos é uma técnica de otimização que minimiza a soma dos quadrados dos resíduos (diferenças entre valores observados e estimados) sujeita a restrições nos parâmetros (LAWSON; HANSON, 1995). No contexto de modelagem térmica, essas restrições garantem que os coeficientes estimados respeitem limites físicos conhecidos, como positividade de coeficientes de transferência de calor ou faixas típicas

enquanto os modelos de *Faiman* e *PVsys* utilizaram evolução diferencial (*differential evolution*) devido à sua natureza não-linear. O modelo de *Ross* segue a formulação da Equação 2.8, onde  $r$  representa o coeficiente de proporcionalidade entre irradiância e elevação de temperatura, otimizado em  $r = 0,0283 \text{ } ^\circ\text{C} \cdot \text{m}^2/\text{W}$ .

Os modelos de *Faiman* e *PVsys* compartilham estrutura análoga de balanço de energia, incorporando explicitamente o efeito de  $V_w$  na dissipação térmica. Para o modelo de *Faiman*, a otimização estocástica resultou nos parâmetros empíricos  $U_0 = 35,26 \text{ W}/(\text{m}^2 \cdot \text{K})$  e  $U_1 = 0,554 \text{ W} \cdot \text{s}/(\text{m}^3 \cdot \text{K})$ . Já no modelo *PVsys*, que contabiliza ativamente as perdas ópticas e a conversão elétrica, adotou-se a absorvidade típica do módulo de silício ( $\alpha = 0,90$ ) e a eficiência nominal da placa ( $\eta = 17\%$ ). Com o acoplamento físico no numerador, os parâmetros de perda térmica otimizados convergiram para  $u_c = 26,34 \text{ W}/(\text{m}^2 \cdot \text{K})$  e  $u_v = 0,414 \text{ W} \cdot \text{s}/(\text{m}^3 \cdot \text{K})$ . A diferença de magnitude entre os coeficientes das duas abordagens reflete a correção do balanço físico no modelo *PVsys*: ao descontar a fração da irradiância não convertida em calor, o modelo passa a exigir coeficientes de dissipação térmica proporcionalmente menores para representar o mesmo aquecimento observado nos dados experimentais.

A Tabela 5 apresenta as métricas de desempenho dos três modelos físico-semiempíricos avaliados no conjunto de teste, correspondente a 50% dos dados temporalmente posteriores ao treinamento.

Tabela 5 – Desempenho dos modelos físico-semiempíricos de referência no conjunto de teste

<b>Modelo</b>	<b>MAE (<math>^\circ\text{C}</math>)</b>	<b>RMSE (<math>^\circ\text{C}</math>)</b>	<b>MBE (<math>^\circ\text{C}</math>)</b>	<b>P95 (<math>^\circ\text{C}</math>)</b>	<b>nRMSE (%)</b>	<b><math>R^2</math></b>
<i>Ross</i>	3,5722	4,9128	-1,8606	10,81	10,09	0,6515
<i>Faiman</i>	3,5692	4,9105	-1,8559	10,82	10,08	0,6518
<i>PVsys</i>	3,5690	4,9110	-1,8560	10,82	10,08	0,6518

Fonte: Elaborado pelo autor.

Os resultados revelam desempenho praticamente equivalente entre os três modelos, com os modelos de *Faiman* e *PVsys* apresentando métricas idênticas até a quarta casa decimal devido à sua formulação matematicamente equivalente. O modelo de *Ross* apresenta desempenho marginalmente inferior, com diferença de apenas  $0,0023^\circ\text{C}$  no RMSE. Todos os modelos físicos alcançaram MAE de aproximadamente  $3,57^\circ\text{C}$  e coeficiente de determinação  $R^2 \approx 0,65$ , indicando capacidade de explicar cerca de 65% da variância de  $T_{mod}$ .

Um aspecto crítico observado é o viés sistemático negativo em todos os modelos

---

reportadas na literatura, evitando soluções matematicamente válidas mas fisicamente inconsistentes.

físicos, com  $MBE \approx -1,86^\circ\text{C}$ , evidenciando tendência consistente de subestimação de  $T_{mod}$ . Este comportamento sugere limitações das formulações físicas simplificadas em capturar completamente os mecanismos de aquecimento, particularmente em condições de alta irradiância solar. O percentil 95 do erro absoluto próximo a  $10,8^\circ\text{C}$  indica que 5% das estimativas apresentam desvios superiores a este valor, representando situações desafiadoras para os modelos baseados exclusivamente em princípios físicos.

O nRMSE de aproximadamente 10% estabelece linha de base para avaliação dos modelos de ML, que deverão superar este desempenho para justificar sua complexidade adicional. A similaridade entre *Faiman* e *PVsyst* confirma que a incorporação explícita de  $V_w$  melhora minimamente a capacidade preditiva em relação ao modelo de *Ross*, sugerindo que outros fatores não modelados como radiação de onda longa, propriedades espectrais da irradiância e efeitos de sujidade possam ser mais relevantes para reduzir o erro sistemático observado.

A Figura 10 apresenta os gráficos de dispersão comparativos entre valores observados e preditos pelos três modelos de referência: *Ross*, *Faiman* e *PVsyst*. Esta análise abrangente permite avaliar o desempenho tanto dos modelos estatísticos lineares quanto dos modelos físicos estabelecidos na literatura.

#### 4.4 Modelagem híbrida

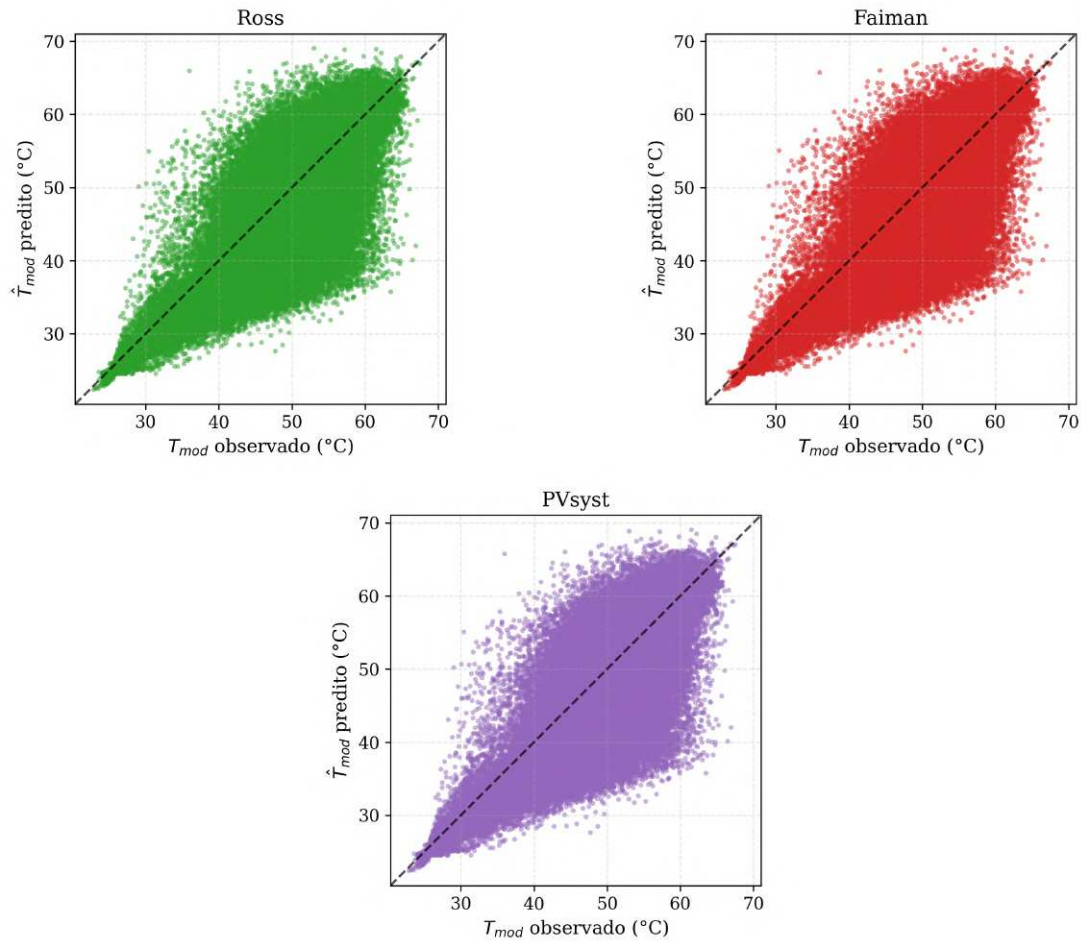
Para explorar o potencial combinado de conhecimento físico e aprendizado de dados, foram desenvolvidos modelos semiempíricos baseados em três formas gerais de correlações descritas na literatura (SANTOS *et al.*, 2021): linear, não-linear e racional. Essas abordagens combinam fundamentos teóricos de transferência de calor com ajustes empíricos calibrados localmente através de otimização numérica robusta, visando balancear precisão preditiva e interpretabilidade física dos parâmetros obtidos.

A correlação linear adapta a forma geral da Equação 2.5, eliminando o termo de temperatura de *back-surface* ( $a_4 = 0$ ) e normalizando  $G$  para melhor condicionamento numérico:

$$\hat{T}_{mod} = a_0 + a_1 T_a + a_2 \left( \frac{G}{1000} \right) + a_3 V_w \quad (4.1)$$

onde a normalização  $G/1000$  reduz a magnitude da irradiância de  $[10, 1250]$   $\text{W}/\text{m}^2$  para  $[0, 001, 1, 25]$ , melhorando a estabilidade do algoritmo de mínimos quadrados restritos. Os limites impostos foram  $a_0 \in [-50, 50]$ ,  $a_1 \in [0, 2]$ ,  $a_2 \in [0, 100]$  e  $a_3 \in [-10, 0]$ , garantindo coeficientes fisicamente

Figura 10 – Comparação entre  $T_{mod}$  e  $\hat{T}_{mod}$  nos modelos de referência.



Fonte: Elaborado pelo autor.

consistentes.

A correlação não-linear simplifica a forma geral da equação 2.6, linearizando os expoentes ( $g_1 = g_2 = g_3 = 1$ ), removendo o termo exponencial com  $V_w$  e mantendo apenas produtos de primeira ordem:

$$\hat{T}_{mod} = b_0 + b_1 T_a + C_1 (d_1 + d_2 V_w + e_1) \left( \frac{G}{1000} \right) C_2 \left( e_2 + f_1 T_a + f_2 \left( \frac{G}{1000} \right) + f_3 V_w \right) \quad (4.2)$$

Esta formulação reduz de 16 para 11 parâmetros otimizáveis [ $b_0, b_1, C_1, d_1, d_2, e_1, C_2, e_2, f_1, f_2, f_3$ ], calibrados via algoritmo de *Levenberg-Marquardt* com até 10.000 avaliações de função. A normalização de  $G$  foi mantida consistentemente para evitar problemas de escala numérica durante a convergência, enquanto a simplificação estrutural preserva a capacidade de capturar interações não lineares entre variáveis.

A correlação racional mantém a estrutura da equação 2.7, mas adota normalização seletiva:  $G$  original no numerador ( $h_2 G$ ) para preservar a proporcionalidade direta com o

aquecimento radiativo, e irradiância normalizada no denominador ( $n_1 G/1000$ ) para estabilidade numérica:

$$\hat{T}_{mod} = \frac{h_1 T_a + h_2 G}{m_1 + n_1 \left(\frac{G}{1000}\right)} \quad (4.3)$$

Esta formulação híbrida com 4 parâmetros  $[h_1, h_2, m_1, n_1]$  oferece limites assintóticos realistas em alta irradiância, evitando divergências numéricas comuns em modelos puramente racionais. Os limites impostos foram  $h_1 \in [0, 1, 3, 0]$ ,  $h_2 \in [0, 001, 0, 1]$  °C · m<sup>2</sup>/W,  $m_1 \in [0, 1, 10, 0]$  e  $n_1 \in [-500, 500]$ , com proteção contra divisão por zero através de *clipping* para  $|\text{denominador}| < 10^{-10}$ .

Funções de validação automática foram implementadas para garantir valores com base no clima local:  $T_a \in [21, 7, 39]$  °C,  $G \in [10, 1250]$  W/m<sup>2</sup> e  $V_w \in [0, 15]$  m/s, eliminando *outliers* e erros de medição antes do processamento.

#### 4.4.1 Desempenho dos modelos semiempíricos com correlações híbridas

A Tabela 6 apresenta o desempenho dos modelos semiempíricos otimizados avaliados no conjunto de teste, correspondente aos 50% dos dados temporalmente posteriores ao treinamento.

Tabela 6 – Desempenho dos modelos semiempíricos com correlações híbridas no conjunto de teste

Modelo	MAE (°C)	RMSE (°C)	MBE (°C)	R <sup>2</sup>	P95( e ) (°C)	nRMSE (%)
Correlação linear (4.1)	2,7833	3,5986	-0,5965	0,8130	7,2400	7,3936
Correlação não-linear (4.2)	2,7833	3,5987	-0,5969	0,8130	7,2418	7,3939
Correlação racional (4.3)	2,8265	3,7097	-0,4373	0,8013	7,5877	7,6218

Fonte: Elaborado pelo autor.

As três correlações híbridas demonstraram desempenho superior aos modelos físicos de referência analisados na seção anterior, com erro médio absoluto consistente entre 2,78 °C e 2,83 °C. A correlação linear (Equação 4.1) e a não-linear (Equação 4.2) alcançaram os melhores desempenhos gerais, capazes de explicar 81,3% da variância de  $T_{mod}$ .

Na correlação linear (Equação 4.1), os parâmetros otimizados  $a_0 = -11,26$ ,  $a_1 = 1,57$ ,  $a_2 = 17,06$  e  $a_3 = -0,007$  revelam intercepto negativo compensado por forte dependência de  $T_a$  (coeficiente próximo a 1,6) e contribuição significativa da irradiância normalizada, enquanto

o efeito do vento é praticamente desprezível ( $a_3 \approx 0$ ), confirmando análises anteriores do microclima urbano.

Apesar da correlação linear compartilhar a mesma natureza matemática do modelo de RL avaliado na Seção 4.2, observa-se uma diferença de desempenho preditivo: a RL obteve RMSE de 3,24 °C, enquanto a correlação linear alcançou 3,60 °C. Esta discrepância é explicada por dois fatores metodológicos fundamentais que ilustram o compromisso entre ajuste estatístico e fidelidade física. Primeiro, a dimensionalidade do espaço de características: o modelo RL foi treinado com cinco variáveis preditoras (incluindo os ângulos zenital e azimutal, que atuam como marcadores da inércia térmica ao longo do dia), enquanto a correlação empírica utiliza apenas as três variáveis termodinâmicas primárias ( $T_a$ ,  $G$  e  $V_w$ ). Segundo, a natureza da otimização: a RL realiza uma minimização irrestrita do erro (puramente orientada aos dados), ao passo que a correlação linear foi otimizada por mínimos quadrados restritos (impondo limites como  $a_3 \leq 0$  para o vento). Isso demonstra que o modelo semiempírico restringe uma pequena fração do ajuste estatístico em prol da consistência termodinâmica estrita dos seus parâmetros.

A correlação não-linear, apesar de sua complexidade com 11 parâmetros otimizáveis, apresentou desempenho estatisticamente equivalente à correlação linear. A diferença de apenas 0,0004 °C no RMSE (sendo 3,5986 °C para a linear e 3,5990 °C para a não-linear) indica que os termos não-lineares adicionais não proporcionaram ganho preditivo significativo para o conjunto de dados estudado. Isso sugere que a relação entre variáveis ambientais e a  $T_{mod}$  pode ser adequadamente capturada por um modelo linear bem calibrado nas condições semiáridas analisadas. Este resultado tem implicações práticas importantes, favorecendo modelos mais simples com menor risco de sobreajuste e maior facilidade de implementação em sistemas de monitoramento em tempo real.

A correlação racional (Equação 4.3) apresentou desempenho ligeiramente inferior aos modelos lineares, com RMSE de 3,7097 °C e  $R^2$  de 0,8013, porém demonstrou vantagem importante na redução do viés sistemático. O MBE de  $-0,4373$  °C representa o menor viés entre todos os modelos avaliados, indicando maior equilíbrio entre subestimação e superestimação ao longo da faixa operacional. Os parâmetros otimizados  $h_1 = 2,76$ ,  $h_2 = 0,0494$  °C · m<sup>2</sup>/W,  $m_1 = 2,30$  e  $n_1 = 0,123$  resultaram em um modelo com comportamento assintótico fisicamente realista em alta irradiância, evitando divergências numéricas comuns em correlações puramente racionais.

A convergência do parâmetro  $n_1$  para valor próximo a zero (0,123) indica que o

denominador  $m_1 + n_1(G/1000) \approx 2,30 + 0,123(G/1000)$  apresenta dependência muito fraca da irradiância. Para  $G = 1000 \text{ W/m}^2$ , o termo  $n_1(G/1000) = 0,123$  representa apenas 5,3% do denominador total, sugerindo que a formulação racional se aproxima de comportamento essencialmente linear nas condições estudadas. Este resultado corrobora a equivalência observada entre as equações 4.1 e 4.2, indicando que correlações lineares otimizadas podem ser tão eficazes quanto modelos não-lineares complexos para estimativa de  $T_{mod}$  em clima semiárido.

#### 4.4.2 Análise da importância das variáveis

Previamente à construção dos modelos híbridos, realizou-se análise sistemática da contribuição relativa de cada variável de entrada para os modelos de ML. Esta etapa é fundamental para orientar a integração entre conhecimento físico e aprendizado empírico, identificando quais variáveis meteorológicas apresentam maior poder preditivo sob perspectiva puramente baseada em dados.

Para os modelos baseados em árvores (RF e *XGBoost*), utilizou-se a métrica nativa de *feature importance*, que quantifica a redução média de impureza de Gini atribuível a cada variável. Para o modelo KNN, empregou-se análise por permutação com 10 repetições, medindo a degradação do erro médio quadrático ao embaralhar aleatoriamente os valores de cada feature.

Os resultados revelaram padrões consistentes entre os diferentes algoritmos não-lineares. O modelo KNN, utilizando análise de permutação, apresentou uma distribuição ligeiramente diferente dos modelos em árvore, atribuindo maior peso relativo à  $T_a$  e aos ângulos solares. Esta diferença pode ser compreendida pela natureza do algoritmo KNN, que é mais sensível a variações locais dos dados e distâncias euclidianas no espaço de features, capturando dependências não-lineares de forma distinta das árvores de decisão.

A Tabela 7 apresenta um resumo das importâncias relativas para os três melhores modelos de ML analisados. Observa-se que  $G$  e  $T_a$  emergem como os fatores dominantes nos modelos baseados em árvores (RF e *XGBoost*), respondendo juntos por uma fatia expressiva da importância total (exatamente 73,78% no RF e 80,66% no *XGBoost*). Os ângulos solares contribuem de forma secundária (25,55% e 18,61% combinados para RF e *XGBoost*, respectivamente), capturando os efeitos dinâmicos da geometria solar, inércia térmica ao longo do dia e possíveis sombreamentos. No modelo KNN, a temperatura ambiente assume o protagonismo isolado (47,63%), embora a soma com a irradiância ainda represente a maior força preditiva do algoritmo (63,47%).

Tabela 7 – Importância relativa das variáveis meteorológicas nos modelos de ML.

Variável	RF (%)	XGBoost (%)	KNN (%)
$G$	41,16	50,81	15,84
$T_a$	32,62	29,85	47,63
$azimuth$	17,26	13,78	18,10
$zenith$	8,29	4,83	10,63
$V_w$	0,67	0,74	1,38

Fonte: Elaborado pelo autor.

Concomitantemente,  $V_w$  apresentou importância estritamente marginal em todos os algoritmos avaliados (0,67%, 0,74% e 1,38% para RF, XGBoost e KNN, respectivamente). Diferentemente do que se poderia supor para um campo aberto em região semiárida, este fenômeno atesta a capacidade dos algoritmos de ML de mapearem independentemente as limitações físicas do microclima local. Conforme elucidado nas Seções 4.1.4 e 4.4, as análises dos dados coletados no LEA-UFC revelaram uma velocidade de vento fortemente atenuada pela rugosidade do ambiente urbano, com média de 0,2 m/s e máximo absoluto limitando-se a 1,5 m/s.

O "desprezo" estatístico da variável eólica por parte dos modelos de ML corrobora perfeitamente as descobertas da modelagem física e semiempírica (onde os coeficientes convectivos  $U_1$  e  $u_v$  e o parâmetro  $a_3$  também convergiram para valores quase nulos). Isso prova de forma inequívoca cruzada que, sob as condições de contorno experimentais deste trabalho, o balanço térmico do módulo FV é ditado em grande parcela pela carga radiativa e por  $T_a$ , sendo a convecção forçada um efeito de resfriamento de segunda ordem.

Estes resultados reforçam a validade das correlações semiempíricas desenvolvidas na Seção 4.4.1, nas quais  $G$  e  $T_a$  são explicitamente parametrizados com coeficientes otimizados, enquanto  $V_w$  aparece com influência baixa ou nula em alguns formalismos, alinhando-se com as descobertas de importância empírica aqui apresentadas.

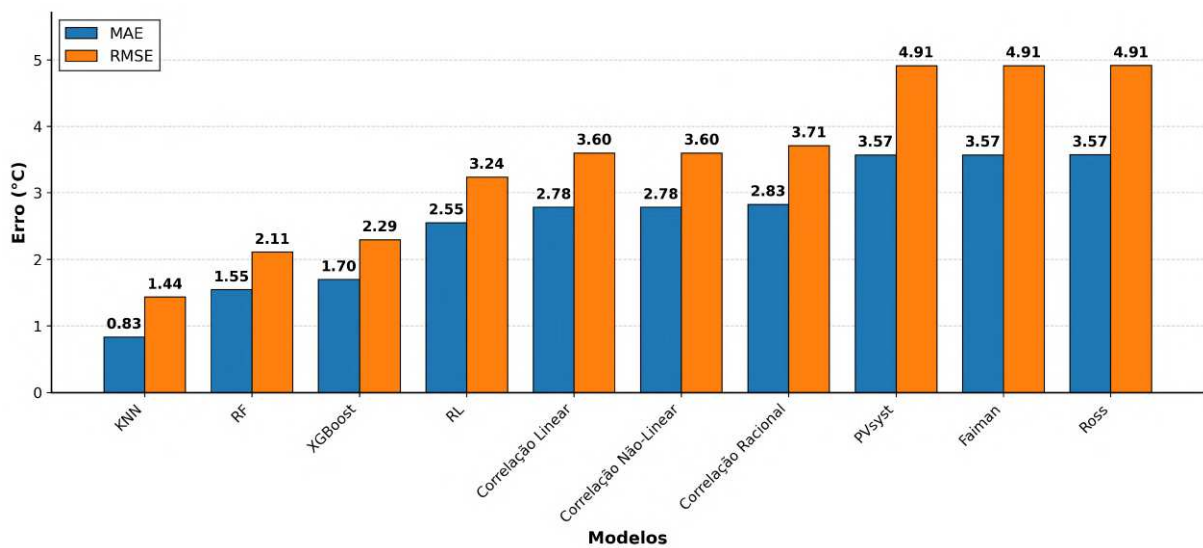
#### 4.5 Análise comparativa de desempenho

Esta seção consolida os resultados obtidos nas etapas anteriores, comparando o desempenho de todos os modelos avaliados: algoritmos de ML (Seção 4.2), modelos físico de referência (Seção 4.3) e correlações semiempíricas híbridas (Seção 4.4). A análise utiliza o conjunto de teste, garantindo avaliação isenta de viés de treinamento. As métricas consideradas incluem MAE, RMSE, MBE,  $R^2$  e nRMSE.

Os resultados evidenciam a superioridade dos algoritmos de ML, com destaque para o KNN e RF. Esta performance representa melhoria substancial comparada ao melhor modelo físico de referência (*PV<sub>sys</sub>*), demonstrando redução de aproximadamente 77% no erro absoluto médio (KNN) e 71% no erro quadrático médio. O algoritmo de RL, embora apresente desempenho inferior aos demais métodos de ML (MAE = 2,55°C, RMSE = 3,24°C), ainda supera significativamente os modelos físicos convencionais, com redução de 29% no MAE.

A Figura 11 apresenta uma comparação visual direta entre todos os modelos avaliados, considerando simultaneamente as métricas MAE e RMSE para análise abrangente do desempenho preditivo. Os modelos de ML demonstram superioridade clara em ambas as métricas, com destaque para KNN, RF e XGBoost. As correlações híbridas apresentam desempenho intermediário, enquanto os modelos físicos de referência exibem os maiores erros.

Figura 11 – Comparação do desempenho de todos os modelos avaliados considerando MAE e RMSE.



Fonte: Elaborado pelo autor.

Os modelos físico-semiempíricos de referência (*Ross*, *Faiman* e *PV<sub>sys</sub>*) apresentaram erros similares entre si (MAE = 3,57°C, RMSE ≈ 4,91°C), com viés sistemático negativo significativo (MBE ≈ -1,86°C), indicando tendência de subestimação da temperatura do módulo. Este comportamento pode ser atribuído à parametrização genérica desses modelos, que não foram especificamente calibrados para as condições climáticas do semiárido cearense características do LEA-UFC, marcadas por alta  $G$  (média de 650 W/m<sup>2</sup> durante estação seca) e baixo  $V_w$  (média de 0,2 m/s devido à configuração urbana do campus).

As correlações semiempíricas híbridas desenvolvidas neste trabalho (Linear, Não-

Linear e Racional) apresentaram desempenho superior aos modelos físicos de referência, com MAE de 2,78–2,83°C, RMSE de 3,60–3,71°C e R<sup>2</sup> de 0,80–0,81, representando melhoria de aproximadamente 22% no MAE comparado aos modelos convencionais. A correlação não-linear (4.2) destacou-se marginalmente com o menor erro (MAE = 2,78°C, RMSE = 3,60°C), seguida pela correlação linear (4.1) com desempenho praticamente equivalente (MAE = 2,78°C, RMSE = 3,60°C). Importante ressaltar que essas correlações apresentaram viés sistemático significativamente reduzido (MBE entre –0,44°C e –0,60°C), representando redução superior a 68% em relação aos modelos físicos de referência, evidenciando melhor adequação às condições locais de operação dos módulos FV estudados. A Tabela 8 sintetiza as vantagens e limitações inerentes a cada algoritmo, avaliadas sob a ótica da aplicabilidade na engenharia. Para padronizar a análise, os critérios foram formulados de modo que maiores pontuações reflitam características operacionais mais desejáveis.

Tabela 8 – Comparação qualitativa entre os modelos de ML avaliados.

<b>Critério de Avaliação</b>	<b>RL</b>	<b>KNN</b>	<b>RF</b>	<b>XGBoost</b>
Desempenho preditivo (Acurácia local)	+	+++	++	++
Interpretabilidade global do modelo	+++	+	++	++
Robustez à escala dos dados	++	+	+++	+++
Facilidade de parametrização	+++	++	+	+
Flexibilidade para novas variáveis	++	++	+++	+++
Eficiência computacional (Inferência)	+++	+	++	+++
Portabilidade e uso de memória	+++	+	++	++

Fonte: Elaborado pelo autor.

**Nota:** A escala indica o grau de vantagem/desempenho do modelo naquele critério: +++ (Alto / Excelente); ++ (Moderado / Bom); + (Baixo / Limitado).

O critério de flexibilidade para incorporação de novas variáveis avalia a capacidade matemática de cada algoritmo em assimilar preditores ambientais extras (como umidade relativa ou índice de sujidade) sem a necessidade de reestruturação analítica. Ao contrário das correlações físicas, que exigem a dedução de uma nova equação governante para cada fenômeno incluído, os modelos de ML requerem apenas a expansão da matriz de *features*. Neste aspecto, os modelos baseados em árvores (RF e XGBoost) recebem pontuação máxima (+++), pois possuem mecanismos nativos de seleção (como o ganho de informação) que isolam preditores ruidosos, além de possuírem alta robustez à escala dos dados, dispensando a normalização. Por outro lado,

a RL e o KNN, embora aceitem novas entradas prontamente, apresentam baixa robustez à escala (+), exigindo re-normalização rigorosa para que o cálculo de distâncias ou os pesos lineares não sejam enviesados por grandezas de magnitudes distintas.

Ademais, a eficiência computacional de inferência e a portabilidade (uso de memória) revelam compromissos importantes para a aplicação industrial. Embora o KNN tenha apresentado o melhor desempenho preditivo neste estudo, sua natureza baseada em instâncias (*lazy learning*) exige que todo o conjunto de dados históricos seja armazenado em memória para o cálculo de distâncias a cada nova predição, limitando sua portabilidade (+). Em contrapartida, algoritmos como a RL e o XGBoost, uma vez treinados, resultam em vetores de pesos ou matrizes de decisão altamente otimizadas, permitindo execução em frações de milissegundo e viabilizando sua implantação direta (*deploy*) em microcontroladores, inversores fotovoltaicos ou sistemas de monitoramento embarcados com restrição de hardware.

A análise qualitativa revela compromissos importantes entre as abordagens de modelagem. Embora os algoritmos de ML apresentem capacidade preditiva indubitavelmente superior, os modelos físicos mantêm vantagens inegáveis em termos de aplicabilidade direta na indústria FV (mediante implementação imediata em *softwares* comerciais de simulação como *PVsyst* (PVsyst SA, 2023; IEA PVPS Task 13, 2017), *HelioScope* (LABS, 2014) e plataformas baseadas em *Meteonorm* (Meteotest AG, 2023)) e interpretabilidade analítica explícita, visto que seus parâmetros possuem significado termodinâmico real.

Neste contexto, embora o modelo KNN tenha alcançado o melhor desempenho preditivo quantitativo (menor RMSE global), o RF emerge como o ponto de equilíbrio metodológico mais favorável da pesquisa. Operacionalmente, a razoável interpretabilidade do RF decorre da sua capacidade matemática nativa de quantificar a importância de cada variável — extraída diretamente da redução da impureza (ganho de informação) nos nós de suas árvores durante o treinamento, o que gerou os percentuais apresentados na Tabela 7. O KNN, em contrapartida, atua como uma "caixa preta" geométrica baseada no cálculo de distâncias, dificultando a extração de sentido físico. Além disso, a flexibilidade operacional do RF permite a incorporação de novas *features* (como índices de sujidade ou umidade) bastando adicioná-las como novas colunas na matriz de dados de treinamento, sem qualquer necessidade de reformular equações governantes ou deduzir novos denominadores físicos.

Por fim, os resultados reforçam a eficácia primária das técnicas de ML para a estimativa de  $T_{mod}$  em condições de clima semiárido. A capacidade computacional desses modelos de

capturar interações não lineares complexas entre variáveis meteorológicas e temporais, contribuiu significativamente para o aumento da acurácia. Isso fica evidenciado pela redução substancial do erro preditivo em comparação aos modelos físicos e semiempíricos otimizados, consolidando os algoritmos de aprendizado supervisionado não apenas como alternativas viáveis, mas como ferramentas analíticas de superioridade comprovada para o monitoramento térmico em plantas FV inseridas em microclimas urbanos.

Por fim, embora os modelos de ML apresentem maior precisão, os modelos físico-semiempíricos mantêm sua relevância como ferramentas de referência, especialmente em contextos de dimensionamento preliminar de sistemas FV, análises de viabilidade técnico-econômica em fase de projeto ou em situações com disponibilidade limitada de dados históricos para treinamento de algoritmos de aprendizado supervisionado.

#### ***4.5.1 Fatores determinantes do desempenho relativo dos modelos***

A análise comparativa apresentada nas seções anteriores revelou diferenças marcantes de desempenho entre modelos de ML, correlações híbridas e formulações físicas convencionais. Esta seção aprofunda os mecanismos subjacentes a essas disparidades, explorando como características específicas do clima semiárido analisado e propriedades intrínsecas de cada abordagem de modelagem contribuíram para os resultados observados.

##### ***4.5.1.1 Previsibilidade climática e cobertura do espaço de estados***

O clima semiárido cearense apresenta padrões sazonais altamente estruturados e repetitivos, conforme caracterizado na Seção 4.1. A sazonalidade pluviométrica bem definida, característica da região, estabelece ciclo binário previsível: períodos de estação seca com céu predominantemente limpo e irradiância média de  $650 \text{ W/m}^2$ , alternando com estação chuvosa caracterizada por aumento de nebulosidade e redução de 20–30% na irradiância incidente. Esta estrutura cíclica, combinada com baixa amplitude térmica anual ( $3\text{--}4^\circ\text{C}$  entre médias mensais) típica de regiões semiáridas, reduz substancialmente a variabilidade interanual e a ocorrência de eventos meteorológicos extremos não previstos.

O conjunto de dados coletado capturou completamente dois ciclos sazonais que mapeiam as condições operacionais típicas dos módulos FV no semiárido. Esta densidade amostral, distribuída ao longo de todas as horas do dia (8h–17h) e estações do ano, permitiu aos modelos de ML explorar praticamente todo o espaço de estados meteorológicos relevantes:

combinações de  $G \in [10, 1250]$  W/m<sup>2</sup>,  $T_a \in [21, 7, 39, 0]$  °C,  $V_w \in [0, 15]$  m/s e posições solares características da latitude 3,74°S.

Durante a fase de teste, os modelos de ML raramente encontraram condições climáticas não representadas no conjunto de treinamento. A estabilidade climática e a repetitividade sazonal implicam que configurações meteorológicas observadas em janeiro de 2024, por exemplo, apresentam alta similaridade com aquelas de janeiro de 2025, permitindo que algoritmos supervisionados realizem interpolação local com boa precisão. Esta característica contrasta fortemente com climas temperados ou de monções irregulares, onde eventos extremos (ondas de calor inesperadas, tempestades convectivas isoladas) podem introduzir estados do sistema fora da distribuição de treino, degradando significativamente o desempenho de modelos puramente empíricos (TUCCI *et al.*, 2024; VOYANT *et al.*, 2017).

#### 4.5.1.2 Auto-calibração matemática versus parametrização generalista

A superioridade dos modelos de ML em relação aos modelos físicos convencionais reflete fundamentalmente diferenças nas estratégias de ajuste aos dados locais. Modelos físicos baseiam-se em equações derivadas de primeiros princípios termodinâmicos (balanço radiativo-convectivo), cujos coeficientes são frequentemente determinados em condições controladas de laboratório ou calibrados para climas temperados.

A análise dos modelos físicos ilustra perfeitamente este fenômeno: o modelo de *Fai-man* com parâmetros originais ( $U_0 = 25,00$  W/(m<sup>2</sup> · K),  $U_1 = 6,84$  W · s/(m<sup>3</sup> · K)) apresentou elevados valores de erro (RMSE = 7,15 °C, MBE = +2,92 °C), superestimando sistematicamente a  $T_{\text{mod}}$  nas condições estudadas. A otimização local via algoritmo de evolução diferencial resultou em parâmetros significativamente diferentes ( $U_0 = 35,26$  W/(m<sup>2</sup> · K),  $U_1 = 0,55$  W · s/(m<sup>3</sup> · K)), reduzindo o RMSE original em 31,3%. A queda drástica do parâmetro  $U_1$  (coeficiente convectivo do vento) de 6,84 para 0,55 ratifica que o resfriamento eólico é substancialmente menos eficaz neste experimento devido à combinação de velocidades de vento severamente atenuadas (média de 0,2 m/s), configuração urbana do LEA-UFC e elevada  $T_a$ , que diminui de forma contínua o gradiente térmico módulo-ambiente.

Os modelos de ML realizam implicitamente essa “auto-calibração” durante o treinamento supervisionado, operando sob diferentes paradigmas de aprendizado adaptativo orientados aos dados. Algoritmos baseados em árvores, por exemplo, ajustam milhares de pesos e estruturas hierárquicas complexas minimizando funções de perda via gradiente descendente ou particiona-

mento recursivo do espaço de *features*. O modelo KNN, por sua vez, alcança a auto-calibração de maneira topológica, baseada em instâncias (*lazy learning*): ele utiliza o próprio histórico de medições locais como um mapa multidimensional do clima, estimando a  $T_{mod}$  ao resgatar as respostas térmicas dos momentos passados mais meteorologicamente semelhantes a cada instante consultado. Independentemente do mecanismo, essa plasticidade inerente ao aprendizado estatístico permite que os algoritmos assimilem as particularidades do microclima do LEA-UFC com uma flexibilidade incomparavelmente superior à dos modelos físicos e suas formulações matemáticas restritas a poucos parâmetros globais.

#### 4.5.1.3 Captura de não-linearidades e interações locais

A relação entre variáveis meteorológicas e  $T_{mod}$  exhibe não-linearidades complexas frequentemente sub-representadas em formulações físicas simplificadas. A correlação linear otimizada (Equação 4.1), embora superior aos modelos físicos de referência (MAE de 2,78 °C contra  $\approx 3,57$  °C dos modelos de balanço térmico calibrados localmente), ainda assume proporcionalidade estrita entre  $T_a$ ,  $G$  e  $T_{mod}$ , ignorando efeitos de saturação térmica, acoplamentos quadráticos e dependências condicionais.

Modelos *ensemble* baseados em árvores (RF, XGBoost) capturam naturalmente essas não-linearidades sem especificação prévia da forma funcional. A Tabela 7 revela que  $G$  e  $T_a$  respondem conjuntamente por 70–80% da importância preditiva nos modelos em árvore, mas a contribuição relativa varia condicionalmente: XGBoost atribui maior peso a  $G$  (50,81%) que a  $T_a$  (29,85%), enquanto KNN inverte essa hierarquia ( $G$ : 15,84%,  $T_a$ : 47,63%). Esta divergência reflete diferenças algorítmicas na detecção de interações: árvores de decisão identificam limiares de irradiância onde o comportamento térmico muda qualitativamente, enquanto KNN captura vizinhanças locais no espaço multivariado onde pequenas variações em  $T_a$  induzem transições de fase no resfriamento convectivo.

Em condições de alta irradiância ( $G > 900$  W/m<sup>2</sup>) e vento nulo ( $V_w < 0,5$  m/s), observadas frequentemente durante estação seca,  $T_{mod}$  pode exceder 60°C (máximo registrado: 67,3°C). Neste regime, a radiação de onda longa emitida pelo módulo cresce com  $T_{mod}^4$  (*Lei de Stefan-Boltzmann*), introduzindo não-linearidade forte não capturada por modelos lineares mas implicitamente aprendida por RF através de *splits* condicionais em nós profundos das árvores. Similarmente, a interação entre baixo  $V_w$  e alta  $T_a$  em ambiente urbano obstruído reduz o coeficiente de transferência de calor convectivo de forma não-linear, efeito detectado

automaticamente por algoritmos de *boosting* mas ignorado em formulações que assumem  $U_1$  constante.

#### 4.5.1.4 Modelos híbridos: equilíbrio entre interpretabilidade física e desempenho

As correlações semiempíricas híbridas (Seção 4.4) demonstraram desempenho intermediário ( $MAE = 2,78\text{--}2,83^\circ\text{C}$ ,  $R^2 = 0,80\text{--}0,81$ ), posicionando-se estrategicamente entre a rigidez dos modelos físicos puros e a opacidade dos algoritmos de ML. A correlação linear (Equação 4.3), com apenas 4 parâmetros otimizados ( $a_0, a_1, a_2, a_3$ ), alcançou RMSE de  $3,60^\circ\text{C}$ , representando redução de 27% em relação ao modelo de *Ross* ( $RMSE = 4,91^\circ\text{C}$ ) mantendo estrutura interpretável: o coeficiente  $a_1 = 1,57$  quantifica diretamente a contribuição térmica de  $T_a$ , enquanto  $a_2 = 17,06$  expressa o aquecimento por unidade de irradiância normalizada.

Este balanceamento entre flexibilidade e transparência física torna modelos híbridos particularmente adequados para aplicações que demandam rastreabilidade de previsões, como certificação de sistemas FV, validação experimental de teorias de transferência de calor ou análises de sensibilidade regulatória. O parâmetro  $a_3 \approx 0$  na correlação linear confirma quantitativamente a baixa influência do vento observada empiricamente, fornecendo evidência numérica direta (não apenas importância relativa em *ensemble black-box*) de que  $V_w$  contribui negligivelmente nas condições específicas do LEA-UFC.

A equivalência de desempenho entre correlação linear e não-linear (RMSE diferindo apenas  $0,0004^\circ\text{C}$ ) sugere que, nas condições semiáridas estudadas, relações entre variáveis meteorológicas e  $T_{mod}$  são suficientemente próximas da linearidade local para que termos de interação complexos não agreguem valor preditivo significativo. Este resultado contrasta com observações em climas temperados variáveis, onde acoplamentos não-lineares entre umidade relativa, precipitação e ventos turbulentos podem exigir formulações de ordem superior. A simplicidade das relações lineares no semiárido cearense reflete a dominância de mecanismos termodinâmicos elementares (irradiância, condução através do encapsulamento) sobre processos secundários (convecção forçada variável, evaporação superficial).

A instalação experimental do LEA-UFC apresenta características de microclima urbano que desafiam a aplicação direta de formulações físicas com parametrização padrão. Como esses modelos de referência foram concebidos predominantemente para condições macroclimáticas de campo aberto e regimes de vento desobstruídos, eles apresentam limitações inerentes ao tentar representar o balanço térmico em um ambiente com barreiras aerodinâmicas locais,

conforme detalhado na Seção 4.1.4. A posição do anemômetro em área circundada por edifícios e vegetação arbórea resulta em rugosidade superficial elevada ( $z_0 \approx 0,5\text{--}1,0$  m, característica de zonas urbanas densas) (WIERINGA, 1992; GRIMMOND; OKE, 1999) que reduz  $V_w$  ao nível dos painéis. Modelos físicos como *Faiman* assumem implicitamente configurações de campo aberto<sup>4</sup>, onde o perfil logarítmico de vento e a camada limite atmosférica seguem formulações clássicas bem estabelecidas (MONIN; OBUKHOV, 1954; BUSINGER *et al.*, 1971).

Esta discrepância entre condições de projeto e realidade operacional introduz erro sistemático nos modelos físicos: o termo  $U_1 \cdot V_w$  na Equação 4.2 superestima o resfriamento convectivo porque (i) os valores de  $U_1$  calibrados em campo aberto não capturam a redução de turbulência em ambiente obstruído e (ii) a velocidade medida ( $V_w \approx 1,2$  m/s) não representa adequadamente o escoamento efetivo sobre a superfície dos módulos, que pode ser ainda menor devido a recirculação e zonas de estagnação induzidas por obstáculos próximos.

Adicionalmente, a interação específica dos ventos alísios de sudeste característicos do litoral cearense com a estrutura da bancada experimental introduz padrões de escoamento tridimensionais (vórtices, separação de camada limite) não capturados por correlações unidimensionais de transferência de calor convectivo. Modelos de ML, ao processar séries temporais de  $V_w$ ,  $G$  e  $T_a$  simultaneamente, aprendem empiricamente essas correlações anômalas sem necessidade de modelagem explícita da fluidodinâmica local.

#### 4.5.1.5 Implicações para generalização e transferibilidade

Embora modelos de ML apresentem desempenho superior no conjunto de dados do LEA-UFC, sua capacidade de generalização para outras instalações FV no semiárido brasileiro ou configurações experimentais distintas permanece incerta sem validação cruzada geográfica (VOYANT *et al.*, 2017; RENKEMA *et al.*, 2024a). A dependência estrita de padrões aprendidos de dados históricos implica que mudanças climáticas de longo prazo (aumento gradual de  $T_a$ , alterações em regimes de precipitação) ou eventos extremos raros (El Niño severo) podem degradar a acurácia preditiva de modelos puramente empíricos (TUCCI *et al.*, 2024; MOHAMMED *et al.*, 2019a).

Em contraste, modelos físicos, embora menos precisos nas condições específicas estudadas, mantêm interpretabilidade termodinâmica que facilita adaptação a novos contextos

<sup>4</sup> Terrenos com baixa rugosidade superficial ( $z_0 < 0,1$  m), típicos de áreas agrícolas ou desérticas com vegetação esparsa (STULL, 1988; KAIMAL; FINNIGAN, 1994).

(FAIMAN, 2008; SKOPLAKI; PALYVOS, 2009): um engenheiro pode ajustar  $U_0$  e  $U_1$  com base em medições curtas de campo ou simulações CFD da geometria local (BARYKINA; HAMMER, 2017a), processo mais direto que re-treinar redes neurais profundas com novas amostras. Este compromisso entre acurácia local e portabilidade sugere estratégia híbrida para aplicações industriais (SANTOS *et al.*, 2024): utilizar modelos físicos para dimensionamento preliminar e análises de viabilidade em fase de projeto (quando dados locais são escassos), seguido por calibração de algoritmos de ML após instalação e início de operação, quando séries temporais de monitoramento acumulam-se.

A análise de importância de variáveis oferece ponte entre ambas as abordagens: a dominância empírica de  $G$  e  $Ta$  valida as premissas físicas centrais de modelos termodinâmicos (SKOPLAKI; PALYVOS, 2009; CUCE *et al.*, 2017), enquanto a baixa contribuição de  $Vw$  justifica simplificações estruturais que eliminam termos convectivos em instalações urbanas obstruídas, reduzindo complexidade sem comprometer fidelidade (PATEL *et al.*, 2024b). Futuras iterações de modelagem híbridas poderiam incorporar esta *insight*, fixando  $a_3 = 0$  na Equação 4.3, eliminando  $Vw$  e redistribuindo graus de liberdade para capturar efeitos de segunda ordem como acoplamento  $Ta-G$  ou dependência angular da emissividade efetiva do céu (KING *et al.*, 2004; MATTEI *et al.*, 2006).

## 5 CONCLUSÕES

Esta dissertação investigou o desenvolvimento e a avaliação comparativa de modelos preditivos para  $T_{mod}$  em condições de clima semiárido no nordeste brasileiro, utilizando dados experimentais coletados ao longo de dois anos no LEA-UFC, em Fortaleza. O estudo confrontou três paradigmas de modelagem: algoritmos de ML, correlações semiempíricas híbridas e modelos físico-empíricos de balanço térmico consolidados na literatura.

Os resultados demonstraram a superioridade clara e consistente dos modelos de ML sobre as abordagens convencionais. O algoritmo KNN alcançou o melhor desempenho no conjunto de teste, com MAE de 0,83 °C, RMSE de 1,44 °C e  $R^2 = 0,97$ , explicando 97% da variância observada em  $T_{mod}$ . Os modelos baseados em árvores de decisão — RF e XGBoost — confirmaram a eficácia de métodos *ensemble* para capturar relações não-lineares complexas, alcançando MAE entre 1,55 °C e 1,70 °C. Este patamar de precisão (RMSE de 1,44 °C a 2,11 °C) insere a presente pesquisa na fronteira do estado da arte da literatura mundial, superando a faixa típica de erro (1,8 °C a 3,5 °C) frequentemente reportada para modelos de predição térmica em outros contextos climáticos. Em contraste, os modelos físicos de referência, mesmo após rigorosa otimização local paramétrica, apresentaram desempenho inferior (MAE de  $\approx 3,57$  °C e RMSE de 4,91 °C), além de um viés sistemático negativo (MBE  $\approx -1,86$  °C), evidenciando tendência persistente de subestimação da temperatura.

A análise revelou que três mecanismos principais determinam esse desempenho diferencial. Primeiro, o clima semiárido cearense apresenta padrões altamente estruturados, com baixa amplitude térmica e irradiação constante. Esta estabilidade, combinada com o robusto conjunto de dados (273.310 observações válidas cobrindo dois ciclos sazonais), permitiu aos modelos de ML mapear exaustivamente as condições operacionais, realizando uma interpolação topológica com alta precisão na fase de teste. Segundo, os modelos de ML realizam uma auto-calibração intrínseca durante o treinamento, ajustando de pesos a limites de decisão para otimizar o desempenho local. Este processo supera os modelos físicos de referência que, apesar de calibrados empiricamente via algoritmos evolutivos, permanecem engessados por equações governantes estruturalmente rígidas concebidas inicialmente para climas temperados. Terceiro, modelos como o *Random Forest* capturam naturalmente dinâmicas complexas sem a necessidade de formulação analítica prévia, contornando a dificuldade de modelar matematicamente as perdas não-lineares de calor do módulo para o ambiente urbano.

As correlações semiempíricas desenvolvidas neste trabalho posicionaram-se estrate-

gicamente em um patamar intermediário, com MAE entre 2,78 °C e 2,83 °C, representando uma melhoria de  $\approx 22\%$  em relação aos modelos físicos clássicos otimizados e uma redução superior a 68% no viés sistemático. É crucial destacar que a equivalência estatística entre as correlações linear e não-linear revelou que a relação multivariada ambiental em Fortaleza é suficientemente contínua para que termos matemáticos de interação complexa não agreguem valor preditivo prático. Este resultado simplifica a implementação na engenharia local, sugerindo a dominância de mecanismos elementares (irradiância e temperatura ambiente) sobre processos convectivos secundários.

A análise de importância de variáveis corroborou perfeitamente esta hipótese termodinâmica:  $G$  e  $T_a$  responderam conjuntamente por 73,8% a 80,7% da importância preditiva nos modelos baseados em árvores. Em contrapartida,  $V_w$  apresentou contribuição estritamente marginal, inferior a 2% em todos os modelos. Este achado reflete fielmente as características do microclima urbano do experimento, onde o arranjo cercado por edifícios e vegetação atua como uma barreira aerodinâmica, resultando em velocidades médias de vento severamente atenuadas (0,2 m/s). A nulidade virtual de  $V_w$  justifica e embasa a adoção de simplificações estruturais na modelagem de usinas fotovoltaicas de telhado (*rooftop*) em ambientes urbanos, reduzindo o custo computacional e de sensoriamento sem comprometer a fidelidade.

Embora os algoritmos de ML apresentem desempenho imbatível no conjunto de dados do LEA-UFC, a dependência estrita do mapeamento de dados históricos implica que sua generalização para geografias distintas ou frente a anomalias climáticas de longo prazo exige retreinamento. Por outro lado, modelos físicos preservam a portabilidade e a dedução teórica. Isso consolida a tese de que a estratégia mais promissora para aplicações na indústria fotovoltaica é a modelagem híbrida e sequencial: empregar modelos de balanço físico calibrados por premissas para a fase de projeto e análise de viabilidade, migrando a planta para sistemas supervisórios guiados por algoritmos de ML à medida que as séries temporais de monitoramento operacional se acumulam e estabilizam.

## 6 PRODUÇÃO CIENTÍFICA DERIVADA DESTA DISSERTAÇÃO

Durante o desenvolvimento deste trabalho, foram produzidos e submetidos artigos científicos relacionados à temática investigada. O primeiro trabalho, intitulado *A Review of Thermal Behavior of Floating Photovoltaic Under Tropical Climates* (COSTA *et al.*, 2025), foi publicado na revista *IEEE Latin America Transactions* e apresenta uma revisão abrangente do comportamento térmico de sistemas fotovoltaicos flutuantes em clima tropical, fundamentando discussões teóricas correlatas a esta pesquisa.

O segundo trabalho, intitulado *Machine learning models for photovoltaic modules operating temperature estimation in semiarid climate* (COSTA *et al.*, 2026), foi aceito para publicação na revista *Solar Energy Advances*. Este artigo é fruto direto da metodologia e dos resultados apresentados nesta dissertação, validando o uso de modelos de aprendizado de máquina para a estimativa da temperatura de operação de módulos fotovoltaicos em condições de clima semiárido.

## REFERÊNCIAS

- AHMED, K.; SHAHID, S.; CHUNG, E.-S.; WANG, X.-j.; HARUN, S. B. Machine learning approaches for meteorological drought prediction: A comparative study for australia. **Computers & Electronics in Agriculture**, Elsevier, v. 173, p. 105446, 2020.
- ALFARO, E.; HIDALGO, H.; MORA, N. P. Climate variability of precipitation in ceara region, northeast of brazil. **Revista Brasileira de Climatologia**, v. 33, n. 19, p. 491–512, 2023.
- ALY, S. P.; BARTH, N.; FIGGIS, B. W.; AHZI, S. A fully transient novel thermal model for in-field photovoltaic modules using developed explicit and implicit finite difference schemes. **Journal of Computational Science**, Elsevier, v. 30, p. 189–204, 2018.
- AMODIO, A.; D'ANGOLA, A.; ENESCU, D.; MUNTEANU, R. E. Semi-analytical electro-thermal modelling of a photovoltaic module for evaluation of spatial temperature distribution. In: **2023 58th International Universities Power Engineering Conference (UPEC)**. [S. l.]: IEEE, 2023. p. 1–6.
- ANTONANZAS, J.; OSORIO, N.; ESCOBAR, R.; URRACA, R.; PISON, F. J. Martinez-de; ANTONANZAS-TORRES, F. Review of photovoltaic power forecasting. **Solar Energy**, Elsevier, v. 136, p. 78–111, 2016.
- ARARIPE, A. T. C.; SANTOS, L. de O.; SOUZA, J. J. S. de; FRATE, C. A.; CARVALHO, P. C. M. de. Monitoramento da temperatura de operação em módulos fotovoltaicos. In: **Anais Congresso Brasileiro de Energia Solar-CBENS**. [S. l.: s. n.], 2024.
- ASSIS, D. P. D.; SANTOS, L. D. O.; PEREIRA, R. I. S.; SOUZA, J. J. S. D.; CARVALHO, P. C. M. D. Scalable data acquisition system for real-time monitoring of photovoltaic plants. **IEEE Access**, IEEE, 2025.
- ASTM INTERNATIONAL. **Standard Test Methods for Electrical Performance of Nonconcentrator Terrestrial Photovoltaic Modules and Arrays Using Reference Cells**. 1996.
- ATLAS SOLAR E EÓLICO. **Atlas Eólico e Solar: Ceará**. Fortaleza: ADECE, 2019. 188 p. Disponível em: <<http://atlas.adece.ce.gov.br/>>. Acesso em: 28 dez. 2025. ISBN 978-85-67342-05-4.
- BAKHT, M. P.; MOHD, M. N. H.; IBRAHIM, B. S. K. S. M.; KHAN, N. Advanced automated machine learning framework for photovoltaic power output prediction using environmental parameters and shap interpretability. **Results in Engineering**, Elsevier, v. 25, p. 103584, 2025.
- BANIK, R.; BISWAS, A. Enhanced renewable power and load forecasting using rf-xgboost stacked ensemble. **Electrical Engineering**, Springer, 2024.
- BARROS, A. P. F. Predição da temperatura operacional das células fotovoltaicas: uma análise dos modelos para as condições do clima semiárido brasileiro. In: . Fortaleza: [S. n.], 2022. Disponível em: <https://repositorio.ufc.br/handle/riufc/65804>. Acesso em: 11 jan. 2026.
- BARROS, A. P. F.; SANTOS, L. d. O.; PEREIRA, R. I. S.; CARVALHO, P. C. M. d. Comparison of temperature forecasting models of photovoltaic modules operating under tropical climate. In: ABCM. **Proceedings of the 26th International Congress of Mechanical Engineering (COBEM 2021)**. Virtual Congress, Brazil, 2021.

BARRY, J.; MUNEEER, T.; KOTAK, Y. Dynamic model of photovoltaic module temperature as a function of atmospheric conditions. **Advances in Science and Research**, v. 17, p. 165–173, 2020. Indica que modelos estacionários como Faiman são derivados de condições de teste padrão (NOCT: 800 W/m<sup>2</sup>, 20°C, 1 m/s) que divergem significativamente de condições tropicais urbanas. Disponível em: <https://asr.copernicus.org/articles/17/165/2020/asr-17-165-2020.pdf>.

BARYKINA, E.; HAMMER, A. Comparison of temperature models for photovoltaic modules. **Energy Procedia**, Elsevier, v. 130, p. 22–28, 2017.

BARYKINA, E.; HAMMER, A. Modeling of photovoltaic module temperature using faiman model: Sensitivity analysis for different climates. **Solar Energy**, v. 146, p. 401–416, 2017.

BREIMAN, L. Random forests. **Machine learning**, Springer, v. 45, n. 1, p. 5–32, 2001.

BUSINGER, J.; WYNGAARD, J.; IZUMI, Y.; BRADLEY, E. Flux-profile relationships in the atmospheric surface layer. In: **Workshop on Micrometeorology**. Boston: American Meteorological Society, 1971. p. 67–100.

BUSSON, B. d. O.; DIAS, P. H. F.; DUPONT, I. M.; CAMPOS, P. H. M.; CARVALHO, P. C. M. d.; BARROSO, E. A. Q. Validação de modelos de comportamento térmico de painéis fotovoltaicos para o semiárido brasileiro. In: **Anais do VII Congresso Brasileiro de Energia Solar (CBENS 2018)**. Gramado, RS: [S. n.], 2018.

BUSSON, B. d. O.; SANTOS, L. d. O.; CARVALHO, P. C. M. d.; FILHO, C. d. O. C. Experimental assessment and modeling of a floating photovoltaic module with heat bridges. **IEEE Latin America Transactions**, IEEE, v. 19, n. 12, p. 2079–2086, 2021.

CAMPOS, R. A.; MARTINS, G. L.; RÜTHER, R. Aplicação de machine learning para a previsão de temperatura de módulos solares fotovoltaicos. In: **Anais Congresso Brasileiro de Energia Solar-CBENS**. [S. l.: s. n.], 2020.

CAMPOS, R. A.; MARTINS, G. L.; RÜTHER, R. Aplicação de machine learning para a previsão de temperatura de módulos solares fotovoltaicos. In: **Anais do VIII Congresso Brasileiro de Energia Solar (CBENS 2020)**. Fortaleza, CE: [S. n.], 2020.

CAMPOS, R. e. a. Prediction of pv module temperature using machine learning and solar position data. **Renewable Energy**, 2020.

CANTILLO-LUNA, S.; MORENO-CHUQUEN, R.; CELEITA, D.; ANDERS, G. Deep and machine learning models to forecast photovoltaic power generation. **Energies**, MDPI, v. 16, n. 10, p. 4097, 2023.

CHANDEL, R.; CHANDEL, S.; KHOSLA, A. Modelling and experimental investigation of cooling of field-operating pv panels using thermoelectric devices for enhanced power generation by industrial solar plants. **Next Energy**, Elsevier, v. 2, p. 100086, 2024.

CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: **Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining**. [S. l.: s. n.], 2016. p. 785–794.

COSKUN, C.; TOYGAR, U.; SARPDAG, O.; OKTAY, Z. Sensitivity analysis of implicit correlations for photovoltaic module temperature: A review. **Journal of Cleaner Production**, v. 164, p. 1474–1485, 2017.

COSTA, A. W. D. da; SANTOS, L. O.; CARVALHO, P. C. M. de. A review of thermal behavior of floating photovoltaic under tropical climates. **IEEE Latin America Transactions**, IEEE, v. 23, n. 11, p. 1050–1059, 2025.

COSTA, A. W. D. da; SANTOS, L. O.; CARVALHO, P. C. M. de. Machine learning models for photovoltaic modules operating temperature estimation in semiarid climate. **Solar Energy Advances**, 2026. Aceito para publicação.

COSTA, S. C. S.; KAZMERSKI, L. L.; DINIZ, A. S. A. C. Impact of soiling on si and cdte pv modules: Case study in different brazil climate zones. **Energy Conversion and Management: X**, Elsevier, v. 10, p. 100070, 2021.

CUCE, E.; GUCLU, T.; CUCE, P. M.; BALI, T. An accurate model for photovoltaic (pv) modules to determine electrical characteristics and thermodynamic performance parameters. **Energy Conversion and Management**, Elsevier, v. 146, p. 205–216, 2017.

DAVIS, M. W.; FANNEY, A. H.; DOUGHERTY, B. P. Prediction of building integrated photovoltaic cell temperatures. **Journal of Solar Energy Engineering**, v. 123, n. 3, p. 200–210, 2001.

DIDAVI, A. B. K.; AGBOKPANZO, R. G. *et al.* Comparative study of decision tree, random forest and xgboost performance in forecasting the power output of a photovoltaic system. In: **2021 4th International Conference on Power Electronics and Renewable Energy Applications (ICPEREA)**. [S. l.]: IEEE, 2021. p. 1–6.

DOWS, R.; GOUGH, E. **PVUSA procurement, acceptance, and rating practices for photovoltaic power plants**. [S. l.], 1995.

DUFFIE, J. A.; BECKMAN, W. A. **Solar engineering of thermal processes**. [S. l.]: John Wiley & Sons, 2013.

FAIMAN, D. Assessing the outdoor operating temperature of photovoltaic modules. **Progress in Photovoltaics**, v. 16, n. 4, p. 307–315, 2008.

FERNANDEZ, E. F.; ALMONACID, F.; RODRIGO, P.; PEREZ-HIGUERAS, P. A two subcell equivalent solar cell model for concentrator photovoltaic systems. **Solar Energy**, v. 92, p. 221–233, 2013.

FILHO, J. F. B. d. F.; NEVES, W. L. A. *et al.* Photovoltaic module temperature prediction model incorporating wind direction and precipitation effects. **IEEE Journal of Photovoltaics**, IEEE, 2025.

FUENTES, M. K. **A Simplified Thermal Model for Flat-Plate Photovoltaic Arrays**. [S. l.], 1987.

FUNCEME. **Atlas Solarimétrico do Ceará**. Fortaleza, Ceará, 2019. Disponível em: <http://www.funceme.br/wp-content/uploads/2019/08/Atlas-Solarim%C3%A9trico-do-Cear%C3%A1.pdf>.

GARCÍA, M. A.; BALENZATEGUI, J. Estimation of photovoltaic module yearly temperature and performance based on nominal operation cell temperature calculations. **Renewable energy**, Elsevier, v. 29, n. 12, p. 1997–2010, 2004.

- GREEN, M. A.; DUNLOP, E. D.; HOHL-EBINGER, J.; YOSHITA, M.; KOPIDAKIS, N.; HAO, X. Solar cell efficiency tables (version 63). **Progress in Photovoltaics: Research and Applications**, v. 32, n. X, p. XX–XX, 2024.
- GRIMMOND, C.; OKE, T. Aerodynamic properties of urban areas derived from analysis of surface form. **Journal of Applied Meteorology**, American Meteorological Society, v. 38, n. 9, p. 1262–1292, 1999.
- GU, W.; MA, T.; SHEN, L.; LI, M.; ZHANG, Y.; ZHANG, W. Coupled electrical-thermal modelling of photovoltaic modules under dynamic conditions. **Energy**, Elsevier, v. 188, p. 116043, 2019.
- HANIFAH, M. M. M.; ZAINUDDIN, H.; NASIR, M. N. M.; IBRAHIM, M. Z. Establishment of standard reference environment for photovoltaic nominal operating cell temperature testing with dedicated approach for tropical region. **Indonesian Journal of Electrical Engineering and Computer Science**, v. 12, n. 3, p. 1051–1058, 2018.
- HANNOUCH, A.; FARAJ, J.; TAHER, R.; MORTAZAVI, M.; KHALED, M. A simplified approach to modeling temperature dynamics in photovoltaic systems—validation, case studies, and parametric analysis. **International Journal of Thermofluids**, Elsevier, v. 22, p. 100609, 2024.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning**. [S. l.]: Springer, 2009.
- HORNUNG, T.; STEINER, M.; NITZ, P. Estimation of the influence of fresnel lens temperature on energy generation of a concentrator photovoltaic system. **Solar Energy Materials and Solar Cells**, Elsevier, v. 99, p. 333–338, 2012.
- IEA PVPS Task 13. **PV Performance Modeling Methods and Practices: Results from the 4th PV Performance Modeling Collaborative Workshop**. [S. l.], 2017. Relatório técnico detalhando métodos de modelagem térmica utilizados na indústria FV, incluindo modelos de Ross, Faïman e PVsyst. Disponível em: [https://iea-pvps.org/wp-content/uploads/2020/01/T13\\_Report\\_PV\\_Performance\\_Modeling\\_Methods\\_and\\_Practices\\_FINAL\\_March\\_2017.pdf](https://iea-pvps.org/wp-content/uploads/2020/01/T13_Report_PV_Performance_Modeling_Methods_and_Practices_FINAL_March_2017.pdf).
- INTERNATIONAL ELECTROTECHNICAL COMMISSION. **IEC 61215-1:2016 – Terrestrial photovoltaic (PV) modules – Design qualification and type approval – Part 1: Test requirements**. Geneva, Switzerland, 2016.
- INTERNATIONAL ENERGY AGENCY. **Renewable Energy Market Update 2024**. 2024. Acesso em: 28 ago. 2025. Disponível em: <https://www.iea.org/>.
- JAIN, M.; ALSKAIF, T.; DEV, S. Are deep learning models more effective against traditional models for load demand forecasting? In: IEEE. **2022 International Conference on Smart Energy Systems and Technologies (SEST)**. [S. l.], 2022. p. 1–6.
- JAKHRANI, A. Q.; OTHMAN, A.-K.; RIGIT, A. R. H.; SAMO, S. R. Determination and comparison of different photovoltaic module temperature models for kuching, sarawak. In: IEEE. **2011 IEEE Conference on Clean Energy and Technology (CET)**. [S. l.], 2011. p. 231–236.
- JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An introduction to statistical learning**. New York: Springer, 2013. v. 112.

- JONES, A.; UNDERWOOD, C. A thermal model for photovoltaic systems. **Solar Energy**, Elsevier, v. 70, n. 4, p. 349–359, 2001.
- KAIMAL, J.; FINNIGAN, J. **Atmospheric Boundary Layer Flows: Their Structure and Measurement**. New York: Oxford University Press, 1994. 289 p. ISBN 978-0-19-506239-7.
- KAPLANIS, S.; KAPLANI, E. A new dynamic model to predict transient and steady state pv temperatures taking into account the environmental conditions. **Energies**, MDPI, v. 12, n. 1, p. 2, 2018.
- KING, D. L.; KRATOCHVIL, J. A.; BOYSON, W. E. Temperature coefficients for pv modules and arrays: measurement methods, difficulties, and results. In: IEEE. **Conference record of the twenty sixth IEEE Photovoltaic Specialists Conference-1997**. [S. l.], 1997. p. 1183–1186.
- KING, D. L.; KRATOCHVIL, J. A.; BOYSON, W. E. **Photovoltaic array performance model**. [S. l.], 2004.
- KLEISSL, J.; WEST, R. Solar radiation and its importance for pv power generation. **Renewable Energy**, Elsevier, v. 75, p. 101–110, 2015.
- KOEHL, M.; HECK, M.; WIESMEIER, S.; WIRTH, J. Modeling of the nominal operating cell temperature based on outdoor weathering. **Solar Energy Materials and Solar Cells**, Elsevier, v. 95, n. 7, p. 1638–1646, 2011.
- KOEHL, M.; HECK, M.; WIESMEIER, S.; WIRTH, J. Outdoor weathering of pv modules—effects of various climates and comparison with accelerated laboratory testing. **Solar Energy Materials and Solar Cells**, Elsevier, v. 95, n. 7, p. 2046–2053, 2011.
- KUITCHE, J.; OH, J.; BRUNGER, A.; INOUE, T.; MULLER, M.; BAUERDICK, C.; ALTHAUS, J.; KIEHN, S.; FENG, V.; THERHAAG, U. *et al.* One year noct round-robin testing per iec 61215 standard. In: IEEE. **2011 37th IEEE Photovoltaic Specialists Conference**. [S. l.], 2011. p. 002380–002385.
- KUO, C.-F. J.; LEE, Y.-W.; UMAR, M. L.; YANG, P.-C. Dynamic modeling, practical verification and energy benefit analysis of a photovoltaic and thermal composite module system. **Energy Conversion and Management**, Elsevier, v. 154, p. 470–481, 2017.
- KURTZ, S.; WHITFIELD, K.; MILLER, D.; JOYCE, J.; WOHLGEMUTH, J.; KEMPE, M.; DHERE, N.; BOSCO, N.; ZGONENA, T. Evaluation of high-temperature exposure of rack-mounted photovoltaic modules. In: IEEE. **2009 34th IEEE Photovoltaic Specialists Conference (PVSC)**. [S. l.], 2009. p. 002399–002404.
- LABS, F. HelioScope: A holistic photovoltaic system design tool. **Solar Industry Magazine**, 2014. Plataforma comercial de design de sistemas FV com modelagem térmica baseada em modelos físico-empíricos. Disponível em: <https://www.helioscopesoftware.com>.
- LAWSON, C. L.; HANSON, R. J. **Solving least squares problems**. Philadelphia, PA: SIAM, 1995. v. 15. (Classics in Applied Mathematics, v. 15). Reedição da obra original de 1974.
- LEDMAOUI, Y.; MAGHRAOUI, A. E.; AROUSSI, M. E.; SAADANE, R. *et al.* Forecasting solar energy production: A comparative study of machine learning algorithms. **Energy Reports**, Elsevier, v. 10, p. 1004–1012, 2023.

- LI, P.; ZHOU, K.; LU, X.; YANG, S. Photovoltaic power forecasting with a hybrid deep learning approach. **IEEE Access**, IEEE, v. 8, p. 175871–175880, 2020.
- LINARDATOS, P.; PAPASTEFANOPOULOS, V.; KOTSIANTIS, S. Explainable ai: A review of machine learning interpretability methods. **Entropy**, v. 23, n. 1, p. 18, 2021. Disponível em: <https://www.mdpi.com/1099-4300/23/1/18>.
- LIU, W.; LIU, Y.; ZHANG, T.; HAN, Y.; ZHOU, X.; XIE, Y.; YOO, S. Use of physics to improve solar forecast: Part ii, machine learning and model interpretability. **Solar Energy**, Elsevier, v. 245, p. 305–318, 2022.
- LOBERA, D. T.; VALKEALAHTI, S. Dynamic thermal model of solar pv systems under varying climatic conditions. **Solar Energy**, Elsevier, v. 93, p. 183–194, 2013.
- LUO, Y.; ZHANG, L.; WU, J.; LIU, Z.; WU, Z.; HE, X. Dynamical simulation of building integrated photovoltaic thermoelectric wall system: Balancing calculation speed and accuracy. **Applied Energy**, Elsevier, v. 203, p. 887–897, 2017.
- MAGALHÃES, G. B.; ZANELLA, M. E. Comportamento climático da região metropolitana de fortaleza. **Mercator**, Universidade Federal do Ceará, v. 10, n. 23, p. 129–145, 2011. ISSN 1984-2201.
- MARION, B.; KROPOSKI, B.; EMERY, K.; CUETO, J. del; MYERS, D.; OSTERWALD, C. **Validation of a Photovoltaic Module Energy Ratings Procedure at NREL**. [S. l.], 1999. Disponível em: <https://www.osti.gov/biblio/12187>.
- MATTEI, M.; NOTTON, G.; CRISTOFARI, C.; MUSELLI, M.; POGGI, P. Calculation of the polycrystalline pv module temperature using a simple method. **Renewable Energy**, v. 31, n. 4, p. 553–567, 2006.
- MEDEIROS, S. E. L.; NILO, P. F.; SILVA, L. P.; SANTOS, C. A. C. *et al.* Influence of climatic variability on the electricity generation potential by renewable sources in the brazilian semi-arid region. **Journal of Arid Environments**, Elsevier, v. 184, p. 104331, 2021.
- MELLIT, A.; PAVAN, A. M.; BENGHANEM, M. Machine learning and deep learning for photovoltaic applications: A comprehensive review. **Applied Energy**, Elsevier, v. 298, p. 117252, 2021.
- MERMOUD, A. **PVsyst: Software for the Study and Simulation of Photovoltaic Systems**. [S. l.], 2012. Disponível em: <http://www.pvsyst.com>.
- Meteotest AG. **Meteonorm 8.2: Global Meteorological Database for Solar Energy and Applied Meteorology**. Bern, Switzerland: [S. n.], 2023. Base de dados meteorológicos utilizada em conjunto com softwares comerciais de simulação FV. Disponível em: <https://meteonorm.com>.
- MOHAMMED, A.; ELBASSIONI, K.; KHONJI, M.; SAXENA, N. Machine learning and deep learning in energy systems: A review. **Sustainable Cities and Society**, Elsevier, v. 44, p. 729–743, 2019.
- MOHAMMED, H.; SINGH, R.; BAREDAR, P. Assessment of different correlations to estimate pv module operating temperature for an indian site. **Energy Science & Engineering**, v. 7, n. 3, p. 1032–1041, 2019.

- MONIN, A.; OBUKHOV, A. Basic laws of turbulent mixing in the surface layer of the atmosphere. **Trudy Geofizicheskogo Instituta, Akademiya Nauk SSSR**, v. 24, n. 151, p. 163–187, 1954.
- MORA, S.; PILIOUGINE, M.; CARRETERO, J.; NOFUENTES, G.; CARDONA, M.-d. Empirical model to predict the operating temperature of the modules of a photovoltaic system. **Int. J. Smart Grid Clean Energy**, v. 6, n. 1, p. 40–46, 2017.
- NASCIMENTO, L. R. do; BRAGA, M.; CAMPOS, R. A. *et al.* Performance assessment of solar photovoltaic technologies under different climatic conditions in Brazil. **Renewable Energy**, Elsevier, v. 146, p. 1070–1082, 2020.
- NESPOLI, A.; MUSSETTA, M.; OGLIARI, E.; LEVA, S.; FERNANDEZ-RAMIREZ, L.; GARCIA-TRIVINO, P. Machine learning techniques for photovoltaic applications: A survey. **Applied Energy**, Elsevier, v. 234, p. 1002–1017, 2019.
- NORDMANN, T.; CLAVADETSCHER, L. Understanding temperature effects on pv system performance. In: **3rd World Conference on Photovoltaic Energy Conversion**. [S. l.: s. n.], 2003. p. 2243–2246.
- OBIORA, C. N.; ALI, A.; HASAN, A. N. Implementing extreme gradient boosting (xgboost) algorithm in predicting solar irradiance. In: **2021 IEEE PES/IAS PowerAfrica**. [S. l.: IEEE, 2021. p. 1–5.
- PACHECO, F. J. S.; CARDONA, M. Sidrach de; CARRETERO, J.; PALACIOS, S. Photovoltaic system distributed monitoring for performance optimization. **Energy Procedia**, v. 74, p. 1417–1426, 2015.
- PATEL, N.; PIETERS, B. E.; BITTKAU, K.; SOVETKIN, E. Assessing the accuracy of two steady-state temperature models for onboard passenger vehicle photovoltaics applications. **Progress in Photovoltaics: Research and Applications**, Wiley Online Library, v. 32, n. 4, p. 234–248, 2024.
- PATEL, P. *et al.* Comprehensive review of pv module temperature models. **Renewable and Sustainable Energy Reviews**, Elsevier, XXX, p. XXX, 2024.
- PEEL, M. C.; FINLAYSON, B. L.; MCMAHON, T. A. Updated world map of the Köppen–Geiger climate classification. **Hydrology and Earth System Sciences**, v. 11, n. 5, p. 1633–1644, 2007.
- PVsyst SA. **PVsyst Version 8: Photovoltaic Software for System Design and Simulation**. 2023. Software comercial amplamente utilizado na indústria FV que implementa modelo de Faiman para simulação de temperatura de módulos. Disponível em: <https://www.pvsyst.com/en/products/pvsyst-8/>.
- PÉREZ-HIGUERAS, P. J.; MUÑOZ, E.; ALMONACID, G.; VIDAL, P. G. Simplified method for estimating direct normal irradiation. **Renewable and Sustainable Energy Reviews**, v. 16, n. 7, p. 5529–5534, 2012.
- RAHMAN, M. M.; HASANUZZAMAN, M.; RAHIM, N. A. Effects of various parameters on pv module power and efficiency. **Energy Conversion and Management**, v. 103, p. 348–358, 2015.

- RAUSCHENBACH, H. S. **Solar Cell Array Design Handbook**. New York: Van Nostrand Reinhold, 1980.
- REN21. **Renewables 2024 Global Status Report**. 2024. Acesso em: 28 ago. 2025. Disponível em: <https://www.ren21.net/>.
- RENKEMA, D.-J. e. a. Statistical approaches for solar module temperature estimation. **Solar Energy**, 2024.
- RENKEMA, W. *et al.* Machine learning for photovoltaic module temperature prediction: A comparative study. **Solar Energy**, Elsevier, XXX, p. XXX–XXX, 2024.
- RENKEMA, Y.; VISSER, L.; ALSKAIF, T. Enhancing the reliability of probabilistic pv power forecasts using conformal prediction. **Solar Energy Advances**, Elsevier, v. 4, p. 100059, 2024.
- RODRIGO, P.; FERNÁNDEZ, E. F.; ALMONACID, F.; PEREZ-HIGUERAS, P. Review of methods for calculation of cell temperature in high concentration pv modules. **Renewable and Sustainable Energy Reviews**, v. 38, p. 478–488, 2014.
- ROSS, R. G. Interface design considerations for terrestrial solar cell modules. In: **12th IEEE Photovoltaic Specialists Conference**. [S. l.: s. n.], 1976. p. 801–806.
- SANTOS, L. d. O.; CARVALHO, P. C. M. d.; FILHO, C. d. O. C. Photovoltaic cell operating temperature models: A review of correlations and parameters. **IEEE Journal of Photovoltaics**, IEEE, v. 12, n. 1, p. 179–190, 2021.
- SANTOS, L. O.; SOUZA, F. A.; FILHO, C. O. C.; CARVALHO, P. C.; ALSKAIF, T.; PEREIRA, R. I. Hybrid modeling for photovoltaic module operating temperature estimation. **IEEE Journal of Photovoltaics**, IEEE, 2024.
- SCHIELZETH, H. Simple means to improve the interpretability of regression coefficients. **Methods in Ecology and Evolution**, Wiley Online Library, v. 1, n. 2, p. 103–113, 2010.
- SEGADO, P. M.; CARRETERO, J.; CARDONA, M. Sidrach-de. Models to predict the operating temperature of different photovoltaic modules in outdoor conditions. **Progress in Photovoltaics: Research and Applications**, Wiley Online Library, v. 23, n. 10, p. 1267–1282, 2015.
- SKOPLAKI, E.; PALLYVOS, J. A. Operating temperature of photovoltaic modules: A survey of pertinent correlations. **Renewable Energy**, Elsevier, v. 34, n. 1, p. 23–29, 2009.
- SOUZA, J. J. S. D.; SANTOS, L. O.; CARVALHO, P. C. Influence of manual cleaning on the thermal behavior of photovoltaic modules. **IEEE Access**, IEEE, 2025.
- SOUZA, J. J. S. de; CARVALHO, P. C. Experimental investigation of the soiling impact on the generation of a photovoltaic plant in an urban area of the brazilian semiarid. **Sustainable Energy Technologies and Assessments**, Elsevier, v. 73, p. 104153, 2025.
- STULL, R. **An Introduction to Boundary Layer Meteorology**. Dordrecht: Springer, 1988. 666 p. ISBN 978-90-277-2769-5.
- STULTZ, J. W.; WEN, L. C. **Thermal performance testing and analysis of photovoltaic modules in natural sunlight**. [S. l.], 1977.

TINA, G. M.; ABATE, R. Experimental verification of thermal behaviour of photovoltaic modules. In: IEEE. **MELECON 2008-The 14th IEEE Mediterranean Electrotechnical Conference**. [S. l.], 2008. p. 579–584.

TODESCHINI, R. k-nearest neighbour method: The influence of data transformations and metrics. **Chemometrics and intelligent laboratory systems**, Elsevier, v. 6, n. 3, p. 213–220, 1989.

TUCCI, M.; PIAZZI, A.; THOMOPULOS, D. Machine learning models for regional photovoltaic power generation forecasting with limited plant-specific data. **Energies**, MDPI, v. 17, n. 10, p. 2346, 2024.

VOYANT, C.; NOTTON, G.; KALOGIROU, S.; NIVET, M.-L.; PAOLI, C.; MOTTE, F.; FOUILLOY, A. Machine learning methods for solar radiation forecasting: A review. **Renewable Energy**, v. 105, p. 569–582, 2017.

WHEELER, A. P.; STEIN, J. S.; KING, B. H. Evaluating cell temperature models and the effect of wind speed on photovoltaic module temperature. **IEEE Journal of Photovoltaics**, v. 11, n. 4, p. 1046–1053, 2021. Demonstra que modelos de Faiman e Ross apresentam desempenho similar sob condições temperadas, mas requerem ajuste de parâmetros  $U_0$  e  $U_1$  para condições locais específicas. Disponível em: <https://docs.nrel.gov/docs/fy21osti/78950.pdf>.

WIERINGA, J. Updating the davenport roughness classification. **Journal of Wind Engineering and Industrial Aerodynamics**, Elsevier, v. 41, n. 1-3, p. 357–368, 1992.

YANDT, M. D.; COOK, J. P.; GORDON, J. M. Estimating cell temperature in a concentrating photovoltaic system. In: AMERICAN INSTITUTE OF PHYSICS. **AIP Conference Proceedings**. [S. l.], 2012. v. 1477, n. 1, p. 311–315.

ZHANG, R.; MIRZAEI, P. A.; CARMELIET, J. Prediction of the surface temperature of bipv modules. **Solar Energy**, v. 147, p. 151–163, 2017.

ZOUINE, M.; AKHSASSI, M.; ERRAISSI, N.; AARICH, N.; BENNOUNA, A.; RAOUFI, M.; OUTZOURHIT, A. Mathematical models calculating pv module temperature using weather data: Experimental study. In: SPRINGER. **Proceedings of the 1st International Conference on Electronic Engineering and Renewable Energy: ICEERE 2018, 15-17 April 2018, Saidia, Morocco 1**. [S. l.], 2019. p. 630–639.